Carlos Ivorra Castillo

LÓGICA MATEMÁTICA

No puedes encontrar la verdad con la lógica si no la has encontrado ya sin ella.

G.K. Chesterton

Índice General

Introducción

En el siglo XVII, las matemáticas iniciaron un rápido desarrollo que llevó en relativamente poco tiempo a los logros más espectaculares a la vez que a las polémicas más desconcertantes. Al mismo tiempo que los matemáticos descubrían las leyes que rigen el movimiento de los planetas o encontraban patrones sorprendentes en la distribución de los números primos, se enzarzaban en discusiones sobre si existen o no números infinitamente pequeños (que se consideraban nulos o no, según conviniera), o sobre si la gráfica de una función podía tener tramos verticales, o sobre cuál es el valor de la suma infinita:

$$1 - 1 + 1 - 1 + 1 - 1 + \cdots$$

sin que ningún argumento en favor de una u otra opinión pudiera considerarse concluyente. El matemático Niels Henrik Abel (1802–1829) llegó a afirmar:

> *Salvo la serie geométrica, no hay en todas las matemáticas una única serie infinita cuya suma se haya determinado rigurosamente.*

A finales del siglo XIX, estos problemas parecían superados. El trabajo de Cauchy, Weierstrass, Riemann, Dedekind, etc. había permitido dar definiciones rigurosas de todos los conceptos matemáticos, a partir de las cuales todas las cuestiones del estilo de las señaladas admitían una respuesta poco menos que trivial. Sin embargo, este proceso paulatino de fundamentación rigurosa de las matemáticas sufrió un grave contratiempo con el descubrimiento de las paradojas de la teoría de conjuntos de Georg Cantor.

Aunque no hubieran hecho énfasis en este concepto, los matemáticos trabajaban con conjuntos desde siempre: el conjunto de los números naturales, el conjunto de los números primos, el conjunto de los puntos del plano, el conjunto de los puntos de una elipse, etc. Y Cantor se había propuesto desarrollar una teoría general sobre conjuntos, pero con ella reveló un problema nada trivial que amenazaba la base misma del razonamiento matemático.

El teorema de Cantor Para entender qué pasó, consideremos un conjunto arbitrario X (que puede ser cualquiera de los que acabamos de citar, como el conjunto de los números naturales, etc.) Con la notación matemática moderna, es habitual escribir $x \in X$ para expresar que x es uno de los elementos del conjunto X. Un conjunto X tiene muchos subconjuntos, es decir, conjuntos A

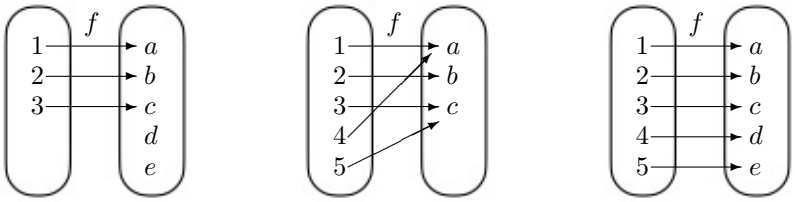
formados por parte de los elementos de X. Con la notación matemática moderna, esto se expresa así: $A \subset X$. Si X es el conjunto de los números naturales, entre sus subconjuntos tenemos el de los números pares, el de los números primos, el de los números menores que 100, etc.

Cantor llamó $\mathcal{P}X$ al conjunto formado por todos los subconjuntos de X, es decir, al conjunto que cumple

$$A \in \mathcal{P}X \quad \text{si y sólo si} \quad A \subset X.$$

En particular, vemos que no hay inconveniente en que los elementos de un conjunto puedan ser otros conjuntos. Tanto si X es finito como si es infinito, Cantor concluyó que $\mathcal{P}X$ es un conjunto estrictamente mayor que X, en el sentido siguiente:

Por una parte, podemos definir lo que en matemáticas se llama una aplicación inyectiva $f : X \longrightarrow \mathcal{P}X$, es decir, un criterio que a cada elemento $x \in X$ le asigna un elemento distinto $f(x) \in \mathcal{P}X$. En general, cuando podemos definir una aplicación inyectiva $f : A \longrightarrow B$ entre dos conjuntos, esto se interpreta como que el número de elementos del primero es menor o igual que el del segundo.



Aplicación inyectiva Aplicación suprayectiva Aplicación biyectiva

Para definir $f : X \longrightarrow \mathcal{P}X$ inyectiva basta definir $f(x) = \{x\}$, es decir, asociar a x el conjunto cuyo único elemento es x. Así pues, $\mathcal{P}X$ tiene al menos tantos elementos como X.

Pero, más aún, podemos decir que $\mathcal{P}X$ tiene un número de elementos estrictamente mayor (aunque ambos conjuntos sean infinitos), en el sentido de que es imposible definir una aplicación biyectiva $f : X \longrightarrow \mathcal{P}X$, es decir, un criterio que empareje los elementos de X con los de $\mathcal{P}X$. En general la existencia de una aplicación biyectiva entre dos conjuntos (aunque sean infinitos) se interpreta como que ambos tienen el mismo número de elementos.

Un poco más en general, vamos a ver que no puede existir ninguna aplicación $f : X \longrightarrow \mathcal{P}X$ suprayectiva: un criterio que a cada elemento $x \in X$ le hace corresponder un subconjunto $f(x) \subset X$ de modo que todo conjunto $A \subset X$ es de la forma $A = f(x)$ para algún $x \in X$ (tal vez para varios).

El argumento de Cantor es el siguiente: supongamos, por reducción al absurdo, que existiera tal aplicación suprayectiva f. Entonces, para cada elemento $x \in X$, podemos considerar $f(x) \in \mathcal{P}X$. Como x es un elemento de X y $f(x)$ es un subconjunto de X, pueden darse dos casos: o bien $x \in f(x)$, o bien $x \notin f(x)$. Por lo tanto, podemos fijarnos en los x que están en el segundo caso y definir el subconjunto $C \subset X$ formado por todos ellos, es decir,

$$x \in C \quad \text{si y sólo si} \quad x \in X \text{ y } x \notin f(x).$$

Pero entonces $C \in \mathcal{P}X$, luego, al ser f suprayectiva, se corresponde a través de f con un cierto elemento c de X, es decir, existe un $c \in X$ tal que $f(c) = C$. Y esto nos lleva a una contradicción: o bien $c \in f(c) = C$, o bien $c \notin f(c) = C$, pero si suponemos el primer caso, por definición de C, la condición $c \in f(c)$ implica que $c \notin C$, lo cual es absurdo, pero si suponemos que $c \notin f(c)$, de nuevo por definición de C, tendría que ser $c \in C$, y de nuevo llegamos a un absurdo.

Esta contradicción viene de suponer la existencia de f, luego concluimos, que, en efecto, es imposible asignar un subconjunto de X a cada elemento de X de forma que los subconjuntos asignados recorran todos los subconjuntos de X.

Esto es un profundo teorema de Cantor que tiene enormes repercusiones en la teoría de conjuntos, pues implica que existen conjuntos infinitos de diferentes tamaños, unos mayores que otros, pese a ser todos infinitos.

La paradoja de Cantor El problema surge cuando aplicamos el teorema de Cantor al conjunto universal V, formado por todos los conjuntos. Según el teorema de Cantor, no debería existir ninguna aplicación $f : V \longrightarrow \mathcal{P}V$ suprayectiva, pero, por otra parte, es muy fácil definir una: si $x \in V$ es un conjunto cuyos elementos son todos conjuntos,[1] es decir, si $x \subset V$, definimos $f(x) = x$, y en caso contrario definimos $f(x)$ de cualquier forma, por ejemplo, $f(x) = V$. Así tenemos definida una aplicación f que claramente es suprayectiva, ya que si $x \in \mathcal{P}V$, es decir, si $x \subset V$, tenemos trivialmente que $x \in V$, y entonces $f(x) = x$.

Esto es lo que se conoce como una *paradoja:* hemos llegado a una contradicción —existe una aplicación $f : V \longrightarrow \mathcal{P}V$ suprayectiva (porque acabamos de construir una) y a la vez no existe ninguna (por el teorema de Cantor)— que no viene de suponer nada por reducción al absurdo. Esta paradoja muestra que el razonamiento matemático que hemos aplicado para llegar a esta conclusión es contradictorio en sí mismo, pese a que no parte de ningún supuesto sospechoso de ser falso.

La paradoja de Russell Puestos a tratar de entender la paradoja de Cantor, lo natural es ver qué sucede al particularizar la demostración del teorema de Cantor al caso concreto del conjunto V de todos los conjuntos, que es el que da lugar a la paradoja. Para ello partimos de la aplicación $f : V \longrightarrow \mathcal{P}V$ que hemos construido y le aplicamos la demostración del teorema de Cantor. Esto supone considerar el conjunto C formado por los conjuntos $x \in V$ tales que $x \notin f(x)$. Si x no es un subconjunto de V, entonces $x \in V = f(x)$, luego no está en C. Así pues, los conjuntos de C son todos conjuntos $x \in \mathcal{P}V$, para los cuales $f(x) = x$, luego C viene dado por

$$x \in C \quad \text{si y sólo si} \quad x \subset V \quad \text{y} \quad x \notin x.$$

[1] En realidad, en la teoría de conjuntos moderna es habitual adoptar el convenio de que los elementos de cualquier conjunto son también conjuntos. En tal caso, todo $x \in V$ cumple $x \subset V$, y el recíproco es obvio (todo $x \subset V$ es un conjunto, luego $x \in V$). Por lo tanto, resulta que $\mathcal{P}V = V$ y trivialmente tenemos una biyección $f : V \longrightarrow \mathcal{P}V$ dada por $f(x) = x$, pero la paradoja de Cantor no depende de que adoptemos o no este convenio.

Si continuamos con el argumento, veremos que, en realidad, la condición $x \subset V$ es superflua,[2] de modo que podemos refinar la paradoja de Cantor considerando simplemente el conjunto R de todos los conjuntos x que no se pertenecen a sí mismos:

$$x \in R \quad \text{si y sólo si} \quad x \notin x.$$

Por ejemplo, es obvio que $V \in V$, luego $V \notin R$, mientras que el conjunto $\mathbb{N}$ de los números naturales no es un número natural, luego $\mathbb{N} \notin \mathbb{N}$ y, por lo tanto, $\mathbb{N} \in R$. Pero la paradoja de Cantor reaparece cuando nos planteamos si $R \in R$ o si, por el contrario, $R \notin R$. Si suponemos que $R \in R$, entonces R es un conjunto que se pertenece a sí mismo, luego no cumple el criterio dado por la definición de R, luego debe ser $R \notin R$, y tenemos una contradicción. Pero si suponemos que $R \notin R$, entonces R es un conjunto que no se pertenece a sí mismo y, por definición de R, debería ser $R \in R$, y tenemos una contradicción en ambos casos.

Esto es lo que se conoce como la *paradoja de Russell* (por el filósofo Bertrand Russell): el conjunto de todos los conjuntos que no se pertenecen a sí mismos es contradictorio por un argumento simple y directo, pese a que tiene perfecto sentido distinguir entre conjuntos que se pertenecen a sí mismos (como V, o el conjunto de todos los conjuntos infinitos, etc.) y conjuntos que no se pertenecen a sí mismos (como $\mathbb{N}$, o el conjunto de todos los conjuntos finitos, etc.).

Cantor, conocedor de varias paradojas de este tipo, introdujo la distinción entre "conjuntos" y "multiplicidades inconsistentes", que eran colecciones de objetos que daban lugar a contradicciones en cuanto se las quería tratar como conjuntos, como es el caso del conjunto V de todos los conjuntos, o del conjunto de Russell R y otros más que aparecían en su teoría de conjuntos, como el conjunto Ω de todos los ordinales, etc., y observó que las "multiplicidades inconsistentes" eran siempre conjuntos muy grandes. Pero esto no era satisfactorio. ¿Cómo podemos estar seguros de que el conjunto $\mathbb{N}$ de los números naturales no es también una multiplicidad inconsistente? Si el razonamiento matemático puede dar lugar a paradojas inexplicables al tratar con ciertos conjuntos, ¿quién nos asegura que no se pueda llegar a contradicciones similares a partir de un conjunto aparentemente inofensivo como $\mathbb{N}$, aunque de momento no las hayamos encontrado?

La teoría axiomática de conjuntos Entre los filósofos de principios del siglo XX saltó la alarma: las matemáticas estaban derrumbándose, pues eran contradictorias, nada en ellas era fiable. Más aún, en las paradojas de la teoría de conjuntos no parecía estar fallando realmente la matemática, pues el contenido matemático de las dos paradojas que hemos discutido es mínimo, sobre todo en el caso de la paradoja de Russell. Lo que fallaba era la propia lógica del razonamiento matemático. Los matemáticos, en general, se lo tomaron menos apasionadamente. En el fondo entendían que para no tener problemas con las

[2] Pues si adoptamos el convenio de que todos los elementos de un conjunto son conjuntos, entonces $x \subset V$ se cumple trivialmente.

paradojas bastaba olvidarse de ellas, y nada les impedía seguir desarrollando el álgebra, el análisis, la topología, etc. como siempre habían hecho, sin preocuparse por estos asuntos. No obstante, sí que era cierto que algo no estaba bien en la base de las matemáticas y que convenía entender el problema y resolverlo.

En 1902, Bertrand Russell recordó que un matemático llamado Gottlob Frege le había enviado hacía un tiempo un manuscrito titulado *Leyes básicas de la aritmética,* en el que pretendía deducir la aritmética a partir de la lógica pura, mediante una teoría axiomática minuciosamente razonada y detallada. De hecho, era tan técnica y enrevesada, que la primera parte —ya publicada— había pasado sin pena ni gloria entre la comunidad matemática, y el propio Russell había dejado en un cajón el manuscrito de la segunda parte durante año y medio.

El trabajo de Frege no era una teoría de conjuntos propiamente dicha, no permitía demostrar en ella los resultados generales de la teoría de conjuntos cantoriana, pero indirectamente hablaba de conjuntos, y Russell, al acordarse del manuscrito, se planteó estudiar cómo evitaba el puntilloso Frege la paradoja que había descubierto, pero se encontró con que no la evitaba. Las leyes lógicas propuestas por Frege daban por válida una variante de la paradoja adaptada al contexto. El resultado era que toda la teoría de Frege era inútil: en ella se podía demostrar igual que $2 + 2 = 4$ o que $2 + 2 = 7$. Y no se veía arreglo posible.

Esto llevó al propio Russell a tratar de hacer por sí mismo lo que Frege había tratado de hacer, pero no para fundamentar la aritmética, sino la teoría de conjuntos cantoriana y, con ello, toda la matemática. Entre 1910 y 1913, junto con Alfred North Whitehead, publicó los *Principia Mathematica,* una teoría axiomática de conjuntos en la que se determinaba con absoluto rigor, al estilo de Frege, lo que había que considerar como un razonamiento matemático válido y que permitía demostrar todos los resultados de la teoría de conjuntos de Cantor sin la posibilidad de definir el conjunto de todos los conjuntos, o el conjunto de Russell, de modo que ninguna paradoja conocida era demostrable en su teoría.

Pero los *Principia Mathematica* eran muy técnicos y prolijos, y unos años antes, en 1908, el matemático Ernst Zermelo había publicado un simplicísimo sistema de axiomas que lograba el mismo propósito: razonando a partir de ellos era posible demostrar los resultados básicos de la teoría de conjuntos, pero no se podía definir el conjunto de todos los conjuntos, o el conjunto de todos los conjuntos que no se pertenecen a sí mismos, ni ninguna de las "multiplicidades inconsistentes" identificadas por Cantor.

En realidad la teoría axiomática de Zermelo requirió unos retoques adicionales para abarcar toda la teoría de Cantor, con lo que se llegó a la teoría hoy conocida como ZFC (teoría de Zermelo-Fraenkel con el axioma de elección), que es la teoría axiomática más usada como referencia hoy en día en cuanto a la fundamentación de las matemáticas y como marco para el estudio de la teoría de conjuntos.

Así, los matemáticos comprendieron que para evitar las paradojas en la teoría de conjuntos bastaba razonar como siempre habían hecho sin más precaución que ceñirse en sus afirmaciones sobre conjuntos a lo establecido por los axiomas

de la teoría de Zermelo (o ZFC). Los profundos análisis lógico-filosóficos de los *Principia* eran prescindibles.

El razonamiento formal Según acabamos de explicar, la modesta teoría axiomática de Zermelo, o mejor, ZFC, conjuraban las paradojas de la teoría de conjuntos cantoriana, y permitieron que la teoría de conjuntos fuera finalmente reconocida como la culminación del proceso de formalización de la matemática. Vamos a analizar con un poco más de detalle en qué consistió esta solución a la crisis de los fundamentos de la matemática.

La teoría de conjuntos de Zermelo, al igual que ZFC, parte de dos conceptos fundamentales, el concepto de "conjunto" y el concepto de "pertenencia", es decir, establece que, entre dos conjuntos x e y puede darse o no una relación de pertenencia que expresamos con la notación $x \in y$. Estos conceptos no se definen, sino que su uso queda determinado por los axiomas. Reproducimos aquí algunos de ellos, para que el lector se haga una idea de en qué consisten:

- Si dos conjuntos tienen los mismos elementos, entonces son iguales.
- Dados dos conjuntos A y B, existe un tercer conjunto C cuyos elementos son exactamente A y B.
- Si A es un conjunto, existe otro conjunto B cuyos elementos son todos los subconjuntos de A.

A partir de éstos y unos pocos axiomas más, es posible definir todos los conceptos y demostrar todos los teoremas que vienen en cualquier libro de álgebra, de geometría, de análisis matemático, etc.

Por ejemplo, si un matemático nos dice que $\log(ab) = \log a + \log b$ y le preguntamos qué es eso de "logaritmo", nos podrá explicar que el logaritmo es una función $\log : \mathbb{R} \longrightarrow \mathbb{R}$ definida de tal forma, y si a continuación le preguntamos qué es "función" y qué es $\mathbb{R}$, nos podrá explicar lo que se entiende por ambos conceptos, y si seguimos preguntándole qué es, qué es, qué es, al final nos habrá reducido todo aquello de lo que nos ha hablado a las nociones de "conjunto" y "pertenencia", y el punto crucial viene cuando le preguntamos qué es un conjunto o qué significa que un conjunto pertenezca a otro.

La respuesta oficial que el matemático debe darnos en ese punto es que no puede darnos una definición de conjunto o pertenencia, pero que no hace falta hacerlo, pues basta saber que los conjuntos y la pertenencia cumplen los axiomas de la teoría de conjuntos. En otras palabras, la matemática se concibe así:

> *Existen unos objetos llamados conjuntos, que no vamos a decir lo que son, entre los cuales está definida una relación de pertenencia que no vamos a decir en qué consiste, pero suponemos que dichos conjuntos y dicha relación cumplen los axiomas de la teoría de conjuntos. Un teorema matemático es cualquier afirmación que pueda deducirse lógicamente de dichos axiomas.*

Si el lector piensa que, después de todo, es fácil definir "conjunto" y "pertenencia" sin más que establecer que un conjunto es una colección de elementos (o de otros conjuntos, según un convenio incluido en ZFC por el que todo es un conjunto), y que $x \in y$ significa simplemente que el conjunto x es uno de los elementos del conjunto y, debe descartar esa idea, porque si aceptamos esa definición, entonces la colección V de todos los conjuntos es una colección de conjuntos, luego es un conjunto, y ello nos lleva irremisiblemente a la paradoja de Cantor. Tenemos que admitir que los conjuntos son *algunas* colecciones de conjuntos, pero no podemos admitir que cualquier colección de conjuntos que podamos concebir se reconozca automáticamente como un conjunto sin incurrir en contradicciones. Casi todos los axiomas de ZFC son precisamente axiomas de existencia, que garantizan que determinadas colecciones de conjuntos son realmente los elementos de un conjunto (como el conjunto $\{x, y\}$ formado por dos conjuntos prefijados x, y, el conjunto $\mathcal{P}x$ de todos los subconjuntos de un conjunto dado x, etc.).

Ahora bien, lo que estamos diciendo es que la teoría axiomática de conjuntos es una especie de "subterfugio legal" que exime a los matemáticos de responder a la pregunta incómoda de qué es un conjunto (o qué es la relación de pertenencia). O más descaradamente: los matemáticos afirman que todos los objetos de los que hablan son conjuntos, y a la vez reconocen que no saben qué son los conjuntos, sino que meramente suponen que cumplen ciertos axiomas que toman como punto de partida de sus razonamientos. Bertrand Russell resumió esta situación con su célebre frase:

> *Las matemáticas podrían definirse como la ciencia en la que nunca sabemos de qué estamos hablando ni si lo que decimos es verdad.*

Pero no conviene que el lector se quede con una idea frívola de la situación. Habitualmente, cuando se dice de alguien que "no sabe de qué está hablando" es para concluir que dice incoherencias o cosas nada fiables, pero ¿estamos diciendo que los matemáticos no saben de lo que hablan y que, por lo tanto, no podemos fiarnos de nada de lo que dicen? Obviamente no, pero ello nos obliga a preguntarnos: ¿cómo puede alguien razonar sensatamente sobre algo y al mismo tiempo admitir que no sabe de qué está hablando? En este punto es donde interviene de forma esencial la lógica matemática. Como primera aproximación a lo que debemos entender por razonamiento lógico podríamos considerar lo siguiente:

> Un razonamiento lógico consiste en obtener unas afirmaciones (llamadas conclusiones) a partir de otras (llamadas premisas) con los criterios adecuados para que podamos tener la garantía de que si las premisas son verdaderas, entonces las conclusiones obtenidas también tienen que serlo necesariamente.

Por ejemplo, el razonamiento siguiente:

	Todos los españoles son europeos,
	Cervantes era español,
luego	*Cervantes era europeo.*

extrae la conclusión *"Cervantes era europeo"* a partir de las dos premisas precedentes y, en efecto, es un razonamiento en el sentido que acabamos de indicar. En cambio, algo muy parecido, como pueda ser

Todos los españoles son europeos,
Shakespeare no era español,
luego *Shakespeare no era europeo.*

no es un razonamiento válido,[3] pues las premisas son verdaderas y, pese a ello, la conclusión es falsa. Aquí es crucial entender que, para que un razonamiento sea válido, (si parte de premisas verdaderas) no basta con que sus conclusiones sean verdaderas, sino que tienen que ser necesariamente verdaderas. Por ejemplo, el razonamiento siguiente no es válido:

Todos los perros tienen cuatro patas,
Una gallina no es un perro,
luego *Una gallina no tiene cuatro patas.*

Es cierto que las gallinas no tienen cuatro patas, pero esto no podemos asegurarlo por el mero hecho de que las premisas sean verdaderas. Si las gallinas tuvieran cuatro patas, las premisas seguirían siendo ciertas, pero la conclusión no lo sería.

Esto nos lleva a que la validez o invalidez de un razonamiento no depende realmente de si las afirmaciones que involucra son verdaderas o falsas, pues sólo requiere que *en el supuesto* (que puede darse o no) de que sus premisas fueran verdaderas, su conclusión también lo sería (sin perjuicio de que tanto las premisas como la conclusión puedan ser falsas). Por ejemplo, el razonamiento siguiente es correcto:

Todos los españoles son americanos,
Molière era español,
luego *Molière era americano.*

Aquí, tanto las premisas como la conclusión son falsas, pero el razonamiento es válido porque si las premisas fueran verdaderas la conclusión también tendría que serlo.

Pero si la validez o invalidez de un razonamiento no depende de si las afirmaciones que involucra son verdaderas o falsas, ¿de qué depende entonces? La respuesta es que depende de la *forma* de las afirmaciones involucradas. Para entender esto consideremos la tabla siguiente:

1)	Todo A es B	2)	Todo A es B	3)	Todo A es B
	C es A		C no es A		C no es B
luego	C es B	luego	C no es B	luego	C no es A

La casilla 1) contiene un ejemplo de *forma* de razonamiento válida. Quiere decir que cualquier razonamiento que tenga esa forma, independientemente de

[3]Notemos que un razonamiento inválido o, mejor dicho, un razonamiento falaz, no es realmente un razonamiento, en el mismo sentido en que una pistola falsa no es una pistola.

qué palabras pongamos en lugar de A, B o C, será válido. De los cuatro ejemplos de razonamientos (válidos o no) que hemos visto más arriba, el primero y el cuarto tienen esta forma. En cambio, el segundo y el tercero tienen la forma de la casilla 2), por lo que, según hemos podido comprobar, 2) no es una forma válida de razonamiento. El hecho de que las premisas de 2) sean ciertas no nos ofrece garantía alguna de que su conclusión vaya a serlo también. Puede ser que sí (como en el caso de las gallinas) o puede ser que no (como en el caso de Shakespeare). Por otra parte, la casilla 3) contiene una forma de razonamiento que, superficialmente, se parece más a la de 2) que a la de 1) y, sin embargo, es una forma válida de razonamiento al igual que 1) y al contrario que 2).

En estos términos, la lógica matemática se encarga de encontrar las formas de razonamiento válidas, es decir, las reglas de inferencia que garantizan que una conclusión será verdadera si lo son las premisas de partida atendiendo únicamente a la forma (a la estructura lógica) de las premisas y la conclusión, de modo que puedan usarse mecánicamente, sin necesidad de preocuparse por el significado de las afirmaciones consideradas.

Llegados a este punto podemos destacar un hecho fundamental: no corresponde a la lógica "definir" lo que es un razonamiento válido. Las formas válidas de razonamiento son las que son, y no estamos en posición de decidir cuáles queremos dar por buenas y cuáles no. El propósito de la lógica es más bien "capturar" el razonamiento, dar criterios precisos que nos permitan distinguir con claridad los razonamientos propiamente dichos de las falacias que aparentan serlo. La palabra técnica en lugar de ese "capturar" que hemos empleado es "formalizar". Formalizar un razonamiento es expresarlo de tal modo que se pueda justificar que es válido atendiendo únicamente a la forma de las afirmaciones involucradas, sin necesidad de considerar para nada, no ya si éstas son verdaderas o falsas, sino siquiera su posible significado.

Ahora podemos precisar que ZFC no es sólo un conjunto de axiomas, sino que es un conjunto de axiomas junto con un cálculo deductivo, un criterio que especifica en qué condiciones podemos decir que una afirmación se deduce formalmente de los axiomas, es decir, sin requerir conocimiento alguno de qué son esos conjuntos (y esa relación de pertenencia) de los que hablan los axiomas.

En los primeros capítulos de este libro presentaremos un cálculo deductivo formal que nos permitirá entender cómo los matemáticos pueden a la vez razonar sobre conjuntos y decir que no saben lo que son los conjuntos sin volverse sospechosos de estar diciendo necedades. Más precisamente, el hecho de que podamos definir con total exactitud qué es un razonamiento formal hace que la "postura oficial" del matemático que se escabulle de explicar qué entiende por "conjunto" o "pertenencia" —alegando que los axiomas le eximen de hacerlo— pueda ser perfectamente defendida sin fisuras, por cínica que pueda parecer, pues determina con todo rigor y objetividad qué es una demostración o un teorema matemático.

La fundamentación de la matemática en términos de ZFC (incluyendo en ZFC el cálculo deductivo formal) levanta un muro de contención entre la mate-

mática y la filosofía de las matemáticas: se puede filosofar cuanto se quiera sobre qué son realmente los conjuntos, si existen objetivamente o son una creación subjetiva, etc., y en general cada matemático suele tener sus ideas al respecto más ingenuas o más elaboradas (sin limitarse necesariamente a la "respuesta oficial" del "a mí no me preguntes qué es un conjunto"), pero ninguna de estas discusiones afecta en nada a lo único relevante en el trabajo de un matemático: determinar si ciertas afirmaciones son o no teoremas matemáticos.

El cálculo deductivo formal es totalmente mecánico y objetivo. En teoría se puede programar a un ordenador para que, si le damos una demostración matemática suficientemente detallada, pueda comprobar que, efectivamente, el razonamiento es correcto sin más que comprobar que cada afirmación del argumento se sigue de otras precedentes mediante reglas formales lógicamente válidas, por lo que es inconcebible que ningún resultado matemático cuya corrección formal esté debidamente verificada pueda ponerse en cuestión a raíz de ninguna clase de consideración filosófica sobre la naturaleza de los conjuntos. El único problema podría ser que se descubriera que los axiomas de la teoría de conjuntos son contradictorios, en cuyo caso sería necesario encontrar una teoría alternativa.

Los teoremas de incompletitud de Gödel Hemos dicho que la teoría de Zermelo, o ZFC, resolvió la crisis de los fundamentos de la matemática abierta con el descubrimiento de las contradicciones de la teoría de conjuntos, pero esto hay que entenderlo en un sentido débil: estas teorías invalidan los argumentos de las paradojas conocidas, pero eso no es garantía de que no puedan existir otros argumentos que permitan deducir contradicciones de los axiomas de la teoría de conjuntos. De hecho, Zermelo publicó su teoría después de haber fracasado en un intento de justificar que no era contradictoria.

En la década de 1920, David Hilbert trazó un ambicioso programa consistente en analizar en profundidad el razonamiento matemático con el fin de encontrar un sistema de axiomas (posiblemente una teoría axiomática de conjuntos, tal vez la propia ZFC) del que pudiera demostrarse que era consistente (es decir, que en él no puedan probarse contradicciones) y completo (que toda afirmación matemática sea demostrable o refutable en él). Este programa fue uno de los incentivos para el desarrollo de la lógica matemática, es decir, del estudio de las leyes de razonamiento matemático más allá del sistema de axiomas de partida considerado, pero en 1931 la comunidad matemática recibió como un jarro de agua fría unos resultados publicados por Kurt Gödel, que esencialmente probaban que el programa de Hilbert era inviable.

Por una parte, Gödel demostró que cualquier teoría matemática que cumpliera unos requisitos mínimos para ser aceptable era necesariamente incompleta: siempre habría afirmaciones elementales sobre números naturales (elementales en el sentido de que no involucran más que la aritmética básica) que no podrían ser demostradas ni refutadas en la teoría. Pero, más aún, Gödel probó que la consistencia de una teoría axiomática razonable nunca puede ser demostrada en la propia teoría, por lo que, en particular, si consideramos una teoría del estilo de ZFC, que es capaz de probar cualquier cosa que un matemático considere que puede probar, resulta que si ZFC es consistente, es imposible

demostrar que lo es, ya que si existiera una demostración, ésta podría expresarse en el propio ZFC, y eso es justo lo que Gödel había probado que era imposible.

En resumen: los teoremas de Gödel demostraban que cualquier teoría matemática suficientemente potente como para aspirar a cumplir el programa de Hilbert es necesariamente incompleta, y que si es consistente es imposible demostrar que así es.

Estos hechos fascinaron a matemáticos y no matemáticos, pero, a la vez, las técnicas de la lógica matemática en general y las demostraciones de los teoremas de Gödel en particular resultaban un tanto crípticas incluso para matemáticos profesionales, por lo que Gödel y sus teoremas quedaron envueltos en una cierta aureola de misticismo y proliferaron los libros divulgativos.

En principio, los teoremas de Gödel son puramente negativos: sólo dicen que hay cosas que los matemáticos nunca podrán lograr, lo cual no ayuda mucho a la investigación en matemáticas. Pero, como veremos a continuación, hay otro aspecto de la fundamentación de las matemáticas en el que la lógica matemática desarrollada en las primeras décadas del siglo XX (e incluso los propios teoremas de incompletitud) resultaba esencial, y que mantuvo la atracción de los matemáticos y contribuyó en gran medida al desarrollo de la lógica matemática como una disciplina matemática más.

La hipótesis del continuo El teorema de Cantor que hemos analizado más arriba es una de las piezas de la teoría de cardinales transfinitos de Cantor. No estamos en condiciones aquí de explicarla con detalle, y lo que sigue ha de verse como una mera panorámica muy superficial.

Del mismo modo que a cada conjunto finito se le puede asociar un número de elementos (lo que usualmente se llama "contar"), Cantor definió un sistema de números para contar conjuntos infinitos, a los que llamó cardinales transfinitos. Los números naturales $0, 1, 2, \ldots$ son los cardinales finitos, pero tras ellos vienen los cardinales transfinitos:

$$0, \quad 1, \quad 2, \quad \ldots \quad \aleph_0, \quad \aleph_1, \quad \aleph_2, \quad \ldots$$

Aquí $\aleph$ (álef) es la primera letra hebrea, que fue la que eligió Cantor para nombrar sus cardinales infinitos.[4]

Cantor demostró que a cada conjunto X, finito o infinito, se le puede asociar un cardinal (un número de elementos), que con la notación moderna se representa por $|X|$, de modo que dos conjuntos (finitos o infinitos) tienen el mismo cardinal si y sólo si se puede establecer una aplicación biyectiva entre ellos, es decir, si se pueden emparejar sus elementos sin que falte ni sobre ninguno. Si X es finito, su cardinal es un número natural y es lo que normalmente se entiende por su número de elementos, mientras que si X es infinito entonces $|X|$ es uno de los cardinales transfinitos.

[4]En realidad la sucesión de los cardinales infinitos es mucho más compleja de lo que la "lista" precedente da a entender. Por ejemplo, por encima de $\aleph_0, \aleph_1, \aleph_2, \ldots$ se encuentran los cardinales transfinitos llamados $\aleph_\omega, \aleph_{\omega+1}, \aleph_{\omega+2}, \ldots$, pero luego de todos ellos viene $\aleph_{\omega\cdot 2}$, y es esta complejidad creciente de la sucesión, que rebasa todo posible límite, la que hace que no se la califique de "sucesión infinita", sino de "sucesión transfinita".

En estos términos, el teorema de Cantor, del que hemos hablado antes, afirma que todo conjunto X cumple la relación $|X| < |\mathcal{P}X|$, pues existe una aplicación inyectiva $X \longrightarrow \mathcal{P}X$, lo que se traduce en que $|X| \leq |\mathcal{P}X|$, pero es imposible establecer una biyección entre ambos, lo que nos da la desigualdad estricta.

El cardinal $\aleph_0$ es el cardinal del conjunto de los números naturales, y es el menor cardinal infinito. Los conjuntos de cardinal $\aleph_0$ son los conjuntos que se pueden poner en correspondencia biunívoca con los números naturales, y por ello se llaman conjuntos numerables (son conjuntos que se pueden "numerar" en el sentido de que se le puede asignar un número natural a cada uno de sus elementos, aunque para ello necesitemos usar todos los números naturales).

Si consideramos los conjuntos numéricos $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$ que contienen, respectivamente, a los números naturales, los enteros, los racionales, los reales y los complejos, Cantor demostró que $|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}| = \aleph_0$, de modo que los tres tienen el mismo número de elementos, aunque cada uno esté estrictamente contenido en el siguiente (con conjuntos infinitos, el todo no es necesariamente mayor que la parte). Por el contrario, Cantor demostró que $|\mathbb{R}| = |\mathbb{C}| = |\mathcal{P}\mathbb{N}|$, y aquí es donde interviene el teorema de Cantor para concluir que estos conjuntos no son numerables. Más precisamente, Cantor definió una aritmética transfinita y demostró que $|\mathcal{P}X| = 2^{|X|}$ (compárese con el caso finito: un conjunto de 3 elementos tiene exactamente $2^3 = 8$ subconjuntos). En resumen:

$$|\mathbb{R}| = |\mathbb{C}| = |\mathcal{P}\mathbb{N}| = 2^{\aleph_0} > \aleph_0.$$

Por otro lado, el cardinal $\aleph_1$ se define como el menor cardinal no numerable. Todo conjunto no numerable tiene al menos $\aleph_1$ elementos, por lo que

$$|\mathbb{R}| = |\mathbb{C}| = |\mathcal{P}\mathbb{N}| = 2^{\aleph_0} \geq \aleph_1.$$

Pero, ¿cuánto vale exactamente $2^{\aleph_0}$? ¿Cuántos números reales, o cuántos conjuntos de números naturales hay? Cantor conjeturó que $2^{\aleph_0} = \aleph_1$ o, lo que es lo mismo, que el cardinal de $\mathcal{P}\mathbb{N}$ es el siguiente cardinal después de $\aleph_0$, de modo que no hay cardinales intermedios. A esta conjetura la llamó *hipótesis del continuo,* porque Cantor llamaba a $\mathbb{R}$ "el continuo". Durante el resto de su vida, Cantor trató de demostrar o refutar la hipótesis del continuo, y encontró muchos argumentos en favor y en contra, pero ninguno era concluyente. En 1900, Hilbert había incluido la hipótesis del continuo como el primero en una lista de los problemas más importantes que, a su juicio, tenía planteada la matemática del siglo XX.

Pero en 1940 Gödel publicó un nuevo artículo revolucionario, en el que demostró que, si la teoría de conjuntos es consistente, también lo es añadir como axioma la hipótesis del continuo o, dicho de otro modo, que es imposible demostrar que la hipótesis del continuo es falsa, y en 1966 Paul Cohen demostró que tampoco es posible demostrar que es verdadera. Más precisamente, Cohen demostró que $2^{\aleph_0}$ puede ser $\aleph_1$, como afirma la hipótesis del continuo, pero también $\aleph_2$, o $\aleph_{17}$, o $\aleph_{\omega^3+13}$, o casi cualquier cardinal no numerable. En resumen: la hipótesis del continuo es un ejemplo de afirmación indecidible en teoría de conjuntos. No existe ningún argumento matemático que permita demostrarla o refutarla.

A partir de ese momento, los especialistas en lógica y en teoría de conjuntos empezaron a encontrar una larga lista de afirmaciones para las que sucedía lo mismo, y no todas eran problemas técnicos de la teoría de conjuntos, sino que hay afirmaciones indecidibles en álgebra, en el análisis matemático, en topología, etc. Éste es el motivo principal por el que la lógica matemática es esencial para entender la matemática: porque hay problemas matemáticos que sólo pueden ser resueltos mediante una prueba de independencia, es decir, una prueba de que es imposible resolverlos, por lo que no tiene sentido seguir buscando una respuesta en un sentido u otro, y ahí la lógica resulta esencial.

Los matemáticos no necesitan estudiar lógica formal para razonar correctamente. Al contrario, es la lógica formal la que se plantea y resuelve el problema de dar una definición de razonamiento formal que se corresponda exactamente con la forma de razonar de los matemáticos. Sin embargo, para demostrar que una afirmación como la hipótesis del continuo no es ni demostrable ni refutable a partir de los axiomas de ZFC, no basta con tomar los axiomas y no ser capaz de concluir nada. La forma de probar la independencia de la hipótesis del continuo es partir de que si fuera demostrable o refutable existiría una demostración de la hipótesis del continuo o de su negación en el sentido preciso que define la lógica matemática, para a continuación razonar en términos de dicha definición que tal demostración no puede existir. En otras palabras: no necesitamos saber qué entendemos exactamente por demostración para demostrar algo, pero es imprescindible tener una definición precisa de demostración y justificar que comprende todas las posibilidades de razonamiento matemático para concluir que no existe tal cosa para una afirmación en concreto (como la hipótesis del continuo o su negación) y de ahí concluir que, por mucho que aguce el ingenio, un matemático no tiene posibilidad alguna de demostrar la afirmación.

Así pues, la lógica matemática no sólo sirve para precisar en qué consiste la fundamentación última de las matemáticas, sino que también es una herramienta indispensable para obtener pruebas de consistencia.

El formalismo Aunque relativamente pocos matemáticos siguieron de cerca la crisis de fundamentos y la lógica matemática o las teorías axiomáticas que surgieron de ella, lo cierto es que este proceso difundió el formalismo entre la comunidad matemática —especialmente en la enseñanza universitaria— introduciendo una noción muy precisa de lo que hay que entender por rigor matemático (o más bien reforzó una línea de pensamiento formalista preexistente):

> Todo razonamiento matemático riguroso debe partir de axiomas explícitos de los que se extraen consecuencias según los principios de la lógica formal. Todos los conceptos involucrados tienen que estar determinados, o bien por dichos axiomas, o bien por definiciones que sólo aludan a otros conceptos previamente determinados (por los axiomas o por definiciones previas). En particular, no son admisibles definiciones o razonamientos basados en ideas o principios "intuitivos". Cualquier concepto o argumento "intuitivo" que quiera usarse en una demostración debe formalizarse debidamente. Cualquier definición o razonamiento "intuitivo" no formalizado es sospechoso de

> error e incluso de contradicción y, por lo tanto, es inadmisible. Es fácil poner una infinidad de ejemplos de cómo afirmaciones "intuitivamente" verdaderas resultan ser falsas a la luz de la matemática formal.

Un ejemplo de afirmación demostrable formalmente, aunque "contraria a la intuición" es la llamada *paradoja de Banach-Tarski* (aunque aquí la palabra "paradoja" sólo hay que entenderla en el sentido débil de "hecho contrario a lo que uno podía pensar", pues se trata de un teorema de ZFC que no da lugar a ninguna contradicción):

> **La paradoja de Banach-Tarski** *Es posible descomponer adecuadamente una esfera maciza en cinco partes de modo que si éstas se reordenan como las piezas de un puzzle, puedan encajarse de nuevo para formar dos esferas macizas del mismo tamaño que la original.*

Cualquiera que aceptara como "intuitivamente evidente" que esto es falso, se estaría equivocando, pues, como ya hemos señalado, la paradoja de Banach-Tarski es un teorema de ZFC. Por ello mismo, no podemos permitirnos el lujo de dar por cierta cualquier cosa que parezca cierta sólo porque pueda parecer inconcebible que sea falsa. Si realmente es cierta, tendrá que poder demostrarse formalmente a partir de los axiomas de ZFC.

Por ejemplo, si alguien pretende definir una aplicación $f : A \longrightarrow B$ como un criterio que a cada elemento $a \in A$ le asigna otro elemento $f(a) \in B$, un matemático consecuente con la concepción formal de la matemática no lo considerará admisible y su réplica será: *¿a qué llamas "criterio"? Dame una definición formal, precisa y rigurosa de lo que llamas "criterio".*

La definición válida en el contexto de la matemática formal es la que establece que una aplicación $f : A \longrightarrow B$ es un conjunto f formado por pares ordenados (a, b), con $a \in A$ y $b \in B$, de modo que, para cada elemento $a \in A$, existe un único elemento $b \in B$ tal que $(a, b) \in f$, y a dicho elemento b se le llama imagen de a por f, y se representa como $b = f(a)$.

De este modo hemos sustituido la palabra "criterio", sospechosa de ser un mero "concepto intuitivo", por la palabra "conjunto", que no necesitamos definir por que los axiomas nos eximen de hacerlo.

Así, muchos matemáticos asimilaron, desde el mismo inicio de su formación como tales, una serie de normas y criterios sobre lo que es riguroso y lo que no lo es, sobre lo que está bien y lo que está mal, sobre lo que es aceptable y lo que no lo es, que son incuestionablemente correctos cuando se aplican al razonamiento formal, pero que se convierten en una aberración si se pretende aplicarlos a cualquier forma de razonamiento matemático.

En efecto, la postura filosófica según la cual los únicos razonamientos matemáticos aceptables son los que respetan estrictamente las normas del razonamiento formal, es lo que podemos llamar *formalismo*. Desde este punto de vista, el proceso de construcción de las matemáticas "desde sus cimientos" tendría que seguir aproximadamente estos pasos:

1. Se fijan los axiomas de ZFC (u otra teoría similar que se considere oportuna), lo que supone aceptar sin definición los conceptos primitivos de "conjunto" y "pertenencia".

2. A partir de ellos se definen los conceptos conjuntistas básicos y se demuestran sus propiedades (unión e intersección de conjuntos, relaciones, funciones, etc.).

3. Se construye el sistema numérico y se demuestran sus propiedades (números naturales, enteros, racionales, reales y complejos).

4. Se definen los conceptos de las distintas ramas de la matemática y se demuestran sus propiedades.

En particular, todos los conceptos matemáticos distintos de "conjunto" y "pertenencia" deben ser definidos formalmente a partir de estos conceptos primitivos o de otros previamente definidos, sin dar cabida a ninguna forma de "definición intuitiva". Cualquier otra forma de proceder no puede considerarse matemáticamente rigurosa ni, por tanto, aceptable. A lo sumo, podrá tener un valor didáctico aceptable siempre y cuando esté probado que es posible llegar a los mismos resultados por procedimientos rigurosos (= estrictamente formales).

El problema es que este "ideario formalista" incapacita a quien lo adopta para entender la fundamentación de la matemática. En efecto, un formalista no es consciente de que lo que toma como punto de partida —fijar los axiomas de ZFC y, tácitamente, también las reglas válidas de razonamiento formal— no es algo que pueda hacerse arbitrariamente de un plumazo. Por el contrario, definir ZFC —o cualquier otra teoría axiomática— con el rigor necesario para que la definición sea operativa, no es tan simple como decir "consideramos estos axiomas":

- En primer lugar es necesario definir un lenguaje formal, es decir, un lenguaje con una sintaxis perfectamente definida que no deje cabida para las ambigüedades que permiten los lenguajes naturales. Esto supone en particular distinguir con rigor cadenas de signos como $+ = xy \wedge \neg$ de otras como $x \neq 0 \to \neg(x + x = x)$, es decir, cadenas asintácticas, sin sentido alguno, de cadenas acordes con la sintaxis del lenguaje formal, que son las únicas que se considerarán en la práctica.

- En segundo lugar es necesario definir y estudiar las propiedades de muchos conceptos sintácticos, como la operación que nos lleva de $x + y = y + x$ a $2 + 3 = 3 + 2$ (es decir, la sustitución de variables por otros términos del lenguaje).

- En tercer lugar es necesario definir qué significa que una afirmación del lenguaje formal sea consecuencia lógica de unas premisas dadas, lo que supone la existencia de una deducción lógica, definida a su vez como una sucesión de afirmaciones que cumpla ciertas condiciones para que la deducción se pueda considerar correcta.

- Pero, más aún, aunque tengamos ya una definición de qué es un razonamiento formal que a un matemático le parezca sensata, tenemos que demostrar que dicha definición incluye realmente toda la capacidad de razonamiento matemático, es decir, que cualquier cosa que un matemático acepte como demostrada admite una demostración en el sentido que hemos definido. De lo contrario, cuando se demuestra que la hipótesis del continuo no se puede demostrar (en el sentido de que no existe una demostración formal de la misma), quedaría la duda de si no podría demostrarse con un argumento no contemplado en nuestra definición de demostración formal, pero que los matemáticos consideren concluyente de todos modos. No es evidente en absoluto cómo puede probarse que una propuesta de definición de razonamiento formal abarca realmente todos los razonamientos que un matemático daría por válidos, pero esto es uno más de los teoremas revolucionarios que probó Gödel, conocido como *teorema de completitud semántica de Gödel.*

Metamatemática Toda la teoría esbozada en los puntos precedentes se conoce como *metamatemática.* La metamatemática es el estudio lógico de las teorías axiomáticas formales necesario para definirlas y estudiar su alcance, lo que incluye resultados positivos, como el teorema de completitud semántica, que asegura que el cálculo deductivo formal captura todas las formas de razonamiento aceptables para un matemático, y también resultados negativos, como los teoremas de incompletitud.

El problema con que tropieza el formalista es que la metamatemática es una teoría sofisticada que requiere tratar con muchos conceptos matemáticos: números naturales, sucesiones, conjuntos, etc. Por ejemplo, necesitamos hablar de los signos de un lenguaje formal, de sucesiones finitas de signos, que forman afirmaciones, de conjuntos de axiomas, que pueden ser infinitos (de hecho, ZFC tiene infinitos axiomas), etc. A menudo es necesario razonar por inducción sobre el número de signos de una afirmación, o sobre el número de afirmaciones de una demostración, etc.

Todo esto resulta desconcertante —o más bien indigerible— al formalista, que pretende que los números naturales sean los objetos construidos rigurosamente en en marco de ZFC, tal y como le enseñaron en sus estudios, y cuya confianza en el principio de inducción se basa en que éste es demostrable en ZFC, y así, al tomar cualquier libro sobre metamatemática, se encuentra con que le hablan de números naturales, de inducción, hasta de conjuntos infinitos, y todo ello antes de haber definido ZFC o cualquier otra teoría axiomática que permita definir, o tratar, estos conceptos *comme il faut.* Y si el formalista decide darle una oportunidad a estos libros y se esfuerza por ver si lo que exponen se puede considerar fiable y riguroso en algún sentido, se encontrará, para su espanto, con que no, con que muchos conceptos matemáticos son tratados con absoluto desprecio a los criterios de rigor que él considera indispensables. Su impresión será que tales libros no son fiables, que sus afirmaciones son cuestionables, ininteligibles, y hasta podría pensar que el autor trata de ocultar maliciosamente el fondo del asunto y que cae cínicamente en un círculo vicioso al usar subrepticiamente ZFC en el proceso con el que pretende construir ZFC.

En otros términos, el problema con el que se encuentra el formalista es que se niega a admitir algo que, bien mirado, es obvio: los razonamientos metamatemáticos no pueden ser razonamientos formales, puesto que el objeto de la metamatemática es determinar qué hay que entender exactamente por razonamiento formal. Si ya tuviéramos a nuestra disposición una teoría axiomática formal para presentar en ella la metamatemática, significaría que ya tendríamos hecho el trabajo que pretendemos hacer.

Por consiguiente, el razonamiento metamatemático, por su propia finalidad, tiene que ser necesariamente *informal,* donde "informal" no hay que entenderlo en el sentido en que lo entiende un formalista, para quien es una forma fina de decir "chapucero" o "carente de rigor", sino que "razonamiento informal" significa simplemente lo contrario de "razonamiento formal": razonamiento en el que el criterio para aceptar una conclusión como válida no es que se obtenga de las premisas aplicando unas reglas de inferencia formales prefijadas, sino atendiendo al significado de las afirmaciones involucradas y dando argumentos concluyentes que justifiquen que son verdaderas. Hay una frase de Raymond Smullyan que resume agudamente la postura formalista:

> *Un formalista es un matemático que es incapaz de entender algo a menos que carezca de significado.*

Podríamos decir que el formalismo es un caso de "deformación profesional". Ciertamente, los criterios de rigor del matemático formalista son los que debe aplicar en el ejercicio de su profesión, que consiste en demostrar teoremas en ZFC ajustándose a las exigencias de su cálculo deductivo formal. Pero pretender que el razonamiento formal es la única forma de razonamiento fiable es considerar como "lo normal" lo que a todas luces es "lo excepcional", pues lo que realmente debería causar sorpresa y requiere justificación es que es posible razonar sin tener en cuenta sobre qué se razona —y en este libro nos ocuparemos de justificarlo—, mientras que el hecho de que es posible razonar tomando como guía el significado de las afirmaciones que se formulan para asegurarse de que cada cosa que se dice es verdad no debería considerarse algo problemático.

Intuición y razonamiento informal En este punto el formalista podría plantearnos esta pregunta maliciosa: *Y si es posible hacer razonamientos matemáticos informales, ¿para qué necesitamos la matemática formal?*

La respuesta es que sólo es posible razonar informalmente de forma fiable cuando tratamos con una parte restringida de los conceptos matemáticos, que, no obstante, es suficiente para fundamentar la metamatemática que a su vez sirve de fundamento para la matemática formal, que abarca todo el razonamiento matemático.

Más precisamente, podemos razonar informalmente sobre aquellos conceptos matemáticos que tienen un contenido intuitivo claro, y aquí hemos empleado una palabra maldita a los ojos de un formalista: la *intuición,* pero nos apresuramos a aclarar que no vamos a usar este término en el sentido en que lo usan los matemáticos que desprecian la intuición. No estamos hablando de "presentimientos", de "ocurrencias infundadas", "perspicacia", ni nada parecido. Nada de eso sería fiable como fundamento de un razonamiento matemático riguroso.

Llamamos "intuición" a nuestra capacidad de formarnos imágenes espaciales y temporales (como la imagen espacial de un cubo), y especialmente —para lo que nos interesa— aquellas que no incluyen ningún elemento de la realidad física. Veamos un ejemplo concreto:

Si alguien desconoce el significado de la palabra "cubo" y le decimos que un cubo es la forma geométrica de un dado, y le mostramos la imagen de la derecha para despejar cualquier ambigüedad (precisando que todos los ángulos son rectos y que todas las aristas tienen la misma longitud), con esto nuestro interlocutor habrá entendido perfectamente lo que es un cubo.

Así, si le preguntamos cuántos vértices tiene un cubo, aunque jamás se lo hubiera planteado antes, no tendrá dificultad en observar que, pensando en un cubo en la posición que muestra la figura, tiene cuatro vértices en un plano inferior y otros cuatro en un plano superior, lo que hace un total de ocho vértices. Y si le preguntamos cuántas aristas tiene, observará que tiene cuatro horizontales en el plano inferior, otras cuatro en el superior y otras cuatro verticales, lo que hace un total de doce aristas. Notemos que para llegar a estas respuestas no es necesario que el interrogado tenga ante sí ningún objeto físico con forma de cubo. Ni siquiera sería necesaria ninguna imagen si ya ha entendido plenamente qué es un cubo y puede formarse la imagen de uno en su mente.

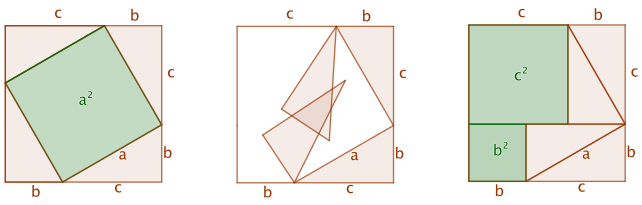
Definir el concepto de "cubo" mostrando una imagen de un cubo (que ni siquiera hace falta que sea una imagen realista, sino que valdría igualmente cualquier boceto mal hecho en un papel, con tal de que expliquemos a nuestro interlocutor cuál es la imagen ideal que debe abstraer) no tiene nada que ver con lo que se entiende por una definición formal en el seno de una teoría axiomática. Sin embargo, cualquiera que entienda esta definición puede responder objetivamente muchas preguntas sobre un cubo, como el número de vértices, caras o aristas que tiene un cubo, o incluso preguntas más sofisticadas, como si se puede cubrir el espacio yuxtaponiendo cubos, sin que queden espacios vacíos entre ellos (cosa que no se puede hacer, por ejemplo, con esferas). Y a todas estas respuestas se llega sin necesidad de ningún razonamiento que parta de ningún axioma. El concepto de "cubo" es un concepto intuitivo. No es un objeto físico. Cuando nos imaginamos un cubo no estamos imaginando —o no necesitamos imaginar— un cubo de madera, o de plástico, ni necesitamos atribuirle un peso, ni lo estamos situando en ningún punto en particular de universo, ni hay riesgo de que alguien nos lo pise sin querer, porque no está en ninguna parte. Es sólo una construcción mental, pero que nos permite razonar informalmente exactamente igual que lo hacemos habitualmente sobre objetos físicos, como cuando, por ejemplo, vemos varios objetos físicos y nos planteamos cómo podemos disponerlos para que quepan en un armario o en un maletero. El hecho de que una imagen de un cubo no sea un objeto físico no dificulta en absoluto nuestras posibilidades de razonamiento sobre los cubos. Al contrario, las facilita, porque un objeto físico puede tener características que nos sean desconocidas, mientras que un cubo no es ni más ni menos que lo que nos imaginamos al imaginar un cubo.

Aprovechamos este ejemplo para señalar un hecho fundamental: no es lo mismo intuir que pensar. Por ejemplo, podemos pensar en un cubo de cuatro dimensiones, y razonar formalmente sobre él, pero no podemos intuirlo. No tenemos ninguna representación intuitiva de un cubo de cuatro dimensiones.[5] Un cubo de cuatro dimensiones tiene grupos de cuatro aristas perpendiculares dos a dos, y eso no nos lo podemos imaginar, porque el espacio de nuestra intuición es el espacio tridimensional euclídeo, y no está en nuestra mano cambiar eso. Establecer esta diferencia es fundamental. Por ejemplo, podemos pensar en un poliedro regular de 10 caras heptagonales, y podemos razonar sobre él. Por ejemplo, podemos razonar que tendría que tener 35 aristas ($10 \times 7 = 70$, pero así contamos dos veces cada arista), luego tendría que tener 27 vértices (por el teorema de Euler sobre poliedros), pero si seguimos analizando las características de este poliedro terminaremos llegando a una contradicción, porque no existen poliedros regulares de 10 caras heptagonales. Pensar en algo no nos da ninguna garantía de que lo que estamos pensando no esconda contradicciones. En cambio, el hecho de que podamos imaginarnos un poliedro regular de 6 caras cuadradas nos da la garantía de que es imposible que el concepto de "cubo" (tridimensional) sea contradictorio. Si los cubos fueran contradictorios, podríamos pensar en ellos, pero no intuirlos.

Similarmente, si contamos mentalmente: "uno, dos, tres, cuatro", y luego avanzamos dos lugares más, pensando "cinco, seis" (con la conciencia de que hemos dado exactamente dos pasos más), habremos constatado que $4 + 2 = 6$ sin que medie ningún razonamiento formal, sin haber deducido nada de ningún axioma. Del mismo modo que a partir de la imagen de un cubo hemos obtenido inmediatamente información matemática (geométrica) sobre los cubos, a partir de una intuición puramente temporal hemos obtenido una afirmación aritmética que podemos dar por cierta fuera de toda duda razonable.

Afirmaciones como "los cubos tienen 12 aristas" o "$4 + 2 = 6$" son ejemplos de afirmaciones intuitivamente evidentes, pero también hay afirmaciones cuya verdad puede razonarse intuitivamente sin que se puedan calificar de evidentes.

Por ejemplo, nadie puede imaginarse un triángulo rectángulo arbitrario y afirmar: obviamente, el cuadrado de la hipotenusa es igual a la suma de los cuadrados de los catetos. Sin embargo, si le mostramos a alguien las figuras siguientes:

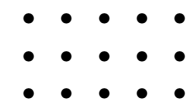


[5] En realidad, podemos intuir una imagen tridimensional en perspectiva de un cubo de cuatro dimensiones, la cual basta para obtener intuitivamente (informalmente) algunas conclusiones sobre los cubos de cuatro dimensiones.

y le explicamos que movemos los triángulos de la primera figura como muestra la segunda hasta situarlos en la posición de la tercera, entonces vemos que el área del cuadrado de la primera figura es la suma de las áreas de los cuadrados de la tercera, lo que demuestra el teorema de Pitágoras (supuesto que ya conozcamos algunos resultados elementales sobre áreas). Lo que sí que podemos dar por evidente es que el razonamiento previo se puede hacer con cualquier triángulo rectángulo dado, sin que importen sus dimensiones. Con ello estamos afirmando que, dado cualquier triángulo rectángulo, siempre seremos capaces de construir mentalmente figuras como las anteriores, y esto nos asegura que el teorema de Pitágoras vale para triángulos rectángulos arbitrarios y no sólo para los que muestre una figura en particular.

Así pues, el teorema de Pitágoras no es intuitivamente evidente, pero es fácil demostrarlo mediante un argumento intuitivo que se basa en que podemos imaginarnos unos triángulos ideales que se mueven de una posición a otra, junto con la conciencia de que esto puede hacerse independientemente de las dimensiones de los triángulos.

Similarmente, basta observar una figura como



para convencernos de que $3 \cdot 5 = 5 \cdot 3$ y, si además nos persuadimos de que podemos formar rectángulos de puntos análogos con cualquier número de puntos en sus lados, nos convencemos de que $mn = nm$ para cualquier par de números naturales m y n (tratando aparte el caso trivial en que alguno de los factores sea 0). Esto es otro ejemplo de razonamiento informal basado en la intuición.

Si el lector es formalista, es probable que ya lleve un rato deseando interpelar al autor para presentarle muchos ejemplos de "razonamientos intuitivos" que llevan a conclusiones falsas y que ponen en evidencia lo poco fiable que es la intuición, a lo que el autor respondería que hay dos clases muy distintas de tales ejemplos:

Por una parte tenemos casos como el de un argumento basado en la figura —por ejemplo— de un triángulo acutángulo por el que alguien concluya erróneamente que una propiedad es cierta para todos los triángulos, cuando bastaría dibujar un triángulo obtusángulo o rectángulo para comprobar que en ese caso no se cumple. Ese caso es totalmente análogo a los errores de razonamiento formal en los que alguien aplica un teorema sin advertir que en el caso considerado no se cumple una de sus hipótesis, y por ello llega a una conclusión errónea, por lo que, si es en esta clase de ejemplos en los que está pensando el lector, lo que sucede es que es muy fácil ver la paja en el ojo ajeno. Es igual de fácil cometer un error razonando formalmente o razonando informalmente, y lo importante es que, cuando alguien se da cuenta y nos señala el error, no se produce un debate sobre si la prueba está bien o está mal, sino que el error queda en evidencia y se llega al acuerdo de que la prueba era errónea (o de que tenía un alcance menor que el pretendido).

No obstante, hay muchos casos en los que la intuición nos informa de su grado de generalidad. Ya hemos señalado dos ejemplos: es intuitivamente inmediato que la demostración geométrica del teorema de Pitágoras que hemos considerado más arriba es aplicable a cualquier triángulo rectángulo, y no sólo a un ejemplo concreto de una figura concreta. Igualmente, es intuitivamente evidente que la disposición rectangular de puntos que nos permite concluir que $3\cdot 5 = 5\cdot 3$ puede modificarse para admitir cualquier número de puntos en cada lado, y así justifica que $mn = nm$. Por supuesto, alguien puede en un momento dado generalizar incorrectamente a partir de una imagen intuitiva, exactamente igual que alguien puede aplicar incorrectamente un teorema formal al no comprobar una de sus hipótesis.

Por otra parte, tenemos casos de naturaleza muy distinta, como el hecho de que alguien podría afirmar "intuitivamente" que no existen funciones continuas no derivables en ningún punto, porque es imposible imaginarse tal cosa, mientras que formalmente se puede demostrar que sí que existen.

En efecto, es imposible imaginarse intuitivamente una función continua no derivable en ningún punto, pero lo que deducimos de ahí es que el concepto formal de "función real de variable real" que se maneja en el análisis matemático es mucho más general que nuestro concepto intuitivo de "curva que no pasa dos veces por la misma recta vertical". Toda curva intuitiva con esta condición es la gráfica de una función en el sentido analítico formal, pero el recíproco no es cierto. Y por ello no podemos pretender que cualquier afirmación sobre funciones que parezca plausible en términos intuitivos tenga que corresponderse con un teorema del cálculo diferencial. No es admisible identificar un concepto intuitivo con un concepto formal por el mero hecho de que algunas instancias concretas del concepto formal tengan una interpretación intuitiva. De hecho, la razón principal por la que muchos matemáticos desprecian los razonamientos intuitivos es porque no tienen en cuenta que a menudo los conceptos abstractos que manejan trascienden con mucho los ejemplos intuitivos que los motivan, lo que invalida todo intento de concluir nada sobre ellos sin más base que la intuición. Así, los matemáticos hacen bien en despreciar la intuición en esos contextos, pero deberían tener presente que hay otros contextos en los que los conceptos involucrados pueden interpretarse fielmente en términos intuitivos, y en esos casos todas las presuntas evidencias en contra de la fiabilidad de la intuición no son aplicables.

Similarmente, el hecho de que no podamos formarnos una imagen intuitiva de cinco fragmentos de esfera que cumplan la paradoja de Banach-Tarski no es motivo para negar su validez, pues nuestra intuición de lo que es un "fragmento de espacio tridimensional" dista mucho del concepto formal de "subconjunto de $\mathbb{R}^3$", que es el que se considera en la demostración de la paradoja.

Así pues, para entender la metamatemática y, con ella, la fundamentación de la matemática formal, es necesario que el lector acabe convenciéndose —no necesariamente en este punto, sino tal vez teniendo en consideración el resto de esta introducción o, incluso tras un análisis crítico del contenido de este libro— de los hechos siguientes:

1. El razonamiento matemático informal que requiere la metamatemática es posible en la medida en que se ciña a afirmaciones a las que podamos atribuir un contenido intuitivo preciso que nos permitan evaluar en todo momento si son verdaderas o falsas.

2. Los límites de la intuición son difusos: por una parte tenemos conceptos con un contenido intuitivo inequívoco (como los triángulos), luego hay conceptos formales con un contenido intuitivo parcial, que no cubren toda su generalidad (como el de función real de variable real), de modo que sería falaz extraer consecuencias "intuitivas" del concepto general, y también conceptos en los que podemos pensar, pero a los que no podemos atribuir ningún significado intuitivo, como un espacio euclídeo de cinco dimensiones.[6]

3. No debemos confundir tener una representación intuitiva de un concepto o de un hecho con pensar en ello. Pensar en algo no nos da ninguna garantía de que lo que pensamos no sea contradictorio, intuir algo sí nos las da. El razonamiento formal es imprescindible a la hora de tratar con afirmaciones matemáticas carentes de contenido intuitivo o, al menos, de las que sólo tenemos una representación intuitiva parcial. Esto se aplica, en particular, a toda la teoría de conjuntos cantoriana, en la que surgieron las contradicciones que provocaron la crisis de fundamentos.

Seguidamente comentaremos algunas características propias del razonamiento matemático informal que lo distinguen del razonamiento formal, pero antes conviene observar que el razonamiento matemático informal (y riguroso) es algo cotidiano en muchos contextos. Por ejemplo, cuando un aficionado al ajedrez se plantea un problema del tipo "juegan blancas y dan mate en tres jugadas", está planteándose un problema de naturaleza matemática. Es posible que los matemáticos profesionales lo consideren de escaso valor, pero, si es así, lo será por juicios basados en una u otra escala de valores, pero no porque la naturaleza del problema sea distinta a la de otros problemas matemáticos mejor considerados. El ajedrecista puede abordar el problema con la ayuda de un tablero de ajedrez físico, a pesar de que la naturaleza física de las piezas es irrelevante, o puede trabajar únicamente con esquemas sobre el papel, o con la ayuda de un ordenador, pero el tablero y las piezas de ajedrez, así como las reglas del juego, son conceptos intuitivos, igual que los números naturales o los grafos finitos. Y

[6] Imaginemos que un topólogo está muy familiarizado con un espacio topológico carente de todo contenido intuitivo, como pueda ser $\beta\mathbb{N}$, la compactificación de Stone-Čech de $\mathbb{N}$, hasta el punto de que puede formular fácilmente conjeturas sobre $\beta\mathbb{N}$ que luego pueden probarse formalmente. Quizá un matemático describiría la situación diciendo que el topólogo es capaz de razonar intuitivamente sobre $\beta\mathbb{N}$, pero no es así si entendemos "intuitivamente" en el sentido que le estamos dando a la palabra. El topólogo no tiene ninguna imagen mental fidedigna de $\beta\mathbb{N}$ (no existe tal cosa). Puede orientarse con esquemas, analogías, usando su experiencia sobre espacios similares, etc., lo cual capacita a su cerebro para presentarle inconscientemente conjeturas acertadas, pero eso no tiene nada que ver con deducir hechos a partir de representaciones intuitivas que se correspondan fielmente con el objeto considerado, no que aporten meras analogías útiles.

nadie echa de menos una teoría axiomática para resolver un problema de ajedrez, ni nadie cuestionará que la solución sea correcta, si está bien razonada, sólo porque no se haya hecho referencia en ella a ningún axioma ni a ninguna regla de inferencia de la lógica formal. De lo que el lector necesita convencerse es de que todo el contenido metamatemático de este libro tiene la misma solidez que una solución bien razonada a un problema típico de ajedrez.

Razonamiento intuitivo vs. razonamiento formal Algunos formalistas que acaban persuadidos a regañadientes de que es inevitable aceptar que el razonamiento metamatemático tiene que ser informal y basado en la intuición, tienden a escribirse sus propias notas procurando que su "razonamiento informal" sea prácticamente indistinguible del razonamiento formal al que están acostumbrados. Eso les lleva a aceptar irreflexivamente afirmaciones que, tal vez sean aceptables, pero que requerirían una cierta atención y, en cambio, cuestionen otras que no son problemáticas en absoluto.

La situación podría compararse a la de alguien que haya vivido toda su vida en una gran ciudad y, de repente, un día se encuentra en medio de una selva y, sin ser consciente de las consecuencias que conlleva este cambio de entorno, espera seguir actuando como siempre lo ha hecho, y así evita tirar una piel de plátano al suelo por si un policía lo ve y le pone una multa (es decir, viendo peligros donde no los hay), y no teme comer cualquier cosa aparentemente comestible que se encuentre dando por hecho que, si está ahí, tendrá su correspondiente certificado de sanidad (no viendo peligros donde los hay).

El primer hecho que pilla desprevenido al formalista reconvertido es que no repara en que en el razonamiento formal todos los conceptos reciben el mismo trato, mientras que al razonar informalmente es necesario distinguir tres clases de conceptos:

1. *Los conceptos plenamente determinados intuitivamente, en el sentido de que sabemos qué significa exactamente que todos los objetos determinados por dicho concepto cumplan una determinada propiedad.*

   Por ejemplo sabemos perfectamente a qué nos referimos cuando hablamos de los números naturales, o de los números pares, o de los números primos, o de los pares de números naturales, etc. Más concretamente, consideremos la afirmación "todo número par mayor que 2 es suma de dos números primos". No sabemos si esa afirmación es cierta o no (es lo que se conoce como la *conjetura de Goldbach*), pero *sabemos exactamente lo que significa,* sabemos lo que quiere decir que todos los números pares mayores que 2 sean suma de dos primos. Podemos programar un ordenador para que vaya buscando descomposiciones de números pares como suma de dos primos. Para cada número par, la existencia de tal descomposición se puede comprobar (o refutar) en un tiempo finito, así que el ordenador puede ir imprimiendo una lista del estilo de:

   $$2 \text{ no}, \quad 4 \text{ sí}, \quad 6 \text{ sí}, \quad 8 \text{ sí}, \quad 10 \text{ sí}, \quad \ldots$$

   No sabemos si, dejando calcular al ordenador durante un tiempo suficiente, terminará encontrando otro "no", aparte del que ha encontrado en

el 2, pero sabemos lo que esto significa, y la conjetura de Goldbach afirma precisamente que eso nunca sucederá. Si realmente nunca sucederá, la conjetura es cierta, y si sucederá en algún momento, es falsa. No sabemos cuál es el caso, pero sabemos lo que estamos diciendo cuando nos planteamos si es cierta o si es falsa.

2. *Los conceptos que podemos identificar como bien definidos en casos concretos, pero de los que no tenemos una representación intuitiva de la totalidad de sus casos particulares.*

   Entre ellos tenemos, por ejemplo, el concepto de "definición". En efecto, es fácil reconocer que ciertas expresiones definen unívocamente números naturales, como:

   > *el menor número natural expresable como suma de dos cubos de dos formas distintas.*

   Puede comprobarse que esta definición determina, concretamente, el número 1 729, y no necesitamos ninguna teoría axiomática para justificar que es así. Sólo tenemos que hacer unos cuantos cálculos que un ordenador puede hacer por nosotros en poco tiempo. Por otra parte, también es claro que "el número natural que da buena suerte" no es una definición válida. Sin embargo, si un formalista nos pregunta cuál es la definición informal de "definición", es decir, qué requisitos pedimos a una presunta definición en virtud de los cuales admitimos como tal al primer ejemplo que hemos dado y no al segundo, tenemos que responderle que no existe tal definición de "definición".

   En efecto, supongamos que pudiéramos distinguir en general qué expresiones (en castellano, por ejemplo) definen un número natural y cuáles no. Entonces, podríamos limitarnos a considerar, por ejemplo, las expresiones castellanas con menos de setenta letras (que son un número finito) y seleccionar entre ellas todas las que definan un número natural. Con esto obtendríamos un número finito de números naturales, luego podríamos calcular:

   > *El menor número natural no definible en castellano con menos de setenta letras.*

   Si llamamos B a este número, resulta que la expresión anterior sería una definición de B, y así tendríamos una contradicción conocida como la *paradoja de Berry*.

   Esta contradicción surge de suponer que podemos distinguir *en general* las expresiones que definen números naturales y las que no definen ningún número natural, por lo que debemos concluir que esto no es posible, pero aquí es fundamental tener presente que el hecho de que no tengamos una definición general de "definición" no es óbice para que podamos reconocer fuera de toda duda que algunas expresiones son definiciones, como "el menor número primo" o "el máximo común divisor de 16 y 100".

3. *Conceptos con los que podemos razonar formalmente en el marco de una teoría axiomática adecuada, pero a los que no podemos atribuir un significado intuitivo preciso.*

   Un ejemplo en este caso es el concepto de "función continua de $\mathbb{R}$ en $\mathbb{R}$", del que ya hemos señalado que tiene algunas interpretaciones intuitivas concretas, pero también muchas otras sin contenido intuitivo (como las funciones continuas que no son derivables en ningún punto) y ello hace que cualquier razonamiento intuitivo sobre funciones continuas puede ser orientativo, pero nunca concluyente por sí mismo, y por ello los recelos de los matemáticos hacia los razonamientos intuitivos en estos contextos están justificados.

Así, los conceptos de tipo 1 son los que más fácilmente pueden ser tratados informalmente, los de tipo 2 pueden manejarse con precauciones y los de tipo 3 son inadmisibles en un contexto informal. Veamos algunos ejemplos más junto con algunas observaciones adicionales.

Otro concepto de tipo 1 es el de "par de números naturales". Podemos razonar intuitivamente sin restricciones con este concepto porque "la totalidad de los pares de números naturales" tiene un significado intuitivo preciso. Con esto no queremos decir que podamos verlos "todos a la vez" o algo así, sino simplemente que sabemos lo que queremos decir cuando decimos que todos los pares de números cumplen algo, o que existe un par de números que cumple algo (siempre y cuando "algo" sea una propiedad intuitivamente bien definida). Esto es así porque podemos enumerar todos los pares de números naturales:

$$(0,0), \quad (0,1), \quad (1,0), \quad (0,2), \quad (1,1), \quad (2,0), \quad \ldots$$

para ello enumeramos primero el único par cuyas componentes suman 0, luego ponemos los dos pares cuyas componentes suman 1, luego los tres pares cuyas componentes suman 2, y así sucesivamente. Afirmar que todos los pares cumplen algo es afirmar que lo cumple el primer par de la lista, y el segundo, y el tercero, y el cuarto, etc. Puede que no tengamos medios de saber si, efectivamente, ninguno fallará, pero sabemos lo que decimos cuando nos planteamos esa posibilidad.

Más en general, sólo admitiremos como "conceptos de tipo 1" los que determinen un conjunto numerable de objetos (finito o infinito). Acabamos de ver que los pares de números naturales están en este caso. Siempre que podemos enumerar unos objetos, sea de forma explícita, o bien podamos describir una forma de enumerarlos aunque en la práctica no podamos determinar cuál es el objeto n-simo, podremos darle sentido intuitivo a una afirmación sobre la totalidad de los objetos: decir que todos cumplen algo es decir que lo cumple el primero de la lista, y el segundo, y el tercero, etc., y decir que existe uno que cumple algo es decir que lo cumple el primero de la lista, o el segundo, o el tercero, etc., lo cual no supone que estemos en condiciones de determinar si se da el caso o no, pero sabemos lo que significa.

¿Significa esto que ningún conjunto no numerable tiene un significado intuitivo preciso? No vamos a afirmarlo categóricamente, pero en este libro nunca

trabajaremos informalmente con conjuntos no numerables. Si alguien considera que puede atribuir un significado objetivo a una afirmación sobre la totalidad de los elementos de un conjunto no numerable, tendrá que explicar en qué sentido es así, porque nosotros no sabríamos hacerlo y por ello nunca lo haremos.

Por ejemplo, el argumento del teorema de Cantor nos asegura que cualquier enumeración que hagamos de conjuntos de números naturales nunca estará completa. Si tenemos bien determinada una sucesión de conjuntos de números naturales

$$A_0, \quad A_1, \quad A_2, \quad \ldots$$

podemos construir un conjunto de números naturales que no está en la lista. Concretamente, el conjunto de los números naturales n tales que n no está en A_n. Por ello no somos capaces de atribuir un significado intuitivo a una afirmación sobre la totalidad de los conjuntos de números naturales. El concepto "conjunto de números naturales" es de tipo 2:

> Podemos trabajar intuitivamente con muchos conjuntos de números naturales, pero no tenemos una representación intuitiva de lo que es la totalidad de ellos.[7] Una cosa es que no veamos ningún impedimento para hablar de la totalidad de los conjuntos de números naturales, pero ahí "hablar de" sólo puede entenderse como "pensar en", no como "intuir". Podemos pensar en un conjunto de números naturales arbitrario, y en que podemos considerar otro distinto, y otro, etc., pero no tenemos forma de precisar qué queremos decir cuando decimos que todos los conjuntos de números naturales tienen una propiedad, o que existe un conjunto de números naturales que tiene una propiedad salvo si tenemos un argumento que lo justifique (en el primer caso) o podemos dar un ejemplo concreto (en el segundo).

Por ejemplo, consideremos esta afirmación:

> *Todo conjunto no vacío formado por números naturales tiene un mínimo elemento.*

Esto es una afirmación sobre la totalidad de los conjuntos de números naturales que podemos justificar como sigue: si A es un conjunto no vacío de números naturales, podemos ir recorriendo los números naturales y preguntarnos:

$$\text{¿0 está en } A?, \quad \text{¿1 está en } A?, \quad \text{¿2 está en } A? \quad \ldots$$

El hecho de que A sea no vacío equivale a que la respuesta tiene que ser afirmativa en algún caso (sin necesidad de que sepamos comprobar en la práctica

[7] Es importante no confundir esto con el hecho de que, por ejemplo, no podemos formarnos una imagen de todos los números naturales. Pero nunca hemos pretendido afirmar que alguien pueda decir "veo todos los números naturales". Lo que afirmamos es que sabemos lo que significa "todos los números naturales tienen tal propiedad", porque eso significa que la tiene el 0, y el 1, y el 2, etc., mientras que no tenemos una forma análoga de interpretar una afirmación del tipo "todos los conjuntos de números naturales tienen tal propiedad".

en qué casos es así). Y si la respuesta tiene que ser afirmativa en algún caso, el primer número n (en el sentido temporal) para el que sea cierto que n está en A, será el mínimo elemento de A.

Este argumento no supone que tengamos una imagen intuitiva de la totalidad de los conjuntos de números naturales. Es más bien un esquema de razonamiento que vemos que es aplicable a cualquier conjunto de números naturales que tengamos bien determinado intuitivamente. La diferencia es sutil, pero está ahí. Insistimos en que si consideramos una propiedad P que puede tener o no un conjunto de números naturales y no podemos demostrar que alguno la tiene o que, por el contrario, ninguno la tiene, no tenemos elementos para sostener que la afirmación "existe un conjunto de números naturales con la propiedad P" es verdadera o falsa, sino que más bien tendremos que decir que no tiene un significado intuitivo preciso.[8] Todo esto se aplica a conceptos similares como "función", "relación", etc. Podemos poner muchos ejemplos concretos intuitivamente bien definidos de funciones $f : \mathbb{N} \longrightarrow \mathbb{N}$, pero no podemos afirmar nada sobre la totalidad de ellas salvo en los casos en los que tengamos un argumento genérico aplicable a cualquiera de ellas.

Por último, una afirmación como $|\mathcal{P}\mathbb{N}| = \aleph_1$, es decir, la hipótesis del continuo, está sin duda en el caso 3. No tiene ningún sentido razonar sobre ella si no es en el marco de una teoría axiomática formal, pues no es posible atribuirle ningún significado intuitivo. Por reformularla de la manera más simple (eliminando la referencia a $\aleph_1$), podemos expresarla así:

> *Si A es un subconjunto infinito de $\mathcal{P}\mathbb{N}$, o bien existe una biyección $f : \mathbb{N} \longrightarrow A$, o bien existe una biyección $f : A \longrightarrow \mathcal{P}\mathbb{N}$.*

Pero el problema no es que no sepamos si esto es cierto o no. Hay muchas afirmaciones con un significado intuitivo preciso que no sabemos si son verdaderas o falsas. El problema es que no podemos atribuir un significado intuitivo a esta afirmación. Podemos hablar intuitivamente de ciertos conjuntos y de ciertas aplicaciones biyectivas, pero no tenemos nada en qué fundamentar que, o bien todo conjunto A cumple lo indicado, o bien existe uno que no lo cumple. Simplemente, tal afirmación trasciende con mucho el marco conceptual determinado por nuestra intuición espacio-temporal. O, al menos, si alguien opina que no es así, y que la hipótesis del continuo tiene un significado preciso que nos permite considerarla al margen de cualquier teoría axiomática formal, tendrá que justificarlo, y en ausencia de tal justificación tendremos que abstenernos de hablar informalmente de la hipótesis del continuo o de cualquier otra afirmación análoga, porque no podemos controlar que lo que vayamos a decir sea cierto.

[8] No estamos en condiciones de dar ningún ejemplo sencillo de esta posibilidad, pero mencionaremos muy superficialmente que en teoría de conjuntos es posible definir algo llamado $0^\sharp$ que admite varias definiciones equivalentes, y una de ellas permite considerar que es un conjunto de números naturales con cierta propiedad P, de tal modo que en ZFC no es posible demostrar que exista el conjunto $0^\sharp$, ni tampoco que no exista. Su existencia tiene repercusiones muy drásticas en la estructura general de los conjuntos y en particular sobre la aritmética cardinal. Pero, ¿tiene sentido plantearse si la afirmación "existe $0^\sharp$" es verdadera o falsa? O, más directamente: ¿tiene esa afirmación un significado intuitivo preciso (u objetivo en algún sentido)? A falta de que alguien lo justifique satisfactoriamente, no es algo que podamos dar por hecho sin más.

Éste es un punto de los que más inducen a error a los matemáticos habituados al razonamiento formal cuando tratan de razonar informalmente: recelan de usar conceptos particulares (como "conjunto" o "definición") porque no podemos definirlos en general (lo cual no invalida su uso en casos concretos intuitivamente precisos) y, en cambio, no muestran el menor recelo a decir "para todo" o "existe" porque estos conceptos se pueden usar sin restricciones en el contexto de la matemática formal, sin tener en cuenta que al razonar informalmente no es lo mismo un concepto de tipo 1 (sobre el que tiene perfecto sentido decir "para todo" o "existe") o un concepto de tipo 2 (que podemos usar en casos concretos, pero no podemos atribuir un significado preciso a una afirmación del tipo de "para todo conjunto" o "existe un conjunto", salvo si contamos con un argumento general particularizable a cada caso particular o sabemos construir un conjunto concreto que pruebe la existencia).

Formalismo, finitismo, platonismo Antes de seguir analizando las características del razonamiento informal, con lo visto hasta ahora podemos comentar brevemente las distintas opiniones de los matemáticos sobre lo que hay al otro lado del "muro formalista" que separa su actividad matemática de la filosofía de las matemáticas. La actitud de decir "no sé ni me importa qué son los conjuntos" puede tacharse de "cínica", como una forma (legal, ciertamente, pero algo "deshonesta") de eludir la responsabilidad de aclarar cuál es la naturaleza de los entes matemáticos. Cuando a los matemáticos se les pregunta cómo conciben esta "solución formalista" de la fundamentación de la matemática, se dividen en un abanico de posturas filosóficas que *grosso modo* se resumen en tres categorías:

1) Como ya hemos indicado, un formalista es quien considera que la matemática no es ni más ni menos que demostrar teoremas a partir de los axiomas de la teoría de conjuntos siguiendo las leyes de la lógica formal. Los formalistas más extremos considerarán que no hay nada más, que las palabras que emplean los matemáticos ("conjunto", "pertenencia", "función", "derivada", "integral", etc.) no son más que palabras, palabras coherentes, pero palabras, como las de una novela bien escrita, sin contradicciones y en la que "todo encaja". En todo caso, es una "novela" especial porque es objetiva, en el sentido de que no está en manos del "autor" decidir "si se salva el protagonista o muere al final". Los más moderados reconocerán que, sí, las afirmaciones matemáticas significan algo, pero todo lo referente al significado posible es poco riguroso, no es científico, es "de andar por casa", y es algo que a lo sumo puede tener interés desde un punto de vista heurístico o didáctico, pero no matemático en sentido estricto.

Esto no quita para que todo matemático, cuando razona sobre sus objetos de estudio (sean números, superficies, espacios topológicos, etc.) piense en ellos como objetos reales, pero esto no tiene por qué ser diferente de lo que hace un novelista cuando escribe una novela fantástica, en la que puede imaginarse con todo detalle a unos elfos luchando contra unos orcos y lo plasma por escrito. El hecho de que se imagine su relato como algo real no supone —si está bien de la cabeza— que no tenga claro que todo es ficción, y que su novela no es más que palabras ordenadas de forma interesante, pero nada más que palabras.

2) En el extremo opuesto encontramos a los *platonistas,* que son quienes consideran —como Gödel, por ejemplo— que la teoría de conjuntos describe una realidad objetiva que va mucho más allá de lo que nuestra mente puede abarcar con precisión suficiente como para razonar sobre ello "de forma natural", es decir, informal, por lo que para estudiarla es necesario razonar formalmente, a partir de axiomas, pero que tiene sentido discutir qué axiomas son "correctos", en el sentido de que describen realmente esa realidad objetiva, y qué axiomas darían lugar a una teoría, tal vez coherente, pero falsa. Si para un formalista las matemáticas son una novela de ficción, para los platonistas la matemática es un relato fiel a una historia real.

3) Entre ambos extremos se sitúan los *finitistas,* —como Hilbert— que son quienes piensan que la matemática es como una novela que combina hechos históricos con hechos de ficción, es decir, que de los conceptos de los que se puede hablar a través de la teoría de conjuntos hay algunos que tienen un significado objetivo intuitivo, como los números naturales, y otros que son meras "ficciones útiles", en el sentido de que usando esos argumentos con objetos ficticios es posible demostrar hechos verdaderos sobre los objetos reales. Un ejemplo sería la demostración del llamado *Último Teorema de Fermat,* es decir, la afirmación de que la ecuación

$$x^n + y^n = z^n$$

no tiene soluciones enteras con $x, y, z \neq 0$ cuando n es un número natural mayor que 2. La única demostración que se conoce de este resultado utiliza conceptos conjuntistas muy sofisticados cuya existencia cuestionaría un finitista y, sin embargo, la conclusión es una afirmación sobre los números naturales que tiene sentido decir que es verdadera. El término "finitista" se debe a que en general la porción de las matemáticas a la que un finitista reconoce un significado es la relacionada con los conceptos que se pueden describir en términos de conjuntos y procesos finitos, aunque un finitista moderado puede aceptar también la existencia objetiva de ciertos conjuntos infinitos, por ejemplo, que sus elementos puedan enumerarse, como hemos discutido más arriba.

La diferencia entre un finitista y un platonista es puramente filosófica, es decir, queda "al otro lado del muro de defensa formalista", y no tiene ninguna consecuencia práctica en su actividad como matemáticos, ni siquiera si tratan cuestiones metamatemáticas.[9] En cambio, ya hemos explicado por qué el formalismo hace aguas cuando trata de abordar la metamatemática.

Cuando Hilbert planteó su programa de encontrar una axiomática formal para la teoría de conjuntos que fuera consistente y completa, tenía claro que los argumentos metamatemáticos necesarios tenían que ser inevitablemente informales, y fijó como requisito que fueran finitistas. Así, Hilbert admitía que es posible razonar informalmente sobre conjuntos finitos y esperaba que las técnicas de razonamiento finitistas bastaran para probar la consistencia y la completitud

[9]Un caso distinto sería el de un "platonista ingenuo" que no viera inconveniente alguno en razonar informalmente sobre conceptos matemáticos arbitrarios, al que habría que pedir explicaciones sobre cómo interpreta el hecho de que razonando informalmente sin restricción alguna se pueda caer en las paradojas de la teoría de conjuntos.

de una teoría axiomática que permitiera formalizar toda la matemática, no sólo la parte finitista, para la que el formalismo resulta superfluo.

Los números naturales Un concepto que resulta imprescindible considerar informalmente en cualquier teoría metamatemática es el de número natural. Según hemos explicado, esto es posible porque tenemos un concepto intuitivo de "número natural" que nos permite razonar sobre ellos con la misma objetividad —o incluso mayor— que con la que podemos razonar sobre cualquier concepto físico. Tan objetivo es afirmar que los planetas más próximos al Sol son Mercurio, Venus, la Tierra y Marte como decir que los primeros números primos son $2, 3, 5, 7$. No es algo que esté en nuestras manos decidir, sino algo objetivo que meramente podemos constatar.

Las teorías axiomáticas que permiten hablar sobre números naturales se llaman *teorías aritméticas.* Si desarrolláramos una teoría aritmética en la que pudiera demostrarse rigurosamente que el 8 es primo, podríamos concluir que, o bien sus axiomas son falsos, o bien la definición de "primo" considerada no se corresponde con la usual. Si entendemos por "primo" lo que usualmente se entiende como tal, no tenemos más remedio que concluir que el 8 no es primo. En cambio, podemos tener dos teorías axiomáticas de conjuntos, en una de las cuales se demuestre que $2^{\aleph_0} = \aleph_1$ —por ejemplo, porque sea uno de sus axiomas— y otra en la que se demuestre que $2^{\aleph_0} = \aleph_7$ (en ambos casos con la mismas definiciones), y en este caso no tenemos ningún argumento para afirmar que una de las dos contradice nuestro conocimiento intuitivo sobre los conjuntos, porque nuestro concepto intuitivo de "conjunto" es muy limitado y no alcanza a dar sentido a una afirmación como la hipótesis del continuo.

Así pues, al contrario de lo que querría un formalista, no podemos afirmar que los números naturales son tales o cuales objetos determinados por una teoría aritmética, porque si la teoría en cuestión dice que el 8 es primo, no concluiremos de ahí que el 8 es primo, sino que la teoría está mal. Por el contrario, los teoremas de cualquier teoría aritmética deben concordar con nuestro concepto intuitivo previo de "número natural" para que podamos aceptarla como válida.

En términos intuitivos, si queremos explicar qué son los números naturales a alguien que nunca haya oído hablar de ellos, debería bastar con esto:

1. Existe un primer número natural, al que llamamos *cero.*
2. Cada número natural tiene asociado otro al que llamamos su *siguiente.*
3. Si vamos enumerando los números naturales empezando con el cero, luego el siguiente de cero, luego el siguiente del siguiente de cero, y así sucesivamente, cada número que añadimos a la sucesión es distinto de los precedentes, y los números naturales son únicamente los que van apareciendo en esa sucesión.

En todo caso, habría que aclarar que los números naturales no son nada más que esto. El cero no es más que el primer número natural, y no tiene otro rasgo distintivo aparte de éste. Cualquier característica adicional que queramos suponer en él será un añadido ajeno al concepto de número natural propiamente

dicho. Por ejemplo, podemos adoptar el convenio de representarlo así: 0, pero esto es un mero convenio. Si alguien decidiera representarlo como $\odot$, estaría hablando del mismo concepto con otro nombre. Igualmente podemos dar nombres arbitrarios a los primeros números naturales y así, llamar 1 al siguiente de 0, y 2 al siguiente de 1, etc. y podemos crear un criterio genérico para nombrar números naturales arbitrarios, como la notación decimal que nos permite interpretar 12 como el siguiente del siguiente del siguiente del siguiente del siguiente del siguiente del siguiente del siguiente del siguiente del siguiente del siguiente de cero, o cualquier otro alternativo, como el que nos permite identificar XII como el mismo número natural. Pero todo esto son convenios lingüísticos (o matemáticos, según lo elaborados que sean) en torno a un mismo concepto intuitivo de número natural.

Por último, para que podamos decir que alguien tiene un pleno conocimiento intuitivo de lo que son los números naturales, tenemos que enseñarle a contar, de modo que al ver

sea capaz de reconocer que en la figura hay 5 puntos. Quien entienda lo dicho hasta aquí, sabe perfectamente lo que son los números naturales, y está en condiciones de convencerse razonadamente de cualquiera de sus propiedades. Por ejemplo, de que todo conjunto de números naturales tiene un mínimo elemento, tal y como hemos razonado en la página xxxiv.

Es posible que un formalista para quien el razonamiento de la página xxxiv sea enfermizo, pero que se ha resignado a admitir que necesita tratar intuitivamente con los números naturales para entender la metamatemática, nos plantee lo siguiente: Vale, los números naturales tienen un sentido intuitivo objetivo, pero ¿qué propiedades necesitamos admitir que cumplen para demostrar a partir de ella los resultados metamatemáticos que queremos probar?

Una variante de la caracterización de Smullyan de los formalistas sería decir que un formalista es un matemático que está dispuesto a aceptar cualquier cosa con tal de que se la presenten como axioma y no le lleve a contradicciones, por lo que no es de extrañar que pretenda tomar como axiomas todas las afirmaciones sobre números naturales cuya demostración intuitiva no le parezca aceptable para deducir formalmente de ellas todas las demás. Aunque es un planteamiento artificial y retorcido, debemos conceder que su pregunta tiene sentido y, de hecho, la respuesta es interesante.

En el siglo XIX se propusieron varias teorías aritméticas formales para desarrollar la aritmética de los números naturales a partir de unos pocos principios, aunque la más famosa es la que publicó Giuseppe Peano en 1889. Peano dio nueve axiomas, pero cuatro de ellos tenían que ver con el uso formal del signo $=$, de modo que los axiomas que trataban propiamente con los números naturales eran los cinco restantes. Prescindiendo de su formalismo, estos axiomas afirmaban lo siguiente:

1. El cero es un número natural.

2. Todo número natural tiene un siguiente (que es otro número natural).

3. Números naturales distintos tienen siguientes distintos.

4. El cero no es el siguiente de ningún número natural.

5. (Principio de inducción) Si un conjunto A contiene al 0 y, supuesto que contenga a un número natural n, también contiene a su siguiente, entonces A contiene a todos los números naturales.

En ZFC se demuestra que estas cinco afirmaciones (que en ZFC no son axiomas) caracterizan al conjunto de los números naturales, en el sentido de que si tenemos un conjunto $\mathbb{N}$ tal que, definiendo "n es un número natural" como $n \in \mathbb{N}$, se cumplen las cinco propiedades anteriores con un cierto $0 \in \mathbb{N}$ y una cierta función "siguiente" $S : \mathbb{N} \longrightarrow \mathbb{N}$ entonces $\mathbb{N}$ es "esencialmente el mismo" que cualquier otro conjunto $\mathbb{N}'$ que cumpla lo mismo con otro $0'$ y otra función S', donde "esencialmente el mismo" o "isomorfo" es un concepto que puede definirse con precisión.

Por ello un formalista estaría encantado de que le aceptáramos que *los números naturales son objetos cualesquiera que cumplan los axiomas de Peano.* Desde un punto de vista intuitivo, el principio de inducción hace referencia a "la totalidad de los conjuntos", lo cual, según hemos visto, hay que entenderlo con precauciones. Un enunciado alternativo es el siguiente:

5. (Principio de inducción) Si el cero tiene una propiedad P y cuando un número natural tiene la propiedad P también la tiene su siguiente, entonces todo número natural tiene la propiedad P.

Es probable que el formalista prefiera la primera versión, porque la palabra "conjunto" le suena más "rigurosa" que "propiedad", pero desde un punto de vista intuitivo son esencialmente lo mismo. En cambio, la segunda versión abre la puerta a una precisión del concepto de inducción consistente en limitarlo a determinadas propiedades que podamos considerar intuitivamente bien definidas.

En principio podríamos pensar en considerar únicamente las propiedades que pueden definirse a partir de los conceptos de "cero" y "siguiente", pero esto es demasiado pobre. Para ir por ese camino debemos añadir algunos hechos básicos a los axiomas de Peano. Si, por abreviar, usamos la letra S para referirnos al siguiente de un número natural y los signos $+$ y $\cdot$ para nombrar la suma y el producto, como es habitual, los nuevos axiomas son:

6. Sobre los números naturales hay definida una suma, de modo que la suma de dos números naturales es un número natural, y además:

$$x + 0 = x, \qquad x + Sy = S(x + y).$$

7. Sobre los números naturales hay definido un producto, de modo que el producto de dos números naturales es un número natural, y además:

$$x \cdot 0 = 0, \qquad x \cdot Sy = x \cdot y + x.$$

Si consideramos que el principio de inducción es aplicable a todas las propiedades P que pueden definirse exclusivamente a partir de los conceptos de "cero", "siguiente", "suma" y "producto" (lo que llamaremos propiedades aritméticas) entonces los axiomas de Peano son suficientes para demostrar prácticamente todo lo que cualquiera sabe sobre los números naturales y mucho más. Una forma de comprobar que esto es así, con la seguridad de que no estamos añadiendo inadvertidamente ningún hecho adicional es estudiar la teoría aritmética cuyos axiomas son ni más ni menos que los axiomas de Peano, que recibe el nombre de *Aritmética de Peano.* Lo que afirmamos es que toda la aritmética básica es demostrable formalmente en la Aritmética de Peano.

Sin embargo, una de las consecuencias más sorprendentes del teorema de completitud semántica de Gödel es que ninguna teoría aritmética caracteriza a los números naturales, en el sentido de que, dada cualquier teoría aritmética consistente, siempre es posible encontrar unos objetos que cumplen sus axiomas entre los cuales hay algunos que no son ni el 0, ni el 1, ni el 2, ni el 3, ni, en general, ninguno de los que resultan de aplicar sucesivamente al 0 la operación "siguiente". Más aún, si la teoría es lo suficientemente potente como para distinguir entre conjuntos finitos y conjuntos infinitos, siempre es posible encontrar objetos que cumplen los axiomas de la teoría de modo que uno de ellos sea un conjunto que satisface la definición considerada de "conjunto finito" a la vez que contiene al 0, y al 1, y al 2, y al 3 y, en general, a todos los números naturales que se obtienen del 0 aplicando sucesivamente la operación "siguiente": el concepto de finitud, como el de "número natural" no pueden caracterizarse formalmente.[10]

La formalización de la metamatemática No es raro que un formalista que —desde sus prejuicios— intente asimilar la metamatemática se quede con la impresión de que lo que hacen los libros que dicen razonar informalmente sea usar ZFC sin decirlo (o sin confesarlo). Esto es así porque si hablamos, por ejemplo, de números naturales, la diferencia entre entender $3 + 2 = 5$ como un teorema demostrado formalmente o como una afirmación intuitivamente verdadera no se aprecia en la propia formulación $3 + 2 = 5$, sino que consiste en los criterios que usamos para aceptarla (sea como teorema formal, sea como afirmación verdadera). Más aún, ZFC es una teoría axiomática suficientemente potente como para formalizar toda la matemática intuitiva y, una vez formalizados en ZFC los resultados intuitivos más básicos, a la hora de probar algo más elaborado, no hay a nivel superficial ninguna diferencia apreciable entre usar un hecho elemental porque es intuitivamente obvio o usarlo porque ya ha sido formalmente demostrado en ZFC, por lo que un razonamiento intuitivo y su formalización resultan superficialmente indistinguibles.

Esto no quita para que, si no queremos caer en un círculo vicioso, el proceso de fundamentación de la matemática deba concebirse como, en primer lugar, una construcción metamatemática informal de un aparato lógico formal que, en particular, nos proporciona ZFC como marco para el razonamiento formal

[10] Estos resultados se prestan a muchas matizaciones que consideraremos en el capítulo IV, cuando podamos demostrar y analizar con detalle el teorema de completitud y sus consecuencias.

y, en segundo lugar, la presentación de toda la matemática en el seno de ZFC, lo cual incluye la formalización de la metamatemática que convierte a la lógica matemática en una disciplina matemática más, al mismo nivel que el álgebra, el análisis, la geometría, etc. que puede desarrollarse empleando las mismas técnicas formales, lo que permite expandir extraordinariamente su alcance respecto a lo que puede hacerse informalmente, cuando nos limitamos a considerar situaciones con pleno significado intuitivo.

La posibilidad de formalizar la metamatemática no es una mera anécdota, sino que muchos resultados de la lógica matemática, entre ellos los teoremas de incompletitud de Gödel, se basan en la relación que existe entre las teorías metamatemáticas y sus formalizaciones correspondientes.

Además, hay otro motivo por el que la formalización de la metamatemática tiene interés. Una de las obsesiones del formalista desubicado que se resigna a digerir una metamatemática intuitiva es que se le aclare cuáles son exactamente los supuestos que la metamatemática asume como verdaderos, sin ser consciente de que ese planteamiento ya es desafortunado en sí mismo. El formalista pedirá "principios generales", cuando un "principio general" será a menudo demasiado general para que un finitista pueda comprometerse a admitirlo como válido siempre que surja la posibilidad. Un finitista juzgará en cada caso en concreto si el principio resulta concluyente.

No obstante, debemos concederle al finitista que inventariar de algún modo los hechos intuitivamente ciertos de los que depende un resultado metamatemático es un problema que tiene sentido y es interesante, y una forma de hacerlo es formalizar cada razonamiento metamatemático en una teoría axiomática formal lo más débil posible. Así, hay resultados metamatemáticos formalizables en teorías aritméticas más débiles incluso que la Aritmética de Peano, y otros que requieren teorías un poco más fuertes. Por otro lado, tratar de formalizar unos resultados (informales o formales) en una teoría axiomática débil complica un poco los argumentos (conseguir algo con menos medios, puede ser posible, pero más complicado), por lo que puede hacerlos más oscuros y artificiales. Por ello, en este libro presentaremos informalmente todos los argumentos metamatemáticos que necesitaremos, y sólo nos ocuparemos de la formalización de la matemática necesaria, por una parte, para demostrar los teoremas de incompletitud y, por otra, para mostrar cómo la lógica matemática puede convertirse en una parte más de la matemática presentable en ZFC.

Libros relacionados En este libro daremos todos los detalles sobre la formalización del razonamiento matemático intuitivo que permite extender su alcance para que podamos razonar con seguridad en contextos que la intuición no alcanza (o sólo alcanza parcialmente), y en particular construiremos y estudiaremos ZFC. Demostraremos varios teoremas fundamentales que determinan hasta qué punto el razonamiento formal cumple las expectativas que tenemos sobre él, entre ellos el teorema de completitud semántica y los teoremas de incompletitud de Gödel, y sentaremos las bases lógicas necesarias para obtener pruebas de consistencia, como la independencia de la hipótesis del continuo (aunque ésta no la veremos aquí).

No entraremos en los aspectos más profundos de la teoría de conjuntos, como la teoría de cardinales transfinitos de Cantor, de la que hemos hablado superficialmente. Ésta la encontrará el lector en nuestro libro *Teoría de conjuntos* [TC], donde presentaremos la teoría de conjuntos sin enfatizar en tecnicismos lógicos, como se exponen habitualmente las matemáticas en cualquier libro, de modo que para todas las cuestiones técnicas de carácter lógico remitiremos al lector a este libro. Si un lector no está muy familiarizado con el razonamiento matemático, no sería descabellado que se planteara leer primero [TC] sin preocuparse por los tecnicismos lógicos y leyera este libro después, cuando ya tuviera cierta familiaridad con la teoría de conjuntos. Por otro lado, en *Pruebas de consistencia* [PC] el lector encontrará la prueba de que la hipótesis del continuo no puede demostrarse ni refutarse en ZFC junto con muchos otros resultados de la misma naturaleza. La lectura de [PC] requiere haber leído antes [TC] y, nuevamente, sólo será necesario recurrir a este libro para rellenar algunos detalles técnicos que no es descabellado pasar por alto, al menos en una primera lectura. Por otro lado, en *La lógica del finitismo* [LF] presentaremos prácticamente todos los resultados metamatemáticos expuestos en este libro, pero no informalmente, sino formalizados en teorías aritméticas débiles, para satisfacer la curiosidad cuasimorbosa del finitista sobre qué hemos empleado concretamente a nivel intuitivo para construir y estudiar la matemática formal. Puede decirse también que este libro pretende exponer la metamatemática que necesitará cualquiera interesado en comprender los fundamentos de la teoría de conjuntos, mientras que [LF] se centra en los fundamentos de la aritmética. Por último, muchos resultados de [LF] requieren de técnicas de la teoría de la demostración más potentes que las que vamos a presentar aquí, para las cuales remitiremos a nuestro libro *El cálculo secuencial de Gentzen* [CS].

Primera parte

Lógica de primer orden

Capítulo I

Lenguajes y modelos

El propósito de la primera parte de este libro es formalizar el razonamiento matemático en el sentido que hemos explicado en la introducción: en principio, lo que determina que un razonamiento sea válido es que podamos garantizar que sus conclusiones son verdaderas bajo el supuesto de que sus premisas lo sean, pero lo que queremos hacer ahora es caracterizar los razonamientos válidos en términos que no exijan en ningún momento estudiar si una afirmación dada es o no verdadera respecto de ninguna interpretación posible de los conceptos que involucra. Sólo así podremos usar la lógica formal para razonar incluso cuando no tengamos una idea clara de qué objetos podrían satisfacer nuestras premisas.

El razonamiento informal se expresa en lenguajes naturales, como el castellano que usamos en este libro. Sin embargo, unas reglas formales de razonamiento totalmente precisas tienen que hacer referencia a la forma sintáctica precisa de las afirmaciones involucradas (sin hacer referencia para nada a su posible significado), pero en las lenguas naturales es prácticamente imposible analizar la estructura sintáctica de una frase sin entrar en consideraciones semánticas.[1]

Por ello, en este primer capítulo definiremos lenguajes formales apropiados para expresar cualquier afirmación de interés matemático y en el capítulo siguiente presentaremos un cálculo deductivo formal (es decir, totalmente independiente del posible significado de las afirmaciones consideradas) que, según demostraremos posteriormente, capturará completamente nuestra capacidad de razonamiento informal.

Para definir y desarrollar los conceptos necesarios para formalizar el razonamiento matemático necesitamos razonar, y es obvio que no podemos razonar formalmente porque precisamente estamos tratando de determinar reglas for-

[1] Un ejemplo clásico es el resultado que obtuvo un programa de ordenador de análisis sintáctico cuando se le dio como entrada la frase *"Time flies like an arrow"*. El programa detectó varias estructuras sintácticas posibles muy dispares entre sí, entre ellas las que corresponderían en castellano a *"El tiempo vuela como una flecha"*, *"Mide la velocidad de las moscas que son como una flecha"*, *"Mide la velocidad de las moscas como lo haría una flecha"* o *"A las moscas del tiempo les gusta una flecha"*. Sólo la semántica nos permite descartar algunas estructuras sintácticas posibles porque dan lugar a interpretaciones surrealistas.

males de razonamiento. Por eso debe quedar claro que razonar informalmente no está reñido en absoluto con razonar con rigor. Lo que sucede es que el rigor de un razonamiento informal (metamatemático) no se garantiza como los matemáticos están acostumbrados a garantizar el rigor de sus razonamientos formales (matemáticos), es decir, ajustándose a unas formas de razonamiento prefijadas, sino que se debe garantizar semánticamente, asegurándonos de que todos los términos que empleamos tienen un significado intuitivo preciso y de que todo lo que decimos es verdad.

1.1 Estructuras

En principio es posible razonar sobre objetos muy diversos, y algunos muy difíciles de tratar. Por ejemplo, uno puede tener necesidad de juzgar, de entre dos personas, cuál es más inteligente. Y es un problema complicado porque no hay un concepto preciso de "inteligencia". Y eso no significa que la pregunta carezca de sentido, pues hay casos en los que se puede concluir sin lugar a dudas que una persona es más inteligente que otra. Por otra parte, a menudo hay que tener presente que una afirmación que podía ser verdadera ayer ya no tiene por qué serlo hoy, porque el mundo cambia (un antibiótico que combatía bien una bacteria hace diez años ya no tiene por qué ser efectivo contra ella hoy).

Afortunadamente, el razonamiento matemático no requiere considerar conceptos imprecisos y variables como los de los ejemplos anteriores, sino que la fundamentación de las matemáticas sólo requiere que seamos capaces de precisar formalmente qué razonamientos son aceptables cuando consideramos únicamente interpretaciones posibles de los conceptos involucrados en "realidades" sencillas, formadas por objetos inmutables con propiedades perfectamente determinadas. Más aún, siempre podremos suponer que todas las propiedades relevantes estarán dadas de antemano (sin perjuicio de que a partir de ellas se puedan definir otras nuevas). Dedicamos esta primera sección a concretar el tipo de "realidades" que pretendemos describir con los lenguajes que vamos a definir.

Conjuntos Cuando queramos atribuir un significado a las afirmaciones de un lenguaje formal, lo primero que tendremos que hacer es especificar el conjunto M de los objetos de los que pretendemos hablar. Tal y como explicamos en la introducción, por "conjunto" entendemos aquí una colección de objetos "bien definida", en el sentido de que no haya duda sobre qué significa que un objeto sea o no uno de sus elementos. Pero, más aún, debemos exigir que no haya duda sobre qué significa que una propiedad se cumpla para todos los objetos de M, o que exista un objeto en M que cumpla una determinada propiedad (supuesto que la propiedad esté bien definida para cada objeto).

Por ejemplo, ya hemos explicado que el conjunto $\mathbb{N}$ de los números naturales cumple estos requisitos. Decir que todo número natural cumple una propiedad P tiene un significado muy concreto (supuesto que lo tenga la propiedad P), pues significa que el 0 tiene la propiedad P y que el 1 la tiene, y que el 2 la tiene y

que por mucho que avancemos en la sucesión de los números naturales nunca encontraremos un número natural que no la tenga. Similarmente, decir que existe un número natural que tiene una propiedad P significa que el 0 la tiene, o que el 1 la tiene, o que el 2 la tiene, de manera que si continuamos avanzando por la sucesión de los naturales tarde o temprano encontraremos algún número que la tenga.

Por el contrario, también hemos visto que no sería admisible tomar como M el "conjunto" de todos los conjuntos de números naturales, ya que, aunque tenemos claro qué es un conjunto de números naturales, no tenemos claro que podamos asignar un significado objetivo a que todo conjunto de números naturales cumpla una determinada propiedad, aunque ésta esté perfectamente definida para cada conjunto concreto.

En muchas ocasiones (en definiciones, en razonamientos) diremos "sea M un conjunto arbitrario", pero esto no deberá entenderse como que sabemos asignarle un significado objetivo a las palabras "conjunto arbitrario", con el cual "nos atrevamos" a razonar informalmente, lo cual no es cierto. Por el contrario, una definición o razonamiento en donde aparezcan esas palabras no deberá entenderse como una auténtica definición o como un auténtico razonamiento, sino como un mero *esquema* de definición o razonamiento que sólo se convertirá en auténticas definiciones o razonamientos cuando se particularice a un conjunto M concreto que podamos considerar bien definido, como es el caso de $\mathbb{N}$, o de un conjunto finito $\{a, b, c\}$. En este último caso, afirmar que todo elemento de M cumple una propiedad P significa que la cumple a, y que la cumple b y que la cumple c. La propiedad P estará bien definida si no queda margen de duda sobre si cada uno de los tres elementos la cumple o no.

Relaciones Si M es un conjunto arbitrario y n es un número natural no nulo, una *relación n-ádica* R en M es cualquier criterio bien definido que asigne un valor de verdad (verdadero o falso) a cada n objetos de M (con posibles repeticiones) en un orden dado. Escribiremos $R(a_1, \ldots, a_n)$ para indicar que la relación R es verdadera sobre los n objetos de M indicados.

Por ejemplo, la relación Pn que se cumple cuando n es primo es una relación monádica en $\mathbb{N}$, la relación $x \mid y$ que se cumple cuando x divide a y es una relación diádica en $\mathbb{N}$, la relación $x \equiv y \pmod{z}$ que se cumple cuando $z \mid x - y$ es una relación triádica, etc.

Nuevamente tenemos la misma situación que con los conjuntos: no existe un concepto preciso de "relación bien definida", pero los tres ejemplos precedentes son ejemplos de relaciones en $\mathbb{N}$ para las que no cabe duda de que están bien definidas. Cuando hablemos de una relación n-ádica arbitraria en un conjunto arbitrario en el contexto de una definición o un razonamiento estaremos de nuevo ante un esquema de definición o razonamiento que sólo dará lugar a auténticas definiciones o razonamientos cuando se particularice a casos concretos de relaciones y conjuntos que podamos considerar bien definidos.

Por ejemplo, si en el conjunto $M = \{a, b, c\}$ consideramos la relación diádica R que se cumple exactamente sobre los pares $(a, b), (a, c), (b, c)$ tenemos

un conjunto bien definido con una relación diádica bien definida sobre él. El conjunto M podría ser un conjunto de tres personas y la relación R podría ser la relación "ser más inteligente que", concretada mediante cualquier criterio arbitrario que no cuestionaremos. Valorar quién es más inteligente queda fuera de nuestro alcance, pero si se ha establecido de algún modo quién es más inteligente, y podemos decir sin peros que a es más inteligente que b y que b es más inteligente que c, entonces con esta relación así precisada podemos tratar sin problema alguno.

En cualquier conjunto podemos considerar la *relación de identidad,* que representaremos por $x \equiv y$ y que es verdadera cuando x e y son el mismo objeto.

Funciones Si M es un conjunto arbitrario y n es un número natural no nulo, una *función n-ádica* f en M es cualquier criterio bien definido que a n objetos de M (con posibles repeticiones) en un orden dado les asigna otro objeto de M, que representaremos por $f(a_1, \ldots, a_n)$.

Aquí se aplican las mismas consideraciones que en los apartados precedentes. Ejemplos de funciones bien definidas sobre el conjunto $\mathbb{N}$ son la función sucesor (que es monádica), la suma y el producto, que son diádicas, o la función triádica dada por $f(x, y, z) \equiv xy + z$.

Estructuras Una *estructura* consiste en un conjunto M en el cual se han fijado varias relaciones y funciones.

Estas estructuras, formadas por un conjunto y unas relaciones y funciones bien definidas, sin ambigüedades, inmutables, son las "realidades simplificadas" que pretendemos estudiar a través de la lógica formal que vamos a introducir.

Por ejemplo, la *estructura de la aritmética de Peano* está formada por el conjunto de los números naturales con las funciones sucesor, suma y producto (que son las que aparecen en los axiomas de Peano).

Observemos que el castellano, como cualquier lengua natural, nos permite hablar sobre los conceptos más diversos, desde los dedos de nuestra mano hasta de astrología. Según de qué hablemos, nuestras palabras pueden tener un significado claro, dudoso o incluso ser mera palabrería sin sentido. Pero si fijamos una estructura en los términos que hemos establecido (un conjunto bien definido con relaciones y funciones bien definidas) y nos comprometemos a hablar en castellano sin hacer referencia a ningún concepto que no sean los elementos del conjunto y las relaciones y funciones fijadas en la estructura, entonces el castellano nos permite razonar con absoluto rigor.

Nuestro propósito es demostrar que todo razonamiento informal en castellano sobre los objetos, relaciones y funciones de una estructura puede expresarse como una sucesión de afirmaciones de un lenguaje formal (es decir, afirmaciones de un lenguaje definido sin tener en consideración el posible significado de sus signos) conectadas por unas reglas de deducción formal fijadas de antemano (es decir, reglas que nos autorizan a pasar de unas afirmaciones a otras, no en

función de su significado, sino de su mera estructura, o forma, sintáctica). El razonamiento formal y el informal serán "lo mismo" en el mismo sentido en que un razonamiento en castellano es "lo mismo" que una traducción fiel al inglés.

1.2 Lenguajes formales y modelos

Vamos ahora a definir lenguajes adecuados para razonar sobre cualquier estructura prefijada. Concretamente, vamos a dar dos definiciones: la de lenguaje formal y la de modelo de un lenguaje formal. La idea es que un lenguaje formal es simplemente un inventario de los signos que vamos a usar, y un modelo de un lenguaje formal es una asignación de significado a cada uno de esos signos. Por razones didácticas empezaremos dando un esbozo provisional de ambas definiciones de forma simultánea, para que el lector se haga una idea aproximada de qué pretenden ser estos dos conceptos, y de la estrecha relación que hay entre ambos. Inmediatamente después daremos las definiciones precisas de forma separada y con algunos matices adicionales.

Como acabamos de decir, un lenguaje formal es un inventario de signos, cada uno de los cuales equivale más o menos a una palabra o expresión de un lenguaje natural. Se trata de listar de antemano todos los signos que aceptaremos como válidos en una afirmación formal. Pero no basta con dar una lista de signos, sino que para cada uno hay que especificar a qué categoría pertenece (el equivalente a la distinción en castellano entre sustantivos, adjetivos, verbos, etc.).

En cuanto a un modelo M de un lenguaje formal $\mathcal{L}$, es un criterio que asigna a cada signo del lenguaje una interpretación posible, que será de un tipo u otro según la categoría del signo. Pero antes de entrar en tales asignaciones el modelo debe fijar un *universo,* es decir, un conjunto cuyos elementos serán los objetos a los que harán referencia todas las afirmaciones del lenguaje formal (respecto del modelo en cuestión). En la práctica usaremos el mismo nombre para referirnos a un modelo y a su universo.

Los signos de un lenguaje formal pueden ser de cualquiera de las categorías siguientes:

Constantes Las constantes son los signos destinados a nombrar objetos (el equivalente de los nombres propios en castellano). Un modelo M de un lenguaje $\mathcal{L}$ debe asignar a cada constante c de $\mathcal{L}$ un objeto $\bar{c}$ de su universo al que llamaremos la *interpretación* de c en M o también el *objeto denotado* por c en M (es decir, el significado atribuido al signo c en M). Cuando convenga explicitar el nombre del modelo escribiremos $M(c)$ en lugar de $\bar{c}$.

Es importante comprender que c y $\bar{c}$ son cosas muy distintas. Por ejemplo, c puede ser un mero "garabato" en un papel, como "$\Diamond$", mientras que $\bar{c}$ puede ser el rey de España. Confundir c con $\bar{c}$ es como confundir un objeto con su nombre. Mi mesa tiene cuatro patas, pero no cuatro letras, la palabra "mesa" tiene cuatro letras, pero no cuatro patas.

Relatores Los relatores son los signos destinados a nombrar relaciones (el equivalente a los verbos en castellano). Al definir un lenguaje formal debemos

asociar a cada signo catalogado como relator un número natural no nulo al que llamaremos su rango. Los relatores de rango n se llaman relatores n-ádicos. Un modelo M de $\mathcal{L}$ debe asignar a cada relator n-ádico R de $\mathcal{L}$ una relación n-ádica $\bar{R}$ sobre su universo, que será la *interpretación* en M del relator R. Cuando convenga explicitar M escribiremos $M(R)$ en lugar de $\bar{R}$.

Por ejemplo, si el universo de un modelo es el conjunto de los números naturales, la interpretación $\bar{R}$ de un relator monádico R puede ser la relación "ser primo". Nuevamente, R no es más que un signo en un papel, por ejemplo, $\triangle$, que en principio no es más que eso, un "triangulito", mientras que $\bar{R}$ es una propiedad que puede tener o no cada número natural.

Exigiremos que todo lenguaje formal tenga al menos un relator diádico que llamaremos igualador, y que representaremos mediante el signo $=$, y en la definición de modelo exigiremos que la interpretación del igualador sea siempre la relación de identidad $\equiv$ en el universo del modelo.

Funtores Los funtores son los signos destinados a nombrar funciones, el equivalente a expresiones del estilo de "el padre de" en castellano. Al definir un lenguaje formal debemos asociar a cada signo catalogado como funtor un número natural no nulo al que llamaremos su rango. Los funtores de rango n se llaman funtores n-ádicos. Un modelo M de un lenguaje $\mathcal{L}$ debe asignar a cada funtor n-ádico f de $\mathcal{L}$ una función n-ádica $\bar{f}$ en su universo, que será la *interpretación* de f en el modelo. Cuando convenga explicitar M escribiremos $M(f)$ en lugar de $\bar{f}$.

Conectores lógicos Los conectores lógicos son signos destinados a construir nuevas afirmaciones a partir de otras dadas. Aunque podrían definirse muchos más, en la práctica consideraremos siempre cinco de ellos: un *negador* $\neg$, un *implicador* $\rightarrow$, un *conjuntor* (o conjunción) $\wedge$, un *disyuntor* (o disyunción) $\vee$ y un *coimplicador* (o bicondicionador) $\leftrightarrow$.

En la definición de modelo no incluiremos ninguna asociación de significado a los conectores lógicos, no porque no vayan a tenerla, sino porque les vamos a asociar siempre la misma, independientemente del modelo considerado. El significado asociado a cada conector no será ni un objeto, ni una relación ni una función, sino una "tabla de verdad". Concretamente, las tablas de verdad de los cinco conectores que estamos considerando serán las siguientes:

p	q	$\neg p$	$p \rightarrow q$	$p \wedge q$	$p \vee q$	$p \leftrightarrow q$
V	V	F	V	V	V	V
V	F	F	F	F	V	F
F	V	V	V	F	V	F
F	F	V	V	F	F	V

Esto significa, por ejemplo, en el caso del negador, que cuando p sea una afirmación verdadera en un modelo M (esto todavía tenemos que definirlo) entonces $\neg p$ será una afirmación falsa por definición (del negador), y viceversa. Podemos abreviar esto diciendo que $\neg$ significa "no".

Similarmente, el conjuntor $\wedge$ da lugar a una afirmación verdadera $p \wedge q$ en un modelo M cuando p y q son verdaderas, y falsa en caso contrario. Esto se resume en que $\wedge$ significa "y".

Igualmente podemos decir que el disyuntor $\vee$ significa "o", pero entendiendo que es un "o" no exclusivo, es decir, que no significa "o lo uno o lo otro, pero no ambas cosas", sino "al menos lo uno o lo otro, o tal vez ambas cosas".

El conector más delicado es el implicador. Traducido a palabras castellanas sería "si p entonces q", pero esto hay que entenderlo bien: La tabla de verdad que le asignamos hace que $p \to q$ sea verdadera salvo cuando p es verdadera y q es falsa. En otras palabras: decimos que $p \to q$ es verdadera si en caso de confirmar que p es verdadera podemos asegurar que q también lo es.[2] Una forma tal vez más clara de expresar el significado de una afirmación de tipo $p \to q$ (sin entrar en hipótesis sobre si p podría ser o no verdadera) es decir que $p \to q$ es verdadera cuando p es falsa o q es verdadera, pues esta disyunción refleja, en efecto, lo que indica la tabla de verdad.

El coimplicador es más simple: es verdadero cuando ambas afirmaciones son verdaderas o ambas son falsas, es decir, cuando ambas tienen el mismo valor de verdad. Significa, por tanto, "si y sólo si".

Variables Las variables son los signos destinados a tener una interpretación variable, como su nombre indica, por lo que no les asociaremos ningún significado en un modelo. Son el equivalente a los pronombres en castellano, los signos que usaremos para formalizar afirmaciones del tipo "sea x un número natural arbitrario" o "tomemos un x que sea primo y congruente con 1 módulo 4", etc.

Cuantificadores En cada lenguaje formal habrá dos signos llamados cuantificadores, un cuantificador universal (o generalizador) $\bigwedge$ y un cuantificador existencial (o particularizador) $\bigvee$. Se usarán siempre en combinación con variables, de modo que $\bigwedge x$ significará "para todo x", mientras que $\bigvee x$ significará "existe un x tal que ...". Tampoco les asignaremos ninguna interpretación en un modelo porque su interpretación será siempre la que acabamos de indicar.

Descriptor El descriptor $|$ será un signo que usaremos para especificar un objeto a partir de una propiedad que lo caracterice. Una expresión de la forma $x|P(x)$ se leerá "el único x que cumple la propiedad $P(x)$". Las expresiones de este tipo se llaman *descripciones.* Naturalmente, puede ocurrir que, bajo una interpretación dada de los signos de un lenguaje $\mathcal{L}$ (es decir, en un modelo de $\mathcal{L}$), no haya un único objeto que cumpla la propiedad $P(x)$, ya sea porque no haya ninguno o ya sea porque haya varios, en cuyo caso la interpretación de $x|P(x)$ en el modelo considerado queda indeterminada. Por ejemplo, "El x tal que x es

[2] Este uso de la implicación difiere del habitual en castellano, donde normalmente se exige una conexión causal entre las dos afirmaciones. Por ejemplo, nadie daría por válida una afirmación como "si tuviera un paraguas la Luna caería sobre la Tierra" sólo por el hecho de que no tengo un paraguas (mientras que, según la tabla de verdad, $F \to F$ es una implicación verdadera). El argumento sería que, si alguien me diera un paraguas, no por ello caería la Luna, pero esta clase de situaciones no se dan en nuestro contexto, pues vamos a tratar únicamente con realidades inmutables en la que o se tiene paraguas o no se tiene, pero no se puede pasar de no tenerlo a tenerlo. Así, en matemáticas una implicación se considera verdadera salvo que sea de la forma $V \to F$.

un número primo par" se interpreta como 2, pero, ¿qué es el x tal que x es un número primo impar, o el x tal que x es múltiplo de 2 impar? Cuando exista un único objeto en el universo del modelo que cumpla la propiedad considerada diremos que la descripción es *propia,* y en caso contrario diremos que es *impropia* (en el modelo considerado).

Para evitar que las descripciones impropias queden indeterminadas exigiremos que cada modelo de $\mathcal{L}$ especifique un objeto de su universo al que llamaremos *descripción impropia* en el modelo, y estableceremos que toda descripción a la que no podemos atribuirle una interpretación por ser impropia denotará por convenio a dicho objeto.

Con esto termina nuestro esbozo de las definiciones de lenguaje formal y modelo de un lenguaje formal. Antes de pasar a las definiciones exactas que vamos a considerar conviene destacar una cuestión sobre los signos y los nombres que les damos:

Sobre el uso y la mención Quizá, cuando hemos introducido los relatores, al lector le haya desconcertado la frase:

> Nuevamente, R no es más que un signo en un papel, por ejemplo, $\triangle$, que en principio no es más que eso, un "triangulito", mientras que $\bar{R}$ es una propiedad que puede tener o no cada número natural.

Uno puede preguntarse ¿en qué quedamos? ¿el relator del que hablamos es R o $\triangle$? Aquí es muy importante distinguir entre los signos y su significado. Vamos a adoptar el convenio de escribir entre comillas un signo cuando hablamos de él y no de su significado (y así, por ejemplo, decimos que España es un país, pero "España" tiene seis letras). Si un lenguaje formal $\mathcal{L}$ tiene una constante "$\bigstar$" que, en un determinado modelo, denota a alguien llamado Juan, entonces podemos decir que "$\bigstar$" significa Juan. Ponemos comillas porque estamos hablando del signo "$\bigstar$". Ahora bien, si decidimos llamar c al signo "$\bigstar$", de modo que "c" es el nombre que damos en castellano al signo "$\bigstar$", entonces podemos decir que c (sin comillas) significa Juan, mientras que "c" significa "$\bigstar$".

En general, si escribimos "$\bigstar$" como una forma de nombrar a Juan, entonces estamos *usando* el signo "$\bigstar$" para nombrar a Juan, mientras que si usamos "c" para nombrar a "$\bigstar$", estamos *mencionando* el signo "$\bigstar$" a través de su nombre (en castellano) "c".

El caso es que, por paradójico que parezca, nunca vamos a usar los signos de un lenguaje formal, sino que siempre vamos a mencionarlos a partir de nombres oportunos. Así, por ejemplo, distintos lenguajes formales pueden tener distintos signos como cuantificador universal: quizá el cuantificador universal de uno sea el signo "$\Diamond$", y en otro el signo "$\boxtimes$", etc., pero importará poco, porque mencionaremos el cuantificador universal de cualquier lenguaje formal con el signo "$\bigwedge$". Así, "$\bigwedge$" no será el cuantificador universal de ningún lenguaje formal (salvo que decidamos definir uno y asignarle precisamente ese signo como cuantificador universal), sino que será el nombre castellano con el que nos referiremos al cuantificador universal de cualquier lenguaje formal.

Es ilustrativo comparar el concepto de "signo de un lenguaje formal" con el de "pieza de ajedrez". Si queremos describir el juego del ajedrez, necesitamos decir que usa 32 piezas. ¿Qué son estas piezas? Pueden ser cualquier cosa: pueden ser figuras talladas en madera o marfil, pero también pueden ser signos que usemos en diagramas, e incluso pueden ser conceptos puramente abstractos: en teoría, dos personas podrían jugar al ajedrez "mentalmente", de manera que una dijera a la otra: "avanzo el peón del rey blanco una casilla" y el otro replicara "pues yo muevo el caballo del rey negro hasta la casilla $C5$", etc. Si así fuera, los peones y los caballos serían objetos de naturaleza similar a los números naturales, es decir, objetos de los que es posible hablar objetivamente, y asignarles objetivamente posiciones en tableros "mentales", y manipularlos según ciertas reglas objetivas, pero sin que tengan ninguna componente física, y en cualquier caso siempre podríamos mencionar las piezas con los nombres de "rey blanco", "rey negro", "reina blanca", etc.

Así pues, dado que en realidad (salvo en los ejemplos que estamos poniendo ahora para fijar ideas) nunca vamos a escribir en un papel los signos de un lenguaje formal, no es necesario que éstos sean realmente signos "plasmables por escrito". Pueden ser perfectamente conceptos abstractos sin soporte físico alguno, como las piezas de ajedrez no vinculadas a objetos físicos o como los números naturales.

Más aún, como la naturaleza de los signos de un lenguaje formal va a ser totalmente irrelevante, nada impide que convengamos en considerar únicamente lenguajes formales cuyos signos sean números naturales. Así, si convenimos en que un lenguaje formal $\mathcal{L}$ tiene por cuantificador universal al número natural 21, cuando usemos el signo "$\bigwedge$" para referirnos a dicho cuantificador universal resultará que "$\bigwedge$" no será más que una forma cómoda de nombrar al 21 cuando queremos pensar en él, no como número natural, sino como signo de un cierto lenguaje formal. Este nombre tiene la ventaja de que nos recuerda que es concretamente su cuantificador universal.

La ventaja de considerar que los signos de un lenguaje formal son simplemente números naturales es que entonces, al hablar de lenguajes formales, no estamos hablando sino de números naturales, y veremos que, interpretadas así, todas las propiedades de los lenguajes formales son afirmaciones aritméticas demostrables (formal o informalmente) a partir de los axiomas de Peano.

Pasamos ahora a dar la definición exacta de lenguaje formal. El lector debe fijarse en que en ella no hacemos ninguna referencia al concepto de modelo ni, más en general, a ningún significado posible de los signos del lenguaje. Ni siquiera al uso que pretendemos hacer de ellos:

Definición 1.1 Un *lenguaje formal de primer orden*[3] $\mathcal{L}$ es una colección de signos divididos en las categorías siguientes y de modo que cumplan las propiedades que se indican:

Variables Un lenguaje $\mathcal{L}$ debe tener infinitas variables. Cada variable debe tener asociado un número natural distinto al que llamaremos su *índice,* de

[3]El lector que quiera saber qué significa exactamente "de primer orden" encontrará la respuesta en la sección 10.1.

tal forma que todo natural es índice de una variable de $\mathcal{L}$. Llamaremos x_i a la variable de índice i de $\mathcal{L}$.

Constantes Un lenguaje $\mathcal{L}$ puede tener cualquier cantidad de constantes, desde ninguna hasta infinitas. En cualquier caso, cada constante debe tener asociado un *índice* natural. Llamaremos c_i a la constante de $\mathcal{L}$ de índice i (si existe) de modo que si $\mathcal{L}$ tiene $m+1$ constantes éstas serán $c_0, c_1, \ldots, c_m$, mientras que si $\mathcal{L}$ tiene infinitas constantes, los índices recorrerán todos los números naturales.

Relatores Cada relator debe tener asociado un número natural no nulo al que llamaremos su *rango.* Llamaremos relatores n-ádicos a los relatores de rango n. El número de relatores n-ádicos de $\mathcal{L}$ puede variar entre ninguno e infinitos. Cada relator n-ádico debe llevar asociado un *índice* distinto. Llamaremos R_i^n al relator n-ádico de índice i de $\mathcal{L}$ (si existe). Así, si $\mathcal{L}$ tiene $m+1$ relatores n-ádicos, éstos serán $R_0^n, \ldots, R_m^n$.

Todo lenguaje formal debe tener al menos el relator diádico R_0^2, al que llamaremos *igualador* o $=$.

Funtores Cada funtor ha de llevar asociado un *rango* y un *índice* en las mismas condiciones que los relatores. Llamaremos f_i^n al funtor n-ádico de índice i de $\mathcal{L}$ (si existe).

Negador Llamaremos $\neg$ al negador de $\mathcal{L}$.

Implicador Llamaremos $\rightarrow$ al implicador de $\mathcal{L}$.

Cuantificador universal (o generalizador) Lo llamaremos $\bigwedge$.

Descriptor Un lenguaje formal $\mathcal{L}$ puede tener o no descriptor y, según el caso, diremos que $\mathcal{L}$ es un lenguaje *con o sin descriptor.* Si existe lo representaremos por $|$.

Cada signo de $\mathcal{L}$ debe pertenecer a una de estas categorías y sólo a una. Las constantes, los funtores y los relatores distintos del igualador se llaman *signos eventuales* de $\mathcal{L}$, mientras que los restantes son *signos obligatorios.*

Si $\mathcal{L}$ es un lenguaje formal con descriptor, llamaremos $\underline{\mathcal{L}}$ al lenguaje que resulta de eliminarle el descriptor.

Quizá ayude comparar la definición precedente con la forma en que describiríamos, por ejemplo, el ajedrez. Si queremos explicarle a alguien cómo se juega al ajedrez, le diremos que se necesita un tablero y 32 piezas divididas en dos grupos de colores de forma que, a una de cada color —no importa cuál— hay que ponerle la etiqueta de rey, a otras la de alfil, etc. Cualquier cosa razonable puede hacer el papel de rey en una partida de ajedrez. Lo que convierte en rey a una pieza no es ninguna característica propia, sino tan sólo el convenio que los jugadores adoptan de usarla como rey. Igualmente, para construir un lenguaje formal necesitamos unos cuantos signos, no importa cuáles, y a cada uno le ponemos una etiqueta. Igual que en ajedrez moveremos de forma distinta

cada pieza según su nombre, también nosotros usaremos e interpretaremos de forma distinta las constantes y los relatores, pero eso no consta en la definición de lenguaje formal.

Observemos que en la definición de lenguaje formal sólo hemos incluido dos conectores y un cuantificador. Los restantes los definiremos más adelante a partir de éstos.

Una vez definido el concepto de lenguaje formal pasamos a definir el de modelo de un lenguaje formal:

Definición 1.2 Un *modelo* M de un lenguaje formal $\mathcal{L}$ viene determinado por:

1. Una colección de objetos U llamada *universo* de M. La colección U ha de tener al menos un objeto.
2. Un criterio que asocie a cada constante c de $\mathcal{L}$ un objeto $\bar{c}$ o $M(c)$ de U.
3. Un criterio que asocie a cada relator n-ádico R_i^n de $\mathcal{L}$ una relación n-ádica $\bar{R}_i^n$ o $M(R_i^n)$ en U. La relación $M(=)$ ha de ser la identidad $\equiv$.
4. Un criterio que asocie a cada funtor n-ádico f_i^n de $\mathcal{L}$ una función n-ádica $\bar{f}_i^n$ o $M(f_i^n)$ en U.
5. (Si $\mathcal{L}$ tiene descriptor) un elemento d de U al que llamaremos *descripción impropia* de M.

Claramente, si $\mathcal{L}$ tiene descriptor, todo modelo de $\mathcal{L}$ lo es de $\underline{\mathcal{L}}$ (olvidando la descripción impropia) y todo modelo de $\underline{\mathcal{L}}$ se convierte en un modelo de $\mathcal{L}$ fijando una descripción impropia.

En la práctica no distinguiremos como hemos hecho aquí entre un modelo M y su universo U, sino que cuando hablemos de un objeto a de M entenderemos que nos referimos a un objeto de su universo.

Ejemplos Para comprobar que la definición no es ambigua, vamos a definir un lenguaje concreto, al que llamaremos $\mathcal{L}_a$. Necesitamos signos. No hemos definido "signo", pero tampoco es necesaria una definición de "pieza" para jugar al ajedrez.

- Para las variables necesitamos infinitos signos. Nos sirven, por ejemplo

  $$|, \quad |-, \quad |--, \quad |---, \quad |----, \quad |-----, \quad |------, \quad \cdots$$

  Es decir, la variable x_0 es una raya vertical y, en general, si $i > 0$, la variable x_i es una raya vertical seguida de i rayas horizontales. Con esto quedan perfectamente definidas las variables de $\mathcal{L}_a$. No hay duda de que tenemos infinitas de ellas. Más concretamente, para cada natural i, sabemos perfectamente cuál es la única variable x_i de $\mathcal{L}_a$ a la que le corresponde el índice i.

- El lenguaje $\mathcal{L}_a$ tiene una única constante, digamos "$\triangle$".
- El lenguaje $\mathcal{L}_a$ tiene como único relator el igualador "$\smile$".
- Consideramos un funtor monádico "$\blacktriangle$" y dos funtores diádicos "$\Diamond$", y "$\Box$", de índices 0 y 1 respectivamente.
- Negador: "$\boxtimes$".
- Implicador: "$\triangleleft$".
- Cuantificador universal: "$\bigstar$".
- Descriptor: "ι".

Con esto, el lenguaje $\mathcal{L}_a$ queda completamente especificado. Si nos preguntan quién es f_0^1 en $\mathcal{L}_a$ la respuesta es clara: el signo "$\blacktriangle$", y si nos preguntan por R_3^5 hemos de contestar que no existe tal signo en $\mathcal{L}_a$.

Ahora consideremos otro lenguaje formal $\mathcal{L}'_a$ determinado por los signos siguientes:

- Como variables tomamos los números naturales que son potencia de dos:

$$1, \quad 2, \quad 4, \quad 8, \quad 16, \ \ldots$$

- La única constante es el número natural 3.
- El único relator es el igualador 5.
- $\mathcal{L}'_a$ tiene un único funtor monádico 6 y dos funtores diádicos 7 y 9, de índices 0 y 1 respectivamente.
- Negador: 10.
- Implicador: 11.
- Cuantificador universal: 12.
- Descriptor: 13.

Observamos que los lenguajes $\mathcal{L}_a$ y $\mathcal{L}'_a$ son esencialmente el mismo lenguaje, porque cada signo de uno se corresponde con un signo del mismo tipo en el otro, y nunca nos va a importar qué es concretamente un signo de un lenguaje formal, sino meramente que esté ahí para poder usarlo. Por ejemplo, nos dará igual si, cuando hablamos de la variable x_3 estamos hablando del signo "$|---$" o del número natural 8, o si el funtor f_1^2 es "$\Box$" o bien el número natural 9, etc.

Más aún, en lugar de usar los nombres que usamos por defecto para mencionar los signos de un lenguaje formal arbitrario, podemos elegir nombres específicos que nos resulten especialmente descriptivos. Por ejemplo, podemos adoptar

el convenio de representar por 0 la única constante c_0 de $\mathcal{L}_a$ (o de $\mathcal{L}'_a$), de representar por S el funtor monádico f^1_0 y de representar por $+$ y $\cdot$ los funtores f^2_0 y f^2_1, respectivamente. ∎

Como la naturaleza misma de los signos de un lenguaje formal es del todo irrelevante, en la práctica, al definir un lenguaje formal, nos limitaremos a especificar qué signos eventuales posee y, a lo sumo, a indicar el nombre con que convendremos referirnos a ellos, en caso de no optar por los nombres que usamos por defecto. Por ejemplo, la forma usual de introducir el lenguaje $\mathcal{L}_a$ del ejemplo anterior es la siguiente:

Definición 1.3 Llamaremos *lenguaje de la aritmética* a un lenguaje formal $\mathcal{L}_a$ (con descriptor) cuyos signos eventuales sean una constante 0, un funtor monádico S y dos funtores diádicos $+$ y $\cdot$.

Así, tanto $\mathcal{L}_a$ como $\mathcal{L}'_a$ satisfacen esta definición, pero nunca va a importar para nada si cuando hablamos de $\mathcal{L}_a$ nos referimos a $\mathcal{L}_a$, a $\mathcal{L}'_a$ o a cualquier otro lenguaje con signos eventuales de los mismos tipos.

Llamaremos *modelo natural de la aritmética* al modelo M de $\mathcal{L}_a$ cuyo universo es el conjunto $\mathbb{N}$ de los números naturales, en el que la constante 0 se interpreta como el número cero, el funtor S se interpreta como la función $\bar{S}$ que a cada natural le asigna su siguiente, y en el que los funtores $+$ y $\cdot$ se interpretan respectivamente como la suma y el producto de números naturales. Como descripción impropia fijamos el cero.

1.3 Expresiones, términos y fórmulas

La finalidad primera de los lenguajes formales es, por supuesto, construir afirmaciones con sus signos. Empezaremos con algunas consideraciones generales sobre las sucesiones de signos.

Definición 1.4 Sea $\mathcal{L}$ un lenguaje formal. Una *cadena de signos* de $\mathcal{L}$ es una sucesión finita de signos de $\mathcal{L}$ repetidos o no y en un cierto orden. Si $\zeta_1, \ldots, \zeta_n$ son cadenas de signos de $\mathcal{L}$ llamaremos $\zeta_1 \cdots \zeta_n$ a la cadena que resulta de yuxtaponer las cadenas $\zeta_1, \ldots, \zeta_n$ en este orden. En particular podemos nombrar una cadena nombrando a cada uno de sus signos en el orden en que aparecen.

Dos cadenas de signos ζ_1 y ζ_2 son *idénticas* si constan de los mismos signos en el mismo orden. Lo indicaremos así: $\zeta_1 \equiv \zeta_2$, con el signo que venimos empleando para referirnos en general a la relación de identidad en cualquier conjunto (y en caso contrario escribiremos $\zeta_1 \not\equiv \zeta_2$). Así pues, $\zeta_1 \equiv \zeta_2$ significa que "ζ_1" y "ζ_2" son dos nombres para la misma cadena de signos.

Si ζ es una cadena de signos de $\mathcal{L}$, llamaremos *longitud* de ζ al número de signos que componen ζ, contando cada uno tantas veces como se repita.

Claramente, si $\mathcal{L}$ es un lenguaje formal con descriptor, toda cadena de signos de $\underline{\mathcal{L}}$ lo es de $\mathcal{L}$ y toda cadena sin descriptores de $\mathcal{L}$ lo es de $\underline{\mathcal{L}}$.

Ejemplo Una cadena de signos de $\mathcal{L}_a$ es

$$\text{“}\boxtimes|--||\square\lhd\lhd\blacktriangle\text{”}$$

Su longitud es 8 (¡recordemos que “$|--$” es un solo signo!). Según el convenio adoptado, podemos referirnos a ella con el nombre “$\neg x_2x_0x_0f_1^2 \to\to f_0^1$”, o también “$\neg x_2x_0x_0\cdot \to\to S$”. Así podemos decir que “$\neg x_2x_0x_0\cdot \to\to S$” es una palabra castellana que nombra a una cadena de signos de $\mathcal{L}_a$ y también que $\neg x_2x_0x_0\cdot \to\to S$ es una cadena de signos de $\mathcal{L}_a$ (¡atención a las comillas!).

Si preferimos pensar que $\mathcal{L}_a$ es en realidad la versión que hemos definido tomando números naturales como signos, entonces esta cadena de signos no es sino la sucesión finita de números naturales

$$10, 4, 1, 1, 9, 11, 11, 6$$

■

Naturalmente, las cadenas de signos del estilo de la que acabamos de considerar no sirven para nada. Ahora hemos de extraer de entre ellas las que tienen “significado”, algo así como seleccionar los movimientos “legales” en el ajedrez de entre todos los movimientos posibles. Hay dos casos distintos en los que una cadena de signos puede tener un significado: bien porque nombre a un objeto, bien porque afirme algo. A las cadenas que nombran objetos las llamaremos términos, mientras que a las que afirman algo las llamaremos fórmulas. Ésta es la idea subyacente, pero no nos sirve como definición porque, además de ser imprecisa, alude a un posible significado de las cadenas de signos, y queremos que la definición sea formal. La definición correcta es la siguiente:

Definición 1.5 Una cadena de signos θ de un lenguaje formal $\mathcal{L}$ es una *expresión* de $\mathcal{L}$ si existe una sucesión finita $\theta_0, \ldots, \theta_m$ de cadenas de signos de $\mathcal{L}$ de modo que $\theta_m \equiv \theta$ y cada θ_k cumple una de las condiciones siguientes:

1. θ_k es una variable.
2. θ_k es una constante.
3. $\theta_k \equiv Rt_1\cdots t_n$, donde R es un relator n-ádico de $\mathcal{L}$ y los t_i son cadenas anteriores de la sucesión cuyo primer signo es una variable, una constante, un funtor o el descriptor.
4. $\theta_k \equiv ft_1\cdots t_n$, donde f es un funtor n-ádico de $\mathcal{L}$ y los t_i son cadenas anteriores de la sucesión cuyo primer signo es una variable, una constante, un funtor o el descriptor.
5. $\theta_k \equiv \neg\alpha$, donde α es una cadena anterior de la sucesión cuyo primer signo es un relator, un conector o el generalizador.
6. $\theta_k \equiv \to\alpha\beta$, donde α y β son cadenas anteriores de la sucesión cuyo primer signo es un relator, un conector o el generalizador.
7. $\theta_k \equiv \bigwedge x\,\alpha$, donde x es una variable y α es una cadena anterior de la sucesión cuyo primer signo es un relator, un conector o el generalizador.
8. $\theta_k \equiv |x\alpha$, donde x es una variable y α es una cadena anterior de la sucesión cuyo primer signo es un relator, un conector o el generalizador.

Diremos que una expresión es un *término* si su primer signo es una variable, una constante, un funtor o el descriptor. Diremos que una expresión es una *fórmula* si su primer signo es un relator, un conector o el generalizador.

Observemos que si una sucesión $\theta_0, \ldots, \theta_m$ cumple los requisitos de la definición anterior, éstos sólo imponen condiciones a cada cadena que dependen de las cadenas que la preceden, luego si eliminamos las últimas cadenas de la sucesión obtenemos una nueva sucesión que sigue cumpliendo las mismas condiciones. Concluimos que toda cadena de la sucesión es una expresión, y no sólo la última, como indica la definición. Teniendo esto en cuenta, así como la definición de términos y fórmulas, la definición de expresión puede simplificarse como sigue:

Una cadena de signos θ de un lenguaje formal $\mathcal{L}$ es una *expresión* de $\mathcal{L}$ si existe una sucesión finita $\theta_0, \ldots, \theta_m$ de cadenas de signos de $\mathcal{L}$ (que serán todas expresiones) de modo que $\theta_m \equiv \theta$ y cada θ_k cumple una de las condiciones siguientes:

1. θ_k es una variable.
2. θ_k es una constante.
3. $\theta_k \equiv Rt_1 \cdots t_n$, donde R es un relator n-ádico de $\mathcal{L}$ y los t_i son términos anteriores en la sucesión.
4. $\theta_k \equiv ft_1 \cdots t_n$, donde f es un funtor n-ádico de $\mathcal{L}$ y los t_i son términos anteriores en la sucesión.
5. $\theta_k \equiv \neg\alpha$, donde α es una fórmula anterior en la sucesión.
6. $\theta_k \equiv \to\alpha\beta$, donde α y β son fórmulas anteriores en la sucesión.
7. $\theta_k \equiv \bigwedge x\, \alpha$, donde x es una variable y α es una fórmula anterior en la sucesión.
8. $\theta_k \equiv |x\alpha$, donde x es una variable y α es una fórmula anterior en la sucesión.

Más claramente, lo que dice esta definición es que las expresiones pueden construirse unas a partir de otras mediante unas reglas muy concretas:

1. x_i es un término.
2. c_i es un término.
3. $R_i^n t_1 \cdots t_n$ es una fórmula.
4. $f_i^n t_1 \cdots t_n$ es un término.
5. $\neg\alpha$ es una fórmula.
6. $\to\alpha\beta$ es una fórmula.
7. $\bigwedge x_i\, \alpha$ es una fórmula.
8. $|x_i\alpha$ es un término (si es que $\mathcal{L}$ tiene descriptor).

Aquí hemos adoptado un convenio de notación que nos simplificará bastante el enunciado de reglas como éstas. Cuando digamos "x_i es un término" queremos decir que "toda variable es un término", en general, usaremos la letra t (con subíndices si hace falta) para referirnos a términos arbitrarios y las letras α, β, γ para referirnos a fórmulas, de modo que el apartado 6) debe leerse como "si α y β son fórmulas, entonces $\to \alpha\beta$ es una fórmula".

Vemos entonces que toda expresión tiene que estar exactamente en uno de los ocho casos anteriores: si empieza por una variable es una variable, y además es un término, si empieza por una constante es una constante y además es un término, si empieza por un relator es de la forma $R_i^n t_1 \cdots t_n$ y es una fórmula, etc. Las fórmulas sin descriptores de tipo $R_i^n t_1 \cdots t_n$ se llaman *atómicas.*

Convenios de notación Aunque el convenio por defecto para nombrar una cadena de signos es nombrar sus signos en el orden en que aparecen en ella, en la práctica adoptaremos otros convenios particulares que nos resulten más cómodos. Por ejemplo, en lugar de escribir $\to \alpha\beta$ escribiremos $(\alpha \to \beta)$. Hay que entender que con esto no estamos "desordenando" los signos de la fórmula, sino simplemente cambiando la forma de nombrarla. El primer signo de $(\alpha \to \beta)$ no es un paréntesis, ni el primer signo de α, sino $\to$. Con esta notación es necesario introducir paréntesis, pues una expresión castellana como $\alpha \to \beta \to \gamma$ es ambigua, ya que no está claro si nombra a

$$(\alpha \to \beta) \to \gamma \equiv \to\to \alpha\beta\gamma, \qquad \text{o bien a} \qquad \alpha \to (\beta \to \gamma) \equiv \to \alpha \to \beta\gamma.$$

No obstante, no siempre harán falta los paréntesis, y cuando se puedan suprimir sin ambigüedad así lo haremos. No necesitamos precisar cuándo ponemos y quitamos paréntesis porque esto no tiene nada que ver con los signos o las expresiones de un lenguaje formal, sino meramente con los convenios que adoptamos para mencionarlos. Con tal de que sepamos sin lugar a dudas qué expresión concreta nombra cada nombre que usemos no necesitamos explicar cómo lo sabemos.

Similarmente, escribiremos

$$(x|\alpha) \equiv |x\alpha, \quad (t_1 = t_2) \equiv = t_1 t_2, \quad (t_1 \neq t_2) \equiv \neg(t_1 = t_2)$$

y cuando tengamos que escribir varios cuantificadores seguidos $\bigwedge x \bigwedge y \bigwedge z$ escribiremos más brevemente $\bigwedge xyz$.

En lenguajes concretos también podemos adoptar convenios concretos. Por ejemplo, en el lenguaje de la aritmética usaremos los convenios de notación

$$t' \equiv St, \quad (t_1 + t_2) \equiv +t_1 t_2, \quad (t_1 t_2) \equiv (t_1 \cdot t_2) \equiv \cdot t_1 t_2.$$

Nuevamente, los paréntesis son necesarios, pues

$$(t_1 + t_2) + t_3 \equiv + + t_1 t_2 t_3, \qquad t_1 + (t_2 + t_3) \equiv + t_1 + t_2 t_3.$$

■

Ejemplo $\alpha \equiv \bigwedge xy(x + y' = (x + y)')$ es una fórmula del lenguaje de la aritmética $\mathcal{L}_a$.

Para justificarlo consideramos la sucesión siguiente de cadenas de signos:

1. x
2. y
3. y'
4. $x + y'$
5. $x + y$
6. $(x + y)'$
7. $x + y' = (x + y)'$
8. $\bigwedge y(x + y' = (x + y)')$
9. $\bigwedge xy(x + y' = (x + y)')$

Es fácil ver que cumple la definición de expresión: la primera es un término porque es una variable, la segunda también, la tercera es un término por ser de la forma $y' \equiv Sy$, es decir, un funtor monádico seguido de un término que ya ha aparecido en la sucesión, la cuarta es un término por ser de la forma $x + y' \equiv +xy'$, es decir, un funtor diádico seguido de dos términos previos. Lo mismo vale para la quinta. La sexta es $(x+y)' \equiv S(x+y)$, un funtor monádico seguido de un término previo, luego es un término. La octava es de la forma $\bigwedge x\beta$, donde β es una fórmula previa, y lo mismo vale para la novena.

Si queremos escribir la fórmula enumerando sus signos en el orden en que aparecen en ella, el resultado es

$$\bigwedge x \bigwedge y = +xSyS + xy, \qquad \text{o también} \qquad 12, 1, 12, 2, 5, 7, 1, 6, 2, 6, 7, 1, 2,$$

si entendemos que los signos de $\mathcal{L}_a$ son los números naturales que hemos convenido. Vemos que su longitud es 13. Obviamente, nunca escribiremos las fórmulas así. ∎

La fórmula del ejemplo anterior es ciertamente una fórmula porque está construida mediante las reglas formales que hemos establecido, tal y como acabamos de comprobar. Dichas reglas son formales porque sólo hacen referencia a la forma de las expresiones: con expresiones de tal forma podemos formar otra expresión de tal otra forma, sin entrar para nada en el posible significado de los signos. Sin embargo, cualquiera ve a simple vista que se trata de una fórmula, y no lo hace construyendo una sucesión de expresiones de acuerdo con la definición, sino que uno se da cuenta de que es una fórmula porque si intenta "leerla" ve que puede hacerlo, que tiene sentido.

Concretamente, si interpretamos los signos del lenguaje de la aritmética en su modelo natural, "vemos" que la fórmula anterior significa que la suma de un número natural arbitrario con el siguiente de otro es igual al siguiente de la suma. Esto es precisamente la versión informal de uno de los axiomas de Peano (extendidos para incorporar la suma y el producto), que queda así formalizado por la fórmula del ejemplo. Decimos, pues, que reconocemos que la

cadena del ejemplo anterior es una fórmula porque tiene sentido, y es que, en efecto, hemos definido las fórmulas pensando en que sean cadenas susceptibles de ser "leídas" como afirmaciones. Similarmente, los términos son las cadenas de signos susceptibles de ser "leídas" como nombres de objetos, cosa que también se reconoce a simple vista. Seguidamente pasamos a precisar estas ideas, es decir, vamos a definir el significado de una expresión de un lenguaje en un modelo.

Para ello tenemos un inconveniente, y es que la definición de modelo no asigna ningún significado a las variables del lenguaje. Ello nos lleva a la definición de valoración:

Definición 1.6 Una *valoración* de un lenguaje formal $\mathcal{L}$ en un modelo M es un criterio v que asigna a cada variable x de $\mathcal{L}$ un objeto $v(x)$ del universo de M.

Así, al considerar un mismo modelo con varias valoraciones estamos permitiendo que las variables varíen de significado sin alterar el modelo.

Si v es una valoración de un lenguaje formal $\mathcal{L}$ en un modelo M, a es un elemento del universo U de M y x es una variable de $\mathcal{L}$, llamaremos v_x^a a la valoración en M dada por

$$v_x^a(y) \equiv \begin{cases} a & \text{si } y \equiv x, \\ v(y) & \text{si } y \not\equiv x. \end{cases}$$

Llamaremos v_{xy}^{ab} a $(v_x^a)_y^b$, llamaremos v_{xyz}^{abc} a $((v_x^a)_y^b)_z^c$, etc.

Es claro que si $x \equiv y$ entonces v_{xy}^{ab} coincide con v_y^b, mientras que si $x \not\equiv y$, entonces v_{xy}^{ab} coincide con v_{yx}^{ba}.

Ahora ya podemos definir el significado de una expresión arbitraria θ de un lenguaje $\mathcal{L}$ respecto a un modelo M. Notemos que si θ es un término su significado ha de ser un objeto del universo de M, mientras que si θ es una fórmula su significado ha de ser un valor de verdad: ha de ser verdadera o falsa.

Definición 1.7 Sea v una valoración de un lenguaje formal $\mathcal{L}$ en un modelo M. Las condiciones siguientes determinan cuándo M *satisface* la fórmula α respecto a la valoración v (abreviado $M \vDash \alpha[v]$) y cuándo un término t *denota* en M respecto a la valoración v a un objeto que representaremos por $\bar{t}$ o $M(t)[v]$:

1. $M(x_i)[v] \equiv v(x_i)$,
2. $M(c_i)[v] \equiv \bar{c}_i$,
3. $M \vDash R_i^n t_1 \cdots t_n[v]$ syss [si y sólo si] $\bar{R}_i^n(\bar{t}_1, \ldots, \bar{t}_n)$,
4. $M(f_i^n t_1 \cdots t_n)[v] \equiv \bar{f}_i^n(\bar{t}_1, \ldots, \bar{t}_n)$,
5. $M \vDash \neg\alpha[v]$ syss no $M \vDash \alpha[v]$,
6. $M \vDash (\alpha \to \beta)[v]$ syss no $M \vDash \alpha[v]$ o $M \vDash \beta[v]$,
7. $M \vDash \bigwedge x_i \alpha[v]$ syss para todo objeto a de M se cumple que $M \vDash \alpha[v_{x_i}^a]$,

8. Si $\mathcal{L}$ tiene descriptor y d es la descripción impropia de M,

$$M(x_i|\alpha)[v] \equiv \begin{cases} \text{el único } a \text{ de } M \text{ tal que } M \vDash \alpha[v^a_{x_i}] & \text{si existe tal } a, \\ d & \text{en otro caso.} \end{cases}$$

Vamos a parafrasear la definición anterior:

1. El objeto denotado por una variable es el que le asigna la valoración considerada.
2. El objeto denotado por una constante es el que le asigna el modelo.
3. Una fórmula de tipo $R^n_i t_1 \cdots t_n$ es satisfecha si los objetos denotados por los términos satisfacen la relación que el modelo asocia al relator.
4. El objeto denotado por un término $f^n_i t_1 \cdots t_n$ es el que resulta de aplicar a los objetos denotados por los términos la función asociada al funtor por el modelo.
5. Una fórmula $\neg\alpha$ es satisfecha si y sólo si α no es satisfecha.
6. Una fórmula $\alpha \to \beta$ es satisfecha cuando α no es satisfecha o β es satisfecha, es decir, siempre salvo si α es satisfecha pero β no.
7. Una fórmula de tipo $\bigwedge x_i \alpha$ es satisfecha si α es satisfecha cualquiera que sea la interpretación de la variable x_i, es decir, para toda interpretación posible de x_i en el universo del modelo.
8. El objeto denotado por una descripción $x_i|\alpha$ es el único objeto a del modelo con el que α es satisfecha si la variable x_i se interpreta como a, si es que existe, o la descripción impropia del modelo si no hay una única forma de interpretar la variable x_i para que α sea satisfecha (sea porque no haya ninguna o porque haya varias).

Esta definición requiere varias reflexiones:

— En primer lugar, vemos que es aquí donde por primera vez establecemos que el negador $\neg$ debe interpretarse siempre como "no" y el implicador $\to$ debe interpretarse como "si... entonces" en el sentido que ya hemos discutido, así como que $\bigwedge x$ significa "para todo x" y $x|$ significa "el x tal que", con el convenio ya explicado para las descripciones impropias.

— En segundo lugar, esta definición no hace sino especificar lo que todos sabemos hacer instintivamente al leer una sentencia formal. Por ejemplo, si consideramos la sentencia

$$\bigwedge xy\ x \cdot x \neq 0'' \cdot (y \cdot y) + 0'''$$

del lenguaje de la aritmética, sin necesidad de definición alguna todos entendemos que ahí dice que no existen números naturales m y n tales que $m^2 - 2n^2 = 3$. Vamos a ver que si aplicamos la definición anterior con el modelo natural llegamos a la misma conclusión. Fijemos una valoración v arbitraria.

- Por la definición 2) tenemos que $\bar{0}$ es el número natural 0.
- Por la definición 4) tenemos que $\overline{0'}$ es el siguiente de $\bar{0}$, o sea, el 1. Similarmente, $\overline{0''} \equiv 2$ y $\overline{0'''} \equiv 3$.
- Aplicando varias veces 1) y 4) concluimos que

$$\overline{x \cdot x} \equiv v(x)^2, \qquad \overline{(0'' \cdot (y \cdot y) + 0''')} \equiv 2 \cdot v(y)^2 + 3.$$

- Por 3) y 5),

$$M \vDash \big(x \cdot x \neq 0'' \cdot (y \cdot y) + 0'''\big)[v] \quad \text{syss} \quad v(x)^2 - 2 \cdot v(y)^2 \not\equiv 3.$$

- En particular, si m y n son dos números naturales cualesquiera

$$M \vDash \big(x \cdot x \neq 0'' \cdot (y \cdot y) + 0'''\big)[v^{mn}_{x\,y}] \quad \text{syss} \quad m^2 - 2 \cdot n^2 \not\equiv 3.$$

- Aplicando dos veces 7) concluimos que $M \vDash \bigwedge xy\ x \cdot x \neq 0'' \cdot (y \cdot y) + 0'''[v]$ syss para todos los números naturales m y n, se cumple $m^2 - 2 \cdot n^2 \not\equiv 3$.

Así pues, en ejemplos concretos nunca necesitaremos aplicar explícitamente la definición de denotación y satisfacción, pues hacerlo nos lleva por un camino más largo al mismo sitio al que llegamos si simplemente "leemos" la expresión dada.

— Otro hecho que hay que destacar es que la definición anterior nos dice qué es el objeto denotado por un término y qué significa que una fórmula sea satisfecha, pero no nos proporciona un algoritmo para decidir si se da o no el caso. Por ejemplo, acabamos de ver que

$$M \vDash \bigwedge xy\ x \cdot x \neq 0'' \cdot (y \cdot y) + 0'''[v]$$

significa que no existen números naturales m y n tales que $m^2 - 2n^2 = 3$. Eso es lo que obtenemos al aplicar la definición de satisfacción, pero un problema muy distinto es decidir si existen o no tales números naturales. Una cosa es saber lo que significa una fórmula y, en particular, constatar que efectivamente significa algo, y otra muy distinta saber si lo que significa es cierto o es falso. Dicho de otro modo: podemos hablar de la satisfacción o no satisfacción de una fórmula con independencia de cualquier razonamiento (formal o informal) que nos convenza de la verdad o falsedad de dicha fórmula. Sabemos lo que significa que no existen números naturales tales que $m^2 - 2n^2 = 3$ con independencia de si sabemos probar que existen o que no existen. Si no fuera así, el modelo M no estaría bien definido.

Más convenios de notación Ahora estamos en condiciones de definir razonadamente los signos lógicos que hemos omitido en la definición de lenguaje formal. Introducimos los convenios de notación siguientes:

1. $(\alpha \vee \beta) \equiv (\neg\alpha \to \beta)$
2. $(\alpha \wedge \beta) \equiv \neg(\neg\alpha \vee \neg\beta)$
3. $(\alpha \leftrightarrow \beta) \equiv \big((\alpha \to \beta) \wedge (\beta \to \alpha)\big)$
4. $\bigvee x_i\, \alpha \equiv \neg\bigwedge x_i \neg\alpha$

Aquí es fundamental comprender que estos convenios no son arbitrarios. Por el contrario, de acuerdo con la definición de satisfacción, podemos demostrar que si M es un modelo de un lenguaje $\mathcal{L}$ y v es una valoración, entonces

$M \vDash (\alpha \vee \beta)[v]$ syss $M \vDash \alpha[v]$ o $M \vDash \beta[v]$,

$M \vDash (\alpha \wedge \beta)[v]$ syss $M \vDash \alpha[v]$ y $M \vDash \beta[v]$,

$M \vDash (\alpha \leftrightarrow \beta)[v]$ syss $M \vDash \alpha[v]$ y $M \vDash \beta[v]$ o por el contrario no $M \vDash \alpha[v]$ y no $M \vDash \beta[v]$,

$M \vDash \bigvee x_i \alpha[v]$ syss existe un objeto a de M tal que $M \vDash \alpha[v^a_{x_i}]$,

En efecto, para probar la primera aplicamos la definición de satisfacción: $M \vDash (\alpha \vee \beta)[v]$ es lo mismo que $M \vDash (\neg\alpha \rightarrow \beta)[v]$, que se cumple cuando no $M \vDash \neg\alpha[v]$ o bien $M \vDash \beta[v]$, y esto equivale a $M \vDash \alpha[v]$ o $M \vDash \beta[v]$.

Así pues, la definición que hemos dado de la disyunción garantiza que su interpretación va a ser siempre la dada por la tabla de verdad que corresponde a la "o" no exclusiva. Similarmente se demuestra que las tablas de verdad (y, por tanto, los significados) de $\wedge$ y $\leftrightarrow$ se corresponden con las que habíamos indicado en la página 8. Veamos por último que el cuantificador existencial que hemos definido realmente significa "existe".

En efecto: $M \vDash \bigvee x_i \alpha[v]$ es lo mismo que $M \vDash \neg\bigwedge x_i \neg\alpha[v]$, y esto se cumple cuando no se cumple $M \vDash \bigwedge x_i \neg\alpha[v]$, es decir, cuando no es cierto que para todo a en M se cumple $M \vDash \neg\alpha[v^a_{x_i}]$. A su vez, esto es lo mismo que decir que existe un a en M para el que no se cumple $M \vDash \neg\alpha[v^a_{x_i}]$, que a su vez equivale a que exista un a en M tal que $M \vDash \alpha[v^a_{x_i}]$.

Así pues, los convenios precedentes, que desde un punto de vista formal son aparentemente arbitrarios, no lo son en absoluto, pues desde un punto de vista semántico se demuestra que son lo que tienen que ser para que los signos correspondientes signifiquen lo que queremos que signifiquen.

Terminamos con la siguiente observación: si $\mathcal{L}$ es un lenguaje con descriptor y d es la descripción impropia fijada por un modelo M, entonces se cumple que $M(x|x = x)[v] \equiv d$.

En efecto, si a es cualquier objeto de M, tenemos que $M \vDash (x = x)[v^a_x]$ equivale a $a \equiv a$, lo cual se cumple siempre, es decir, se cumple para todo a. Entonces tenemos dos posibilidades: si el universo de M tiene más de un objeto, entonces la descripción $x|x = x$ es impropia, pues $M \vDash (x = x)[v^a_x]$ no lo cumple un único objeto de M, sino que lo cumplen todos, y entonces $M(x|x = x)[v] \equiv d$, por definición (d es la interpretación de todas las descripciones impropias). Por el contrario, si M tiene un único objeto, éste tiene que ser tanto d como el objeto denotado por $x|x = x$, luego también se da la igualdad.

Así pues, el término $x|x = x$ es una forma de referirnos formalmente a *la* descripción impropia, de modo que toda descripción impropia denotará en cualquier modelo al mismo objeto que $x|x = x$.

1.4 Variables libres y ligadas

Consideremos una fórmula como $\bigvee y\ x = y'$, del lenguaje de la aritmética. Diremos que en ella la variable y está ligada, porque está bajo el alcance del cuantificador $\bigvee y$, mientras que la variable x está libre porque no está afectada por ningún cuantificador ni descriptor. Una definición precisa y operativa de estos conceptos es la siguiente:

Definición 1.8 Sea $\mathcal{L}$ un lenguaje formal. Diremos que una variable x está *libre* en una expresión de $\mathcal{L}$ si se puede probar que lo está a partir de las reglas siguientes:

1. x está libre en x_i syss $x \equiv x_i$.
2. x nunca está libre en c_i.
3. x está libre en $R_i^n t_1 \cdots t_n$ syss lo está en algún t_j.
4. x está libre en $f_i^n t_1 \cdots t_n$ syss lo está en algún t_j.
5. x está libre en $\neg\alpha$ syss lo está en α.
6. x está libre en $\alpha \to \beta$ syss lo está en α o en β.
7. x está libre en $\bigwedge x_i \alpha$ syss lo está en α y $x \not\equiv x_i$.
8. x está libre en $x_i|\alpha$ syss lo está en α y $x \not\equiv x_i$ (en el caso en que $\mathcal{L}$ tenga descriptor).

Diremos que una variable x está *ligada* en una expresión de $\mathcal{L}$ si se puede probar que lo está a partir de las reglas siguientes:

1. x nunca está ligada en x_i.
2. x nunca está ligada en c_i.
3. x está ligada en $R_i^n t_1 \cdots t_n$ syss lo está en algún t_j.
4. x está ligada en $f_i^n t_1 \cdots t_n$ syss lo está en algún t_j.
5. x está ligada en $\neg\alpha$ syss lo está en α.
6. x está ligada en $\alpha \to \beta$ syss lo está en α o en β.
7. x está ligada en $\bigwedge x_i \alpha$ syss lo está en α o $x \equiv x_i$.
8. x está ligada en $x_i|\alpha$ syss lo está en α o $x \equiv x_i$ (en el caso en que $\mathcal{L}$ tenga descriptor).

El "si se puede probar que lo está" puede parecer ambiguo, pero lo que sucede es que, dada una expresión θ, siempre podemos tomar una sucesión de expresiones $\theta_0, \ldots, \theta_m \equiv \theta$ de acuerdo con la definición de expresión, y las dos definiciones precedentes nos permiten saber si una variable dada está libre o ligada en cada expresión θ_i a partir de si lo está o no en las expresiones precedentes, por lo que tenemos un algoritmo que siempre decide en un número finito de pasos cuál es el caso.

Observaciones Una variable x está libre o ligada en $\alpha \vee \beta$, $\alpha \wedge \beta$ o $\alpha \leftrightarrow \beta$ syss lo está en α o en β. Así mismo, x está libre en $\bigvee x_i \alpha$ syss lo está en α y $x \not\equiv x_i$, y x está ligada en $\bigvee x_i \alpha$ syss lo está en α o $x \equiv x_i$. Estas observaciones no forman parte de la definición de variable libre y ligada, sino que se deducen inmediatamente de las definiciones de $\alpha \vee \beta$, etc.

Una variable está en una expresión θ (es decir, es uno de los signos que componen θ) si y sólo si está libre o ligada en θ.

Si θ es una expresión sin descriptores de un lenguaje $\mathcal{L}$ con descriptor, entonces una variable x está libre o ligada en θ considerada como expresión de $\mathcal{L}$ si y sólo si lo está considerada como expresión de $\underline{\mathcal{L}}$. ∎

Ejemplos Observemos que una variable puede estar a la vez libre y ligada en una expresión, así como no estar ni libre ni ligada. Los ejemplos siguientes muestran las cuatro posibilidades para una misma variable x:

$u = v$	x no está ni libre ni ligada.
$u = x$	x está libre y no ligada.
$\bigvee x\ u = x$	x está ligada y no libre.
$x = 0 \wedge \bigvee x\ x = 0'$	x está libre y ligada.

(Suponemos que las variables x, y, u, v son distintas). ∎

Una expresión es *abierta* si tiene variables libres. En caso contrario es *cerrada.* Un *designador* es un término cerrado. Una *sentencia* es una fórmula cerrada. Por lo tanto las cadenas de signos quedan clasificadas como sigue:

- cadenas de signos
  - expresiones
    - términos
      - designadores
      - términos abiertos
    - fórmulas
      - sentencias
      - fórmulas abiertas
  - no expresivas

La distinción entre variables libres y ligadas tiene una interpretación semántica. Consideremos, por ejemplo, el lenguaje de la aritmética y su modelo natural M. Se cumple

$$M \vDash (\bigvee y\ 0''' \cdot y = x)[v],$$

si y sólo si $v(x)$ es múltiplo de 3. Así pues, para saber si la fórmula $\bigvee y\ 0''' \cdot y = x$ es satisfecha o no por una valoración sólo necesitamos saber cómo actúa la valoración sobre la variable x. Su valor sobre las demás variables es irrelevante. Esto no es casual:

Teorema 1.9 *Si v y w son valoraciones de un lenguaje formal $\mathcal{L}$ en un modelo M que coinciden sobre las variables libres de una expresión θ, entonces si θ es un término $M(\theta)[v] \equiv M(\theta)[w]$ y si θ es una fórmula $M \vDash \theta[v]$ syss $M \vDash \theta[w]$.*

DEMOSTRACIÓN: Por inducción sobre la longitud de θ.

Si $\theta \equiv x$ entonces x está libre en θ, luego $M(\theta)[v] \equiv v(x) \equiv w(x) \equiv M(\theta)[w]$.

Si $\theta \equiv c$ entonces $M(\theta)[v] \equiv M(c) \equiv M(\theta)[w]$.

Si $\theta \equiv R_i^n t_1 \cdots t_n$, por hipótesis de inducción $M(t_j)[v] \equiv M(t_j)[w]$. Llamamos $\bar{t}_j$ a este objeto. Entonces

$$M \vDash \theta[v] \text{ syss } \bar{R}_i^n(\bar{t}_1, \ldots, \bar{t}_n) \text{ syss } M \vDash \theta[w].$$

Si $\theta \equiv f_i^n t_1 \cdots t_n$, por hipótesis de inducción $M(t_j)[v] \equiv M(t_j)[w]$. Llamando como antes $\bar{t}_j$ a este objeto,

$$M(\theta)[v] \equiv \bar{f}_i^n(\bar{t}_1, \ldots, \bar{t}_n) \equiv M(\theta)[w].$$

Si $\theta \equiv \neg\alpha$, por hipótesis de inducción $M \vDash \alpha[v]$ syss $M \vDash \alpha[w]$, luego no $M \vDash \alpha[v]$ syss no $M \vDash \alpha[w]$, o sea $M \vDash \neg\alpha[v]$ syss $M \vDash \neg\alpha[w]$.

Si $\theta \equiv \alpha \to \beta$, por hipótesis de inducción $M \vDash \alpha[v]$ syss $M \vDash \alpha[w]$ y $M \vDash \beta[v]$ syss $M \vDash \beta[w]$. Entonces $M \vDash (\alpha \to \beta)[v]$ syss no $M \vDash \alpha[v]$ o $M \vDash \beta[v]$ syss no $M \vDash \alpha[w]$ o $M \vDash \beta[w]$ syss $M \vDash (\alpha \to \beta)[w]$.

Si $\theta \equiv \bigwedge x\alpha$, sea a un objeto de M. Si y está libre en α entonces y está libre en θ o $y \equiv x$. En cualquier caso $v_x^a(y) \equiv w_x^a(y)$, luego v_x^a y w_x^a coinciden en las variables libres de α. Por hipótesis de inducción $M \vDash \alpha[v_x^a]$ syss $M \vDash \alpha[w_x^a]$ para todo objeto a de M.

En particular $M \vDash \alpha[v_x^a]$ para todo objeto a de M syss $M \vDash \alpha[w_x^a]$ para todo objeto a de M, es decir, $M \vDash \bigwedge x\alpha[v]$ syss $M \vDash \bigwedge x\alpha[w]$.

Si $\theta \equiv x|\alpha$, razonando como antes, para todo objeto a de M tenemos que $M \vDash \alpha[v_x^a]$ syss $M \vDash \alpha[w_x^a]$. Por lo tanto hay un único a en M tal que $M \vDash \alpha[v_x^a]$ syss hay un único a en M tal que $M \vDash \alpha[w_x^a]$, además en tal caso son el mismo objeto a.

Si se da la unicidad $M(\theta)[v] \equiv a \equiv M(\theta)[w]$. Si no se da la unicidad $M(\theta)[v] \equiv d \equiv M(\theta)[w]$. ∎

Así pues, la interpretación de una expresión en un modelo respecto de una valoración depende únicamente de cómo actúa ésta sobre las variables libres en la expresión.

Como aplicación de este teorema podemos justificar que la definición siguiente es razonable:

$$\overset{1}{\bigvee} x_1 \cdots x_n \alpha \equiv \bigvee y_1 \cdots y_n \bigwedge x_1 \cdots x_n (\alpha \leftrightarrow y_1 = x_1 \wedge \cdots \wedge y_n = x_n),$$

donde $y_1, \ldots, y_n$ son las variables de menor índice que no están en $\bigvee x_1 \cdots x_n \alpha$. Vamos a probar que $\overset{1}{\bigvee} x_1 \cdots x_n$ así definido se interpreta en cualquier modelo como "existen unos únicos $x_1, \ldots, x_n$ tales que ..."

Teorema 1.10 *Si M es un modelo de un lenguaje formal $\mathcal{L}$ y v es una valoración en M, entonces, para toda fórmula α de $\mathcal{L}$, $M \vDash \bigvee^1 x_1 \cdots x_n \alpha[v]$ syss existen unos únicos $a_1, \ldots, a_n$ en M tal que $M \vDash \alpha[v^{a_1 \cdots a_n}_{x_1 \cdots x_n}]$.*

DEMOSTRACIÓN: Tenemos que $M \vDash \bigvee^1 x_1 \cdots x_n \alpha[v]$ es lo mismo que

$$M \vDash \bigvee y_1 \cdots y_n \bigwedge x_1 \cdots x_n (\alpha \leftrightarrow y_1 = x_1 \wedge \cdots \wedge y_n = x_n)[v],$$

donde las variables y_i son distintas de las x_i y no están en α. Esto a su vez equivale a que existen $a_1, \ldots, a_n$ en M tales que

$$M \vDash \bigwedge x_1 \cdots x_n (\alpha \leftrightarrow y_1 = x_1 \wedge \cdots \wedge y_n = x_n)[v^{a_1 \cdots a_n}_{y_1 \cdots y_n}],$$

que a su vez equivale a que existen $a_1, \ldots, a_n$ en M tales que, para todos los $b_1, \ldots, b_n$ de M, se cumple

$$M \vDash (\alpha \leftrightarrow y_1 = x_1 \wedge \cdots \wedge y_n = x_n)[v^{a_1 \cdots a_n b_1 \cdots b_n}_{y_1 \cdots y_n x_1 \cdots x_n}],$$

que equivale a $M \vDash \alpha[v^{a_1 \cdots a_n b_1 \cdots b_n}_{y_1 \cdots y_n x_1 \cdots x_n}]$ syss $a_1 \equiv b_1, \ldots, a_n \equiv b_n$. Ahora aplicamos el teorema anterior, que nos dice que, como las variables y_i no están en α, el valor de $v^{a_1 \cdots a_n b_1 \cdots b_n}_{y_1 \cdots y_n x_1 \cdots x_n}$ sobre las variables y_i es irrelevante. En total tenemos, pues, que existen $a_1, \ldots, a_n$ en M que son los únicos elementos $b_1, \ldots, b_n$ de M que cumplen $M \vDash \alpha[v^{b_1 \cdots b_n}_{x_1 \cdots x_n}]$. ∎

1.5 Sustitución

Nos falta discutir un último aspecto sobre los lenguajes formales antes de poder usarlos para formalizar razonamientos. Se trata del concepto de sustitución. Es un concepto que los matemáticos usan cotidiana e instintivamente. Por ejemplo, cuando un matemático escribe

$$A = \{x \mid \alpha(x)\}$$

y luego tiene que $3 \in A$, de ahí deduce $\alpha(3)$. Al escribir esto ha sustituido la variable x por el término 3 en la fórmula α. Parece una operación trivial que apenas requiere un comentario para ser definida, pero no es exactamente así, como vamos a ver ahora. En esencia es ciertamente trivial, pero hay una cuestión técnica que debemos tener presente a la hora de tratar con sustituciones.

Para empezar, a la hora de tratar teóricamente con la sustitución, es preferible usar una notación más explícita, en la que se vean todos los elementos involucrados. En lugar de escribir $\alpha(x)$ y $\alpha(3)$ escribiremos α (porque una fórmula puede tener varias variables libres, y cualquiera de ellas puede ser sustituida en cualquier momento por un término) y escribiremos $\mathtt{S}^3_x\alpha$ para la sustitución (de modo que se indica claramente que estamos sustituyendo la variable x por el término 3 en la fórmula α). El objetivo de esta sección es dar una definición precisa de $\mathtt{S}^t_x\theta$ donde θ es una expresión, x es una variable y t es un término.

Como siempre en estos contextos, la idea subyacente al concepto que queremos definir se expresa de forma natural en términos semánticos, es decir, en términos de lo que pretendemos que signifique lo que queremos definir, supuesto que hayamos fijado un modelo para el lenguaje formal considerado, pero al final daremos una definición puramente sintáctica (formal) que no haga referencia a ningún modelo posible, pero de tal modo que cuando se fije un modelo se corresponda con lo que pretendíamos.

En términos semánticos, el propósito de la sustitución es el siguiente:

> Si θ es un término, una sustitución $\mathtt{S}^t_x\theta$ debe ser otro término con la propiedad de que el objeto que denote en un modelo dado sea el objeto que denota el término θ cuando la variable x se interpreta como el objeto denotado por el término t.
>
> Si θ es una fórmula, una sustitución $\mathtt{S}^t_x\theta$ debe ser otra fórmula que sea satisfecha en un modelo si θ es satisfecha cuando la variable x se interpreta como el objeto denotado por t.

Veamos un ejemplo concreto con el lenguaje de la aritmética (y su modelo natural):

Si $\theta \equiv x'$, se trata de un término que significa "el siguiente del número natural x". El objeto que denota en el modelo natural depende de la valoración v que consideremos, pues será el siguiente del número natural $v(x)$.

¿Qué queremos que sea $\mathtt{S}^{0''}_x x'$? Queremos que sea un término que signifique lo mismo que θ, es decir, "el siguiente de x", pero cuando la variable x se interpreta concretamente como el objeto denotado por $0''$, es decir, como el número natural 2. Así pues, $\mathtt{S}^{0''}_x x'$ debe denotar al siguiente del número natural 2, es decir, al 3. Por lo tanto, lo natural es tomar $\mathtt{S}^{0''}_x x' \equiv 0'''$, que es simplemente lo que resulta de quitar la variable x y poner en su lugar el término $0''$, es decir, lo que comúnmente se entiende por "sustituir".

Veamos ahora otro ejemplo con la fórmula

$$\alpha \equiv \bigwedge y(\bigvee z\ x = y \cdot z \wedge y \neq 0' \to \bigvee w\ y = w \cdot 0'').$$

Esta fórmula dice que todo divisor de x distinto de 1 es par. Más precisamente, esta fórmula es satisfecha en el modelo natural del lenguaje de la aritmética si y sólo si el objeto denotado por x, es decir, $v(x)$, tiene todos sus divisores no triviales pares o, equivalentemente, si y sólo si $v(x)$ es potencia de dos.

¿Qué queremos que sea $\mathtt{S}^{x'''}_x\alpha$? Tiene que ser una fórmula que ya no sea satisfecha cuando $v(x)$ sea potencia de dos, sino cuando el objeto denotado por x''' sea potencia de dos, es decir, cuando $v(x) + 3$ sea potencia de dos. La fórmula obvia que cumple esto es:

$$\mathtt{S}^{x'''}_x\alpha \equiv \bigwedge y(\bigvee z\ x''' = y \cdot z \wedge y \neq 0' \to \bigvee w\ y = w \cdot 0'').$$

Una vez más, se trata de la fórmula que se obtiene quitando x y poniendo en su lugar x''', es decir, "sustituyendo", en el sentido usual de la palabra.

Esto podría llevarnos a pensar que una definición puramente formal de $\mathtt{S}_x^t\theta$ (sin hacer referencia a modelos) es definirlo como la expresión (término o fórmula, según lo sea θ) que resulta de cambiar cada x que aparezca en θ por el término t. Sin embargo, dar esa definición no sería una buena idea, al menos sin tener presentes un par de cuestiones técnicas.

1) En primer lugar tenemos las fórmulas en las que una misma variable puede aparecer a la vez libre y ligada. La fórmula siguiente es lógicamente equivalente a la que hemos puesto como ejemplo, en el sentido de que será satisfecha en cualquier modelo y con cualquier valoración si y sólo si lo es la que hemos dado:

$$\alpha^* \equiv \bigwedge y(\bigvee z\ x = y \cdot z \wedge y \neq 0' \to \bigvee x\ y = x \cdot 0'').$$

Cuando interpretamos esta fórmula en un modelo con una valoración, la primera x que aparece se interpretará como $v(x)$, mientras que la segunda x, al estar ligada por el cuantificador $\bigvee x$, se interpretará como convenga (si es posible) para que se cumpla la fórmula que sigue al cuantificador, y no será necesario asignarle precisamente el valor $v(x)$.

En la práctica, los matemáticos evitan este tipo de fórmulas, y en lugar de expresar así la propiedad "tener sólo divisores pares", la expresan con la fórmula inicial, donde la variable x no representa dos papeles distintos. Pero lo cierto es que estas fórmulas existen, y nuestra definición de sustitución debe dar cuenta de ellas, y si seguimos al pie de la letra nuestra propuesta de definición de sustitución obtendríamos este "engendro" que ni siquiera es una fórmula:

$$\mathtt{S}_x^{x'''}\alpha^* \equiv \bigwedge y(\bigvee z\ x''' = y \cdot z \wedge y \neq 0' \to \bigvee x'''\ y = x''' \cdot 0'').$$

Observemos que esto no es una fórmula porque las reglas sintácticas no permiten escribir $\bigvee x'''$, sino que detrás de un cuantificador sólo puede ir una variable. Pero aunque corrigiéramos nuestra definición para no tocar las variables que siguen a los cuantificadores y calculáramos esto:

$$\mathtt{S}_x^{x'''}\alpha^* \equiv \bigwedge y(\bigvee z\ x''' = y \cdot z \wedge y \neq 0' \to \bigvee x\ y = x''' \cdot 0'')$$

seguiríamos sin acertar, porque esta fórmula no significa que el objeto denotado por x''' es potencia de dos. Aquí dice que todo divisor de x''' distinto de 1 es un número par mayor o igual que 6, cosa que no cumple, por ejemplo, el 1, a pesar de que $1 + 3 = 4$ es potencia de dos. La sustitución correcta es:

$$\mathtt{S}_x^{x'''}\alpha^* \equiv \bigwedge y(\bigvee z\ x''' = y \cdot z \wedge y \neq 0' \to \bigvee x\ y = x \cdot 0''),$$

y la moraleja que extraemos de este ejemplo es que la sustitución $\mathtt{S}_x^t\alpha$ debe definirse de modo que las apariciones ligadas de x no se alteren en absoluto. Sustituir una variable por un término es sustituir únicamente las apariciones libres de la variable.

Sin embargo, esta precisión era la menor de las dos consideraciones que nos vemos obligados a hacer si queremos llegar a un concepto de "sustitución" que se corresponda realmente con nuestro objetivo en todos los casos.

2) Volvamos a nuestra fórmula α original, sin el doble papel que la x tenía en la fórmula α^*, pero supongamos ahora que queremos calcular $\mathtt{S}_x^{y+z}\alpha$.

Puesto que α significa "x es potencia de dos", queremos que $\mathtt{S}_x^{y+z}\alpha$ signifique "$y+z$ es potencia de dos", pero si nos limitamos a sustituir cada x (libre) que aparece en α por $y+z$ lo que nos sale es:

$$\mathtt{S}_x^{y+z}\alpha \equiv \bigwedge y(\bigvee z\ y+z = y\cdot z \wedge y \neq 0' \to \bigvee w\ y = w\cdot 0''),$$

y esto es un "enredo" que nada tiene que ver con que los divisores de $y+z$ sean pares. Ahí dice algo que es cierto, algo así como que, para todo número y, si existe un z que cumple la ecuación $y+z = yz$ e y no es 1, entonces y es par (y es cierto porque la condición sólo se da cuando y y z se interpretan por 0 o por 2). ¿Qué ha fallado? Que las variables y, z estaban libres en $y+z$, pero al meter este término en el lugar de x, han quedado en el radio de alcance de los cuantificadores $\bigwedge y$, $\bigvee z$, y se han mezclado con otras y, z que ya estaban en la fórmula α, y el resultado ha sido completamente imprevisible.

Los libros de lógica evitan estas sustituciones incontroladas de dos formas distintas:

A) Quizá la solución más aceptada sea prohibir que se realice la sustitución en este caso. Para ello definen que una variable x es sustituible por un término t en una expresión θ si ninguna variable que está libre en t aparece ligada en θ, y sólo consideran definida la sustitución $\mathtt{S}_x^t\theta$ cuando se cumple esta condición.

En nuestro ejemplo, la variable x no es sustituible por el término $y+z$ en la fórmula α porque las variables y, z están libres en $y+z$ y están ligadas en α.

La solución A), con ser, como decimos, la más habitual y totalmente operativa, tiene dos defectos. Uno es que obliga a poner como hipótesis en todos los teoremas que involucren una sustitución que tal o cual variable debe ser sustituible por tal o cual término en tal o cual fórmula, lo cual supone mantener en un constante primer plano lo que es una mera anécdota que nunca se encuentra uno en la práctica.

El segundo inconveniente es que esto de que ciertas variables no puedan ser sustituidas por ciertos términos es algo totalmente ajeno a la práctica habitual del matemático que razona competentemente sin conocer los tecnicismos de la lógica. Si un matemático ve $\phi(x)$ y tiene un t a mano y le interesa decir que t cumple $\phi(x)$, no concibe que no pueda escribir $\phi(t)$. ¿Qué hace en la práctica un matemático? La realidad es que lo que hace es evitar que se le den casos como el que nos ocupa eligiendo prudentemente las variables, pero imaginemos que, por un descuido, un matemático ha escrito la fórmula α como nosotros lo hemos hecho y, en el curso de un razonamiento está trabajando con una y y una z y se ve en la necesidad de decir que $y+z$ cumple la propiedad α. ¿Qué haría? Vería α:

$$\bigwedge y(\bigvee z\ x = y\cdot z \wedge y \neq 0' \to \bigvee w\ y = w\cdot 0''),$$

se daría cuenta de que no puede meter ahí $y+z$ sin hacer un estropicio y, en lugar de detenerse, corregiría la fórmula α, cambiándola por

$$\bigwedge u(\bigvee v\ x = u\cdot v \wedge u \neq 0' \to \bigvee w\ u = w\cdot 0''),$$

pues sabe que las y, z que aparecen ligadas en la fórmula no tienen nada que ver con las y, z concretas que está manejando y que le aparecen en su término $y+z$, y sabe que si en la fórmula cambia las y, z ligadas por unas u, v ligadas pasa a otra fórmula que significa lo mismo, que le vale igual, y ahora puede sustituir con tranquilidad y escribir que

$$\mathtt{S}_x^{y+z}\alpha \equiv \bigwedge u(\bigvee v\ y+z = u\cdot v \wedge u \neq 0' \rightarrow \bigvee w\ u = w\cdot 0'').$$

Así obtiene una fórmula que realmente significa "$y+z$ es potencia de dos".

En definitiva, la solución B) al problema consiste en definir la sustitución incluyendo instrucciones para sustituir variables ligadas cuando éstas entran en conflicto con variables del término que queremos sustituir. Con esto complicamos la definición de sustitución, pero eso afecta únicamente a los tres o cuatro resultados que hay que demostrar basándose en la definición de sustitución y, a partir de ahí, como las sustituciones se tratarán apelando a esos tres o cuatro resultados ya demostrados, en la práctica usual este caso patológico no deja rastro alguno. Ni hace falta añadir hipótesis a los teoremas recordando siempre lo que no es más que un caso extraño, ni hace falta contradecir la idea natural de que cualquier variable es sustituible en cualquier fórmula por cualquier término si uno tiene dos dedos de frente a la hora de hacerlo.

Naturalmente, adoptaremos la solución B), y tras esta discusión ya estamos en condiciones de dar una definición formal no ingenua de sustitución de una variable por un término.

Definición 1.11 Sea $\mathcal{L}$ un lenguaje formal. Definimos la *sustitución* de una variable x por un término t en una expresión θ de $\mathcal{L}$ como la expresión $\mathtt{S}_x^t\theta$ determinada por las reglas siguientes:

1. $\mathtt{S}_x^t x_i \equiv \begin{cases} t & \text{si } x \equiv x_i, \\ x_i & \text{si } x \not\equiv x_i. \end{cases}$
2. $\mathtt{S}_x^t c_i \equiv c_i$.
3. $\mathtt{S}_x^t R_i^n t_1 \cdots t_n \equiv R_i^n S_x^t t_1 \cdots S_x^t t_n$.
4. $\mathtt{S}_x^t f_i^n t_1 \cdots t_n \equiv f_i^n S_x^t t_1 \cdots S_x^t t_n$.
5. $\mathtt{S}_x^t \neg\alpha \equiv \neg \mathtt{S}_x^t \alpha$.
6. $\mathtt{S}_x^t(\alpha \rightarrow \beta) \equiv \mathtt{S}_x^t\alpha \rightarrow \mathtt{S}_x^t\beta$.
7. $\mathtt{S}_x^t \bigwedge x_i\alpha \equiv \begin{cases} \bigwedge x_i\alpha & \text{si } x \text{ no está libre en } \bigwedge x_i\alpha, \\ \bigwedge x_i \mathtt{S}_x^t\alpha & \text{si } x \text{ está libre en } \bigwedge x_i\alpha \text{ y } x_i \text{ no lo está en } t, \\ \bigwedge x_j \mathtt{S}_x^t \mathtt{S}_{x_i}^{x_j}\alpha & \text{si } x \text{ está libre en } \bigwedge x_i\alpha,\ x_i \text{ está libre en } t \\ & \text{y } j \text{ es el menor índice tal que } x_j \text{ no está} \\ & \text{en } \bigwedge x_i\alpha \text{ ni en } t. \end{cases}$

$$8.\ \mathtt{S}^t_x(x_i|\alpha) \equiv \begin{cases} x_i|\alpha & \text{si } x \text{ no está libre en } x_i|\alpha, \\ x_i|\mathtt{S}^t_x\alpha & \text{si } x \text{ está libre en } x_i|\alpha \text{ y } x_i \text{ no lo está en } t, \\ x_j|\mathtt{S}^t_x\mathtt{S}^{x_j}_{x_i}\alpha & \text{si } x \text{ está libre en } x_i|\alpha,\ x_i \text{ está libre en } t \\ & \text{y } j \text{ es el menor índice tal que } x_j \text{ no está} \\ & \text{en } x_i|\alpha \text{ ni en } t. \end{cases}$$

La condición 1) dice que la variable x_i se cambia por t si es la variable x, la condición 2) dice que las constantes no se modifican, las condiciones 3), 4), 5), 6) dicen que los relatores, funtores y conectores no se modifican. La condición 7) dice que, según las observaciones precedentes, si x no está libre en $\bigwedge x_i\alpha$ entonces no se sustituye nada. Si está libre y no hay conflicto con el cuantificador $\bigwedge x_i$, entonces se sustituye en α y se deja igual el $\bigwedge x_i$, pero si hay conflicto con el cuantificador, es decir, si x_i está libre en t y se ligaría al meterla dentro del alcance de $\bigwedge x_i$, antes de sustituir se cambia x_i por una variable nueva x_j. El apartado 8) distingue los mismos casos que el 7) para el descriptor.

A partir de la definición del disyuntor, etc. se obtiene fácilmente que

$$\mathtt{S}^t_x(\alpha \vee \beta) \equiv \mathtt{S}^t_x\alpha \vee \mathtt{S}^t_x\beta, \quad \mathtt{S}^t_x(\alpha \wedge \beta) \equiv \mathtt{S}^t_x\alpha \wedge \mathtt{S}^t_x\beta, \quad \mathtt{S}^t_x(\alpha \leftrightarrow \beta) \equiv \mathtt{S}^t_x\alpha \leftrightarrow \mathtt{S}^t_x\beta,$$

$$\mathtt{S}^t_x\bigvee x_i\alpha \equiv \begin{cases} \bigvee x_i\alpha & \text{si } x \text{ no está libre en } \bigvee x_i\alpha, \\ \bigvee x_i\mathtt{S}^t_x\alpha & \text{si } x \text{ está libre en } \bigvee x_i\alpha \text{ y } x_i \text{ no lo está en } t, \\ \bigvee x_j\mathtt{S}^t_x\mathtt{S}^{x_j}_{x_i}\alpha & \text{si } x \text{ está libre en } \bigvee x_i\alpha,\ x_i \text{ está libre en } t \\ & \text{y } j \text{ es el menor índice tal que } x_j \text{ no está} \\ & \text{en } \bigvee x_i\alpha \text{ ni en } t. \end{cases}$$

El lector debería considerar "razonable" la definición de sustitución que hemos dado a la luz de los comentarios precedentes, pero también debe comprender que "razonable" no es suficiente. No podemos definir "sustitución" como queramos, sino que la definición debe cumplir el propósito que nos hemos marcado. Hemos visto dos problemas que podrían hacer que no fuera así y los hemos tenido en cuenta, pero ¿eran los únicos problemas a tener en cuenta? ¿Seguro que no hay otro inconveniente que no hemos tenido en cuenta y que puede hacer que la definición que hemos dado siga sin ser adecuada a pesar de las precauciones que hemos tomado? El teorema siguiente demuestra que nuestra definición de sustitución es correcta:

Teorema 1.12 *Sea v una valoración de un lenguaje formal $\mathcal{L}$ en un modelo M. Sea t un término de $\mathcal{L}$, sea θ una expresión y x una variable. Entonces si θ es un término*

$$M(\mathtt{S}^t_x\theta)[v] \equiv M(\theta)[v^{M(t)[v]}_x]$$

y si θ es una fórmula

$$M \vDash \mathtt{S}^t_x\theta[v] \quad \textit{syss} \quad M \vDash \theta[v^{M(t)[v]}_x].$$

La primera parte dice que el objeto denotado por $\mathtt{S}^t_x\theta$ es el objeto denotado por θ cuando la variable x se interpreta como el objeto denotado por t. La segunda parte dice que $\mathtt{S}^t_x\theta$ es satisfecha si y sólo si θ es satisfecha cuando la variable x se interpreta como el objeto denotado por t. Son estos hechos los que nos permiten decir que la sustitución está "bien definida" en el sentido de que es una expresión definida "como haga falta" para que al final signifique lo que tiene que significar.

DEMOSTRACIÓN: Por inducción sobre la longitud de θ.

Si $\theta \equiv y$ distinguimos dos casos:
a) si $x \not\equiv y$ entonces $M(\mathtt{S}^t_x\theta)[v] \equiv M(y)[v] \equiv M(\theta)[v_x^{M(t)[v]}]$,
b) si $x \equiv y$ entonces $M(\mathtt{S}^t_x\theta)[v] \equiv M(t)[v] \equiv M(\theta)[v_x^{M(t)[v]}]$.

Si $\theta \equiv c$ entonces $M(\mathtt{S}^t_x\theta)[v] \equiv M(c)[v] \equiv M(\theta)[v_x^{M(t)[v]}]$.

Si $\theta \equiv R^n_i t_1 \cdots t_n$ entonces

$$M \vDash \mathtt{S}^t_x\theta[v] \text{ syss } M \vDash R^n_i \mathtt{S}^t_x t_1 \cdots \mathtt{S}^t_x t_n[v] \text{ syss}$$
$$M(R^n_i)(M(\mathtt{S}^t_x t_1)[v], \ldots, M(\mathtt{S}^t_x t_n)[v]) \text{ syss}$$
$$M(R^n_i)(M(t_1)[v_x^{M(t)[v]}], \ldots, M(t_n)[v_x^{M(t)[v]}]) \text{ syss } M \vDash \theta[v_x^{M(t)[v]}].$$

Si $\theta \equiv f^n_i t_1 \cdots t_n$ entonces

$$M(\mathtt{S}^t_x\theta)[v] \equiv M(f^n_i \mathtt{S}^t_x t_1 \cdots \mathtt{S}^t_x t_n)[v] \equiv M(f^n_i)(M(\mathtt{S}^t_x t_1)[v], \ldots, M(\mathtt{S}^t_x t_n)[v])$$
$$\equiv M(f^n_i)(M(t_1)[v_x^{M(t)[v]}], \ldots, M(t_n)[v_x^{M(t)[v]}]) \equiv M(\theta)[v_x^{M(t)[v]}].$$

Si $\theta \equiv \neg\alpha$ entonces $M \vDash \mathtt{S}^t_x\theta[v]$ syss $M \vDash \neg\mathtt{S}^t_x\alpha[v]$ syss no $M \vDash \mathtt{S}^t_x\alpha[v]$ syss no $M \vDash \alpha[v_x^{M(t)[v]}]$ syss $M \vDash \theta[v_x^{M(t)[v]}]$.

Si $\theta \equiv \alpha \to \beta$ entonces $M \vDash \mathtt{S}^t_x\theta[v]$ syss $M \vDash (\mathtt{S}^t_x\alpha \to \mathtt{S}^t_x\beta)[v]$ syss no $M \vDash \mathtt{S}^t_x\alpha[v]$ o $M \vDash \mathtt{S}^t_x\beta[v]$ syss no $M \vDash \alpha[v_x^{M(t)[v]}]$ o $M \vDash \beta[v_x^{M(t)[v]}]$ syss $M \vDash \theta[v_x^{M(t)[v]}]$.

Si $\theta \equiv \bigwedge y\alpha$ distinguimos tres casos:

a) Si x no está libre en $\bigwedge y\alpha$, entonces $M \vDash \mathtt{S}^t_x\theta[v]$ syss $M \vDash \bigwedge y\alpha[v]$ syss $M \vDash \bigwedge y\alpha[v_x^{M(t)[v]}]$ por el teorema 1.9.

b) Si x está libre en $\bigwedge y\alpha$ e y no lo está en t, fijemos un objeto a en M. Entonces $M \vDash \mathtt{S}^t_x\alpha[v^a_y]$ syss (hip. de ind.) $M \vDash \alpha[v_{y\,x}^{a\,M(t)[v^a_y]}]$ syss (por 1.9) $M \vDash \alpha[v_{y\,x}^{a\,M(t)[v]}]$ syss $M \vDash \alpha[v_{x\quad y}^{M(t)[v]a}]$ (notar que $x \not\equiv y$ pues x está libre en $\bigwedge y\alpha$ e y no lo está).

Por lo tanto, $M \vDash \mathtt{S}^t_x\theta[v]$ syss $M \vDash \bigwedge y\mathtt{S}^t_x\alpha[v]$ syss para todo a de M se cumple que $M \vDash \mathtt{S}^t_x\alpha[v^a_y]$ syss para todo a de M se cumple que $M \vDash \alpha[v_{x\quad y}^{M(t)[v]a}]$ syss $M \vDash \bigwedge y\alpha[v_x^{M(t)[v]}]$.

c) Si x está libre en $\bigwedge y\alpha$, y está libre en t y z es la variable de menor índice que no está en $\bigwedge y\alpha$ ni en t, fijemos un objeto a en M. Entonces $M \vDash \mathtt{S}^t_x\mathtt{S}^z_y\alpha[v^a_z]$ syss (hip. de ind.) $M \vDash \mathtt{S}^z_y\alpha[v^{aM(t)[v^a_z]}_{z\ x}]$ syss (1.9) $M \vDash \mathtt{S}^z_y\alpha[v^{aM(t)[v]}_{z\ x}]$ syss $M \vDash \mathtt{S}^z_y\alpha[v^{M(t)[v]a}_{x\ \ z}]$ syss (hip. de ind.) $M \vDash \alpha[v^{M(t)[v]aa}_{x\ \ zy}]$ syss (1.9) $M \vDash \alpha[v^{M(t)[v]a}_{x\ \ y}]$.

Por lo tanto $M \vDash \mathtt{S}^t_x\theta[v]$ syss $M \vDash \bigwedge z\mathtt{S}^t_x\mathtt{S}^z_y\alpha[v]$ syss para todo a de M se cumple que $M \vDash \mathtt{S}^t_x\mathtt{S}^z_y\alpha[v^a_z]$ syss para todo a de M se cumple que $M \vDash \alpha[v^{M(t)[v]a}_{x\ \ y}]$ syss $M \vDash \bigwedge y\alpha[v^{M(t)[v]}_x]$.

Si $\theta \equiv y|\alpha$ distinguimos tres casos:

a) Si x no está libre en $y|\alpha$ entonces, usando el teorema 1.9,

$$M(\mathtt{S}^t_x\theta)[v] \equiv M(y|\alpha)[v] \equiv M(\theta)[v^{M(t)[v]}_x].$$

b) Si x está libre en $y|\alpha$ e y no lo está en t, entonces, fijando un objeto a en M, como en el apartado b) del caso anterior concluimos que $M \vDash \mathtt{S}^t_x\alpha[v^a_y]$ syss $M \vDash \alpha[v^{M(t)[v]a}_{x\ \ y}]$, luego existe un único a en M tal que $M \vDash \mathtt{S}^t_x\alpha[v^a_y]$ syss existe un único a en M tal que $M \vDash \alpha[v^{M(t)[v]a}_{x\ \ y}]$, y en tal caso son el mismo.

Si se da la unicidad entonces $M(\mathtt{S}^t_x\theta)[v] \equiv M(y|\mathtt{S}^t_x\alpha)[v] \equiv a \equiv M(\theta)[v^{M(t)[v]}_x]$. En otro caso $M(\mathtt{S}^t_x\theta)[v] \equiv d \equiv M(\theta)[v^{M(t)[v]}_x]$.

c) Si x está libre en $y|\alpha$, y está libre en t y z es la variable de menor índice que no está en $y|\alpha$ ni en t, fijamos un objeto a en M y como en el apartado c) del caso anterior se prueba que $M \vDash \mathtt{S}^t_x\mathtt{S}^z_y\alpha[v^a_z]$ syss $M \vDash \alpha[v^{M(t)[v]a}_{x\ \ y}]$.

Ahora razonamos igual que en el apartado b) de este caso. ∎

Nota En el tercer caso de la definición de $\mathtt{S}^t_x\bigwedge x_i\alpha$ y de $\mathtt{S}^t_x(x_i|\alpha)$ hemos elegido una variable x_j como la de menor índice tal que x_j no está en $x_i|\alpha$ ni en t, pero en la demostración del teorema anterior no hemos usado en ningún momento que z sea precisamente la variable de menor índice con tal propiedad, sino que el argumento vale para cualquier variable z con dicha propiedad. Esto no es casual. Fijar concretamente la variable de menor índice es sólo un criterio arbitrario para elegir una fórmula concreta a la que llamar $\mathtt{S}^t_x\theta$, pero precisamente el hecho de que el teorema anterior no se vea afectado si la variable x_j no se elige precisamente como la de menor índice prueba que, en realidad, cualquier otra elección es igualmente legítima y aceptable.

Por ello, en lo sucesivo, cuando hablemos de una expresión $\mathtt{S}^t_x\theta$ entenderemos que nos referimos a cualquier expresión calculada según la definición 1.11 pero sin exigir necesariamente que las variables x_j se escogen como las de menor índice, sino que admitimos elecciones arbitrarias.

De este modo, cuando planteemos una identidad entre expresiones que contienen sustituciones, habrá que entender que la identidad se da si las variables x_j se eligen consistentemente en ambos miembros. Por ejemplo, en el teorema siguiente vale si se escogen en ambos miembros con el criterio del mínimo índice.

∎

El teorema siguiente contiene las propiedades más importantes de la sustitución:

Teorema 1.13 *Se cumple:*

1. $\mathtt{S}^x_x\theta \equiv \theta$.
2. *Si x no está libre en θ, entonces* $\mathtt{S}^t_x\theta \equiv \theta$.
3. *Una variable y está libre en $\mathtt{S}^t_x\theta$ syss o bien y está libre en θ e $y \not\equiv x$, o bien x está libre en θ e y está libre en t.*
4. *Si y no está en θ entonces* $\mathtt{S}^x_y\mathtt{S}^y_x\theta \equiv \theta$.
5. *Si x está libre en θ, entonces las variables libres de t y las variables libres de θ distintas de x están libres en* $\mathtt{S}^t_x\theta$.

Demostración: Todos los apartados se demuestran igual. Veamos sólo la propiedad 4). Razonamos por inducción sobre la longitud de θ.

Observamos en primer lugar que $\mathtt{S}^x_y\mathtt{S}^y_x x_i \equiv x_i$. Para ello hay que distinguir dos casos: si $x \equiv x_i$ entonces $\mathtt{S}^x_y\mathtt{S}^y_x x_i \equiv \mathtt{S}^{x_i}_y y \equiv x_i$, mientras que si $x \not\equiv x_i$ entonces $\mathtt{S}^x_y\mathtt{S}^y_x x_i \equiv \mathtt{S}^x_y x_i \equiv x_i$ (porque, por hipótesis $y \not\equiv x_i$).

El caso de c_i es trivial, porque las constantes permanecen invariables. También son inmediatos los casos correspondientes a relatores, funtores, el negador y el implicador. Veamos el caso del generalizador. Supongamos que $\theta \equiv \bigwedge x_i\alpha$.

$$\mathtt{S}^x_y\mathtt{S}^y_x\bigwedge x_i\alpha \equiv \begin{cases} \mathtt{S}^x_y\bigwedge x_i\alpha & \text{si } x \text{ no está libre en } \bigwedge x_i\alpha, \\ \mathtt{S}^x_y\bigwedge x_i\mathtt{S}^y_x\alpha & \text{si } x \text{ está libre en } \bigwedge x_i\alpha. \end{cases}$$

El tercer caso de la definición de sustitución no puede darse, porque exigiría que $x_i \equiv y$, pero estamos suponiendo que y no está en θ. Por este mismo motivo, en el primer caso llegamos a $\bigwedge x_i\alpha$. En el segundo caso es claro que si x está libre en $\bigwedge x_i\alpha$, también lo está en α, luego, por 3), y está libre en $\mathtt{S}^y_x\alpha$, y como $y \not\equiv x_i$, también está libre en $\bigwedge x_i\mathtt{S}^y_x\alpha$. Por consiguiente la sustitución es $\mathtt{S}^x_y\bigwedge x_i\mathtt{S}^y_x\alpha \equiv \bigwedge x_i\mathtt{S}^x_y\mathtt{S}^y_x\alpha \equiv \bigwedge x_i\alpha$ por hipótesis de inducción. El caso del descriptor es muy simliar. ∎

La notación matemática Aunque la notación que hemos empleado para la sustitución es la más conveniente para nuestros fines, en matemáticas es habitual escribir $\theta(x)$ para indicar que $\theta(t) \equiv \mathtt{S}^t_x\theta$. No hay que entender, salvo que se diga explícitamente, que la variable x está libre en θ ni que sea la única variable libre de θ. El escribir $\theta(x)$ simplemente nos indica qué variable hay que sustituir por t para interpretar $\theta(t)$. Generalizar esto a varias variables requiere una precaución:

Si escribimos $\theta(y_1, \ldots, y_n)$ en lugar de θ, donde $y_1, \ldots, y_n$ son variables cualesquiera, entonces se define

$$\theta(t_1, \ldots, t_n) \equiv \mathtt{S}^{t_1}_{z_1} \cdots \mathtt{S}^{t_n}_{z_n}\mathtt{S}^{z_1}_{y_1} \cdots \mathtt{S}^{z_n}_{y_n}\theta,$$

donde $z_1, \ldots, z_n$ son las variables de menor índice (o más en general, variables cualesquiera) que no estén ni en $t_1, \ldots, t_n$ ni en θ ni en $y_1, \ldots, y_n$.

Notemos que no podemos definir $\theta(t_1, t_2) \equiv \mathbf{S}^{t_1}_{y_1}\mathbf{S}^{t_2}_{y_2}\theta$ porque entonces estaríamos sustituyendo por t_1, no sólo las variables y_1 que hubiera en θ, sino también las que hubiera en t_2. ∎

1.6 Fórmulas verdaderas y falsas

El tratamiento de las variables libres nos plantea un problema técnico, y es que los matemáticos, en su uso cotidiano, tratan a veces las variables libres como que representan a objetos arbitrarios (y sus conclusiones valen entonces para todo valor de las variables) y a veces como que representan a objetos particulares (y sus conclusiones valen entonces para un cierto valor de las variables). La diferencia depende del contexto. Si un matemático ha empezado un razonamiento diciendo "tomemos un número real arbitrario x", entonces la variable x es genérica, mientras que si la introduce en la forma "podemos asegurar que esta ecuación tiene al menos una solución x", en lo sucesivo "recuerda" que la variable x es particular.

De momento no estamos en condiciones de tener en cuenta estos contextos, así que vamos a adoptar una interpretación "por defecto" de las variables libres (la genérica), pero teniendo en cuenta que más adelante nos las ingeniaremos para considerar variables particulares sin contradecir estrictamente nada de lo que vamos a decir ahora.

Definición 1.14 Una fórmula α de un lenguaje formal $\mathcal{L}$ es *verdadera* en un modelo M si $M \models \alpha[v]$ cualquiera que sea la valoración v de $\mathcal{L}$ en M. Lo representaremos $M \models \alpha$. Diremos que α es *falsa* en M si ninguna valoración v de $\mathcal{L}$ en M cumple $M \models \alpha[v]$. Si Γ es un conjunto de fórmulas escribiremos $M \models \Gamma$ para indicar que todas las fórmulas de Γ son verdaderas en M. Diremos también que M es un *modelo* de Γ.

Nota Esta definición presupone algo que debe ser matizado: que cuando hablamos de la totalidad de las valoraciones de un lenguaje en un modelo sabemos lo que estamos diciendo. Esto no está claro en absoluto: cuando hablamos de que todas las fórmulas de un lenguaje cumplen algo sabemos lo que queremos decir: es fácil enumerarlas explícitamente, y entonces nuestra afirmación significa que la primera cumple lo indicado, y la segunda también, etc. (con independencia de si sabemos probar o no que así es). Por el contrario, no tenemos ninguna representación similar que nos permita atribuir un significado a las afirmaciones que hagamos sobre la totalidad de las valoraciones.

Pese a esto, la definición anterior tiene un sentido preciso gracias al teorema 1.9. Los únicos modelos que vamos a considerar, tanto a nivel teórico como a nivel práctico, (sin entrar en la cuestión de si tendría sentido hablar de modelos más generales) van a ser modelos cuyo universo es un conjunto numerable, es decir, tal que sabemos establecer una correspondencia biunívoca entre sus objetos y los números naturales (no necesariamente calculable en la práctica). En tal caso, una afirmación sobre la totalidad de los objetos del modelo se entiende como una afirmación válida para el primer objeto, y para el

segundo, etc. En estas circunstancias —que, según lo dicho, son las únicas en las que vamos a trabajar—, también sabemos dar un sentido preciso a cualquier afirmación sobre la totalidad de los grupos $(a_1, \ldots, a_n)$ de n objetos del modelo en un orden dado y con posibles repeticiones. En efecto, podemos enumerar explícitamente todas las n-tuplas de números naturales (ponemos primero todas las formadas por números que sumen 0 (hay una sola), luego todas las formadas por números que sumen 1 (hay n), etc.) De este modo, una afirmación sobre la totalidad de los grupos de n objetos es verdadera si se cumple con el primer grupo de n objetos, y con el segundo, etc.

Ahora sólo queda observar que en virtud del teorema 1.9 podríamos haber definido que una fórmula α es verdadera en un modelo M como que α es satisfecha para todas las interpretaciones posibles de sus variables libres, lo cual sí sabemos lo que significa porque son un número finito. Si sucede esto, tendremos que $M \models \alpha[v]$ para cualquier valoración v que consideremos, tal y como exige la definición de verdad que hemos dado.

Ejemplos Consideremos las fórmulas de $\mathcal{L}_a$:

$$x + y = y + x, \qquad x + 0' = 0, \qquad x + 0'' = 0'''.$$

La primera es verdadera en el modelo natural del lenguaje de la aritmética, mientras que la segunda es falsa y la tercera no es ni verdadera ni falsa, ya que es satisfecha por las valoraciones que cumplen $v(x) \equiv 1$ y no lo es por las que no cumplen esto. ∎

Claramente se cumplen los hechos siguientes:

1. Una fórmula no puede ser verdadera y falsa en un mismo modelo (pues tomando una valoración cualquiera, será satisfecha o no lo será).

2. Toda sentencia es verdadera o falsa en un modelo (por el teorema 1.9).

3. Una fórmula α es verdadera en un modelo M syss $\neg\alpha$ es falsa, y α es falsa syss $\neg\alpha$ es verdadera.

4. Si α no tiene descriptores, entonces $M \models \alpha$ se cumple o no independientemente de cuál sea la descripción impropia de M. En particular, $M \models \alpha$ en $\mathcal{L}$ syss $M \models \alpha$ en $\underline{\mathcal{L}}$.

1.7 Consideraciones finales

Al margen de los tecnicismos en los que necesariamente hemos tenido que incurrir, lo más importante que el lector debe extraer de este capítulo es que ahora tenemos un concepto preciso de lo que es formalizar una afirmación informal. Significa diseñar un lenguaje formal con los signos adecuados y definir en él una fórmula que, en un modelo adecuado, signifique precisamente la afirmación que pretendíamos formalizar.

Por ejemplo, ahora sabemos formalizar los axiomas de Peano, considerando el lenguaje de la aritmética que hemos definido:

$$\begin{array}{ll}
\text{(AP1)} & \bigwedge x\ x' \neq 0 \\
\text{(AP2)} & \bigwedge xy(x' = y' \to x = y) \\
\text{(AP3)} & \bigwedge x\ x + 0 = x \\
\text{(AP4)} & \bigwedge xy(x + y' = (x + y)') \\
\text{(AP5)} & \bigwedge x\ x \cdot 0 = 0 \\
\text{(AP6)} & \bigwedge xy(xy' = xy + x) \\
\text{(AP7)} & \phi(0) \wedge \bigwedge x(\phi(x) \to \phi(x')) \to \bigwedge x\phi(x),
\end{array}$$

donde ϕ es cualquier fórmula, tal vez con más variables libres aparte de x (notemos que estamos usando la notación matemática, de modo que $\phi(0) \equiv \mathtt{S}^0_x\phi$, $\phi(x') \equiv \mathtt{S}^{x'}_x\,\phi$).

En efecto, es claro que en el modelo natural de la aritmética, AP1 significa que el cero no es el siguiente de ningún número natural, AP2 significa que números naturales distintos tienen siguientes distintos, etc. Notemos que los dos primeros axiomas de Peano "originales", es decir, que el cero es un número natural y que el siguiente de un número natural es también un número natural, no tienen cabida en este contexto, porque son triviales: en la interpretación natural de $\mathcal{L}_a$ todos los objetos son números naturales, luego el objeto denotado por 0, o el denotado por cualquier término x', es necesariamente un número natural, por lo que no podemos enunciar un axioma que exija estos hechos: simplemente no hay nada que exigir.

En el capítulo siguiente veremos qué es exactamente formalizar un razonamiento informal.

Otro hecho fundamental es que en este capítulo hemos presentado simultáneamente los resultados sintácticos (sobre lenguajes formales) y los semánticos (sobre modelos), pero perfectamente habríamos podido dar las definiciones de lenguaje formal, términos, fórmulas, variables libres y ligadas, la sustitución, el lenguaje de la aritmética, los axiomas de Peano, etc. sin haber definido siquiera lo que es un modelo. Por eso podemos decir que todos estos conceptos son formales, porque pueden definirse considerando a los signos de un lenguaje como meros signos sin significado, únicamente a partir de la forma en que se combinan unos con otros.

Esto es muy importante, porque significa que cuando uno define un lenguaje formal y selecciona unas fórmulas en él (como puedan ser los axiomas de Peano en el lenguaje de la aritmética, o los axiomas de la teoría de conjuntos en el lenguaje de la teoría de conjuntos, de la que hablaremos más adelante) no está obligado en absoluto a determinar un modelo de su lenguaje, sino que puede trabajar con todo rigor con sus términos y sus fórmulas sin atribuirles ninguna interpretación en particular.

Ahora bien, si en este capítulo hubiéramos presentado únicamente los conceptos sintácticos relacionados omitiendo toda referencia a la semántica, el lector podría haberse quedado con la falsa idea de que todas las definiciones son arbitrarias, de modo que, igual que hemos definido $\alpha \vee \beta \equiv \neg\alpha \to \beta$, podríamos

haber definido $\alpha \vee \beta \equiv \alpha \rightarrow \neg\beta$. Precisamente para evitar eso hemos optado por el tratamiento paralelo de la sintaxis y la semántica, para hacer el máximo hincapié en que las definiciones formales no son arbitrarias, sino que son las que son para que signifiquen lo que queremos que signifiquen (el disyuntor se define para que signifique "o", el particularizador se define para que signifique "existe", la sustitución se define para que signifique lo que tiene que significar, etc.).

Cuando los lenguajes formales se presentan sólo formalmente, sin referencias a la semántica, uno tiende a pensar que la parte formal es "lo riguroso" y que la semántica es algo "de andar por casa", y aunque posteriormente se estudie la semántica, no es raro que el lector se quede con la idea de que la semántica es "un mero complemento" a la sintaxis, en lugar de la razón de fondo que justifica que las definiciones formales adoptadas sean las que son y no otras distintas.

Por último, señalemos que en la sección 3.2 de [LF] mostramos que toda la parte sintáctica (no la semántica) que hemos expuesto informalmente en este capítulo puede formalizarse en la Aritmética Recursiva Primitiva (ARP), que es una teoría axiomática formal mucho más débil que la Aritmética de Peano que sólo permite formalizar argumentos finitistas en el sentido más estricto. En cambio, formalizar el concepto de modelo requiere teorías más fuertes (véase [LF 7.29, 7.45]).

Capítulo II

El cálculo deductivo

En este capítulo formalizaremos el razonamiento matemático. Es fundamental comprender que no vamos a "definir" el razonamiento matemático, en el sentido de dar una definición arbitraria que pueda parecer una opción entre una infinidad de alternativas ni mejores ni peores, sino que vamos a definir un concepto de razonamiento formal que, en el contexto de las afirmaciones formalizables en un determinado lenguaje formal, sea totalmente equivalente a nuestra capacidad de razonamiento informal que nos caracteriza como seres racionales y, en particular, que nos permite seguir los razonamientos informales que hemos venido haciendo hasta aquí. Este razonamiento formal que vamos a definir no es ni más ni menos que el razonamiento que usan cotidianamente los matemáticos.

2.1 Reglas de inferencia semánticas

En la introducción dijimos que razonar (informalmente) es deducir consecuencias de unas premisas de tal modo que tengamos la garantía de que, si las premisas son verdaderas, también lo serán las consecuencias que extraigamos de ellas. Los conceptos presentados en el capítulo anterior nos permiten precisar esta idea:

Definición 2.1 Sea $\mathcal{L}$ un lenguaje formal y consideremos fórmulas $\alpha_1, \ldots, \alpha_n, \alpha$ de $\mathcal{L}$. Escribiremos $\alpha_1, \ldots, \alpha_n \vDash \alpha$ para expresar que es posible razonar que cuando $\alpha_1, \ldots, \alpha_n$ son verdaderas en un modelo M, entonces necesariamente α es también verdadera en M. A los resultados de este tipo los llamaremos *reglas de inferencia semánticas.*

En estos términos, razonar informalmente sobre un lenguaje formal es justificar reglas de inferencia semánticas. Un ejemplo de regla de inferencia semántica es

$$\alpha \to \beta, \alpha \vDash \beta.$$

Esta regla se conoce como *modus ponendo ponens* (es decir, la regla que afirmando (α) afirma (β)). Vamos a demostrarla de dos formas distintas:

1) Supongamos que M es un modelo del lenguaje considerado y que se cumple $M \vDash (\alpha \to \beta)$ y $M \vDash \alpha$. Esto significa que, para toda valoración v en M, se cumple $M \vDash (\alpha \to \beta)[v]$ y $M \vDash \alpha[v]$. Por definición de satisfacción, lo primero significa que o bien no $M \vDash \alpha[v]$, o bien $M \vDash \beta[v]$, pero la primera posibilidad no puede darse, pues tenemos que $M \vDash \alpha[v]$. Así pues, tiene que ser $M \vDash \beta[v]$. Como esto vale para toda valoración v, concluimos que $M \vDash \beta$, como queríamos probar.

2) Cuando no hay cuantificadores involucrados, podemos usar tablas de verdad. En este caso tenemos:

α	β	$\alpha \to \beta$
V	V	**V**
V	F	F
F	V	V
F	F	V

Observamos que en el único caso en que tanto $\alpha \to \beta$ como α son verdaderas, sucede que β también lo es. Como la tabla contempla todas las posibilidades para las tres fórmulas, queda claro que es imposible que $\alpha \to \beta$ y α sean verdaderas en un modelo sin que β lo sea. ∎

Otro ejemplo es el *modus tollendo ponens* (es decir, la regla que negando (α) afirma (β)):

$$\alpha \vee \beta, \neg\alpha \vDash \beta.$$

Para demostrarla basta considerar la tabla de verdad siguiente:

α	β	$\neg\alpha$	$\alpha \vee \beta$
V	V	F	V
V	F	F	V
F	V	**V**	**V**
F	F	V	F

vemos que en el único caso en que las dos premisas son verdaderas, la conclusión también lo es, por lo que si tomamos un modelo en el que tanto $\alpha \vee \beta$ como $\neg\alpha$ sean verdaderas, es claro que β también tiene que serlo. ∎

Dejamos como ejercicio al lector la justificación con la tabla de verdad correspondiente del *modus tollendo tollens* (la regla que negando (β) niega (α)):

$$\alpha \to \beta, \neg\beta \vDash \neg\alpha.$$

Una regla de inferencia no tiene por qué tener premisas. Es el caso de la regla del *tercio excluido* o, en latín, *tertium non datur:* $\vDash \alpha \vee \neg\alpha$. Lo que afirma esta regla es que la fórmula $\alpha \vee \neg\alpha$ es verdadera en cualquier modelo que consideremos, como se prueba trivialmente a partir de su tabla de verdad.

Veamos ahora alguna regla de inferencia que no puede demostrarse mediante tablas de verdad. Por ejemplo la de *transitividad de la igualdad:*

$$t_1 = t_2, t_2 = t_3 \vDash t_1 = t_3.$$

En efecto, supongamos que M es un modelo en el que $M \models t_1 = t_2$ y $M \models t_2 = t_3$. Esto significa que, para cualquier valoración v, se cumple que $M \models (t_1 = t_2)[v]$ y $M \models (t_2 = t_3)[v]$. A su vez, esto equivale a que se cumpla $M(t_1)[v] \equiv M(t_2)[v]$ y $M(t_2)[v] \equiv M(t_3)[v]$, es decir, a que $M(t_1)[v], M(t_2)[v]$, y $M(t_3)[v]$ son el mismo objeto, luego en particular $M(t_1)[v] \equiv M(t_3)[v]$, lo cual equivale a $M \models (t_1 = t_3)[v]$. Como vale para toda valoración, $M \models t_1 = t_3$. ∎

Tampoco podemos usar tablas de verdad cuando hay cuantificadores involucrados. Por ejemplo en la regla de eliminación del generalizador:

$$\bigwedge x\alpha \models \mathtt{S}^t_x\alpha.$$

En efecto, supongamos que M es un modelo en el que $M \models \bigwedge x\alpha$. Esto significa que, para toda valoración v, se cumple $M \models \bigwedge x\alpha[v]$. A su vez, esto significa que, para todo a en M, se cumple $M \models \alpha[v^a_x]$. Aplicamos esto concretamente a $a \equiv M(t)[v]$. Entonces $M \models \alpha[v_x^{M(t)[v]}]$, y por el teorema 1.12 esto equivale a $M \models \mathtt{S}^t_x\alpha[v]$. Como vale para toda valoración, concluimos que $M \models \mathtt{S}^t_x\alpha$. ∎

Veamos un ejemplo más sofisticado que involucra al igualador y a un cuantificador. Si la variable x no está libre en el término t, entonces:

$$\models (\bigwedge x(x = t \to \alpha) \leftrightarrow \mathtt{S}^t_x\alpha).$$

En efecto, tomamos un modelo M y hemos de probar que la fórmula indicada es verdadera en él. Esto equivale a que para toda valoración v se tiene que cumplir

$$M \models (\bigwedge x(x = t \to \alpha) \leftrightarrow \mathtt{S}^t_x\alpha)[v].$$

Por la tabla de verdad del coimplicador esto equivale a que se cumple

$$M \models \bigwedge x(x = t \to \alpha)[v]$$

si y sólo si se cumple $M \models \mathtt{S}^t_x\alpha[v]$.

Supongamos en primer lugar que $M \models \bigwedge x(x = t \to \alpha)[v]$. Esto quiere decir que, para todo a en el universo del modelo (y tomamos concretamente $a \equiv M(t)[v]$) se cumple $M \models (x = t \to \alpha)[v_x^{M(t)[v]}]$. Esto significa que, o bien no se cumple $M \models (x = t)[v_x^{M(t)[v]}]$, o bien $M \models \alpha[v_x^{M(t)[v]}]$.

Ahora bien, $M \models (x = t)[v_x^{M(t)[v]}]$ equivale a $v_x^{M(t)[v]}(x) \equiv M(t)[v]$ (donde usamos que x no está libre en t), y esto sí que se cumple, luego $M \models \alpha[v_x^{M(t)[v]}]$. El teorema 1.12 nos da que esto equivale a $M \models \mathtt{S}^t_x\alpha[v]$.

Recíprocamente, si se cumple $M \models \mathtt{S}^t_x\alpha[v]$, para probar que también se tiene $M \models \bigwedge x(x = t \to \alpha)[v]$ hemos de probar que para todo a en el universo del modelo, se cumple $M \models (x = t \to \alpha)[v^a_x]$. Para asegurar que esto sucede hemos de ver que si $M \models (x = t)[v^a_x]$, entonces $M \models \alpha[v^a_x]$.

Suponemos, pues, que $M \models (x = t)[v^a_x]$, es decir, que $a \equiv M(t)[v^a_x]$. Como x no está libre en t, se cumple que $a \equiv M(t)[v^a_x] \equiv M(t)[v]$.

Ahora usamos que $M \vDash \mathtt{S}^t_x\alpha[v]$, que por el teorema 1.12 es lo mismo que $M \vDash \alpha[v_x^{M(t)[v]}]$, o también que $M \vDash \alpha[v_x^a]$, que es justo lo que queríamos probar. ∎

Hay una regla de inferencia que involucra el cuantificador universal que requiere especial atención. Es la de *introducción del generalizador:*

$$\alpha \vDash \bigwedge x\alpha.$$

En efecto, si se cumple $M \vDash \alpha$, esto significa que se cumple $M \vDash \alpha[v]$, para toda valoración v. En particular, para cualquier a en M, lo anterior se cumple para la valoración v_x^a, es decir, $M \vDash \alpha[v_x^a]$, y esto es justo lo que significa que $M \vDash \bigwedge x\alpha[v]$. Como vale para toda valoración, tenemos que $M \vDash \bigwedge x\alpha$. ∎

Su peculiaridad es que se cumple en virtud del convenio que hemos adoptado de considerar que una fórmula es verdadera cuando es satisfecha por cualquier valoración, lo cual equivale a que se cumple si y sólo si se cumple para toda posible interpretación de sus variables libres, y en particular si y sólo si se cumple "para todo x". Ahora bien, ya hemos comentado que los matemáticos no siempre consideran que lo que dice una fórmula hay que entenderlo como que vale para cualquier interpretación de sus variables, sino que distinguen (según el contexto) variables que representan objetos generales y variables que representan objetos particulares. Al adoptar el convenio sobre interpretación de las variables libres nos hemos apartado ligeramente de los hábitos matemáticos y la regla de introducción del generalizador pone de manifiesto esa divergencia. Más adelante veremos cómo podemos "congraciarnos" de nuevo con la práctica usual de los matemáticos sin desdecirnos de lo dicho.

Las reglas de inferencia semánticas nos permiten desglosar los razonamientos complejos en razonamientos más simples y reducirlos a "piezas" ya chequeadas y que no es necesario volver a chequear. Por ejemplo, consideremos de nuevo el ejemplo que pusimos de razonamiento (informal) válido en la introducción:

	Todos los españoles son europeos
	Cervantes era español
luego	Cervantes era europeo.

Podemos formalizarlo así:

$$\bigwedge x(\mathrm{Es}\, x \to \mathrm{Eu}\, x),\ \mathrm{Es}\, C \vDash \mathrm{Eu}\, C.$$

Aquí estamos considerando un lenguaje formal $\mathcal{L}$ con una constante C y dos relatores monádicos Es y Eu. Si consideramos el modelo de $\mathcal{L}$ cuyo universo está formado, digamos, por los seres humanos que han poblado el mundo desde el siglo XVI hasta la actualidad, interpretamos la constante C como Cervantes y los relatores como las relaciones "ser español" y "ser europeo", entonces la interpretación de las tres fórmulas que estamos considerado es justo la de las fórmulas de nuestro razonamiento informal, pero podemos demostrar que forman una regla de inferencia semántica, es decir, que la conclusión será verdadera en cualquier

modelo de $\mathcal{L}$ en el que las premisas sean verdaderas. No importa si el universo del modelo lo forman seres humanos o protozoos. Podríamos razonar que esta regla de inferencia es válida con la misma clase de argumentos semánticos que hemos venido empleando hasta ahora, pero otra opción es encadenar reglas de inferencia ya demostradas, así:

1)	$\bigwedge x(\mathrm{Es}\, x \to \mathrm{Eu}\, x)$	Premisa
2)	$\mathrm{Es}\, C$	Premisa
3)	$\mathrm{Es}\, C \to \mathrm{Eu}\, C$	Eliminación del generalizador 1
4)	$\mathrm{Eu}\, C$	Modus Ponens 2,3

¿Cómo hay que entender esto? Supongamos que las dos premisas son verdaderas en un modelo M. Entonces podemos asegurar que la fórmula 3) también será verdadera en M, porque resulta de aplicar la regla de inferencia de eliminación del generalizador, que ya hemos demostrado (notemos que la fórmula 3 resulta de calcular $\mathtt{S}_x^C$ sobre la fórmula tras $\bigwedge x$ en 1). A su vez 4) tiene que ser verdadera en M porque se obtiene de 2 y 3 por la regla Modus Ponens, que también está demostrada.

Así hemos reducido un razonamiento que podríamos haber improvisado directamente a la aplicación de unas reglas de inferencia, que no son sino razonamientos ya comprobados y que no hace falta comprobar cada vez. Más aún, para construir la sucesión de fórmulas precedente no ha hecho falta considerar para nada ningún modelo. Sólo ha sido un problema de encajar adecuadamente las reglas oportunas, como un puzzle.

Ahora estamos en condiciones de precisar nuestro objetivo inmediato: ¿es posible seleccionar un número finito de reglas de inferencia semánticas (lo cual supone demostrarlas) de modo que siempre que queramos probar una afirmación del tipo $\alpha_1, \ldots, \alpha_n \vDash \alpha$ (es decir, que el hecho de que unas premisas sean verdaderas en un modelo obliga a que también lo sea la conclusión) podamos hacerlo aplicando oportunamente a las premisas las reglas de inferencia seleccionadas, sin necesidad de mencionar modelos para nada?

La respuesta es afirmativa (pero no es nada trivial que así sea, y esto es una de las joyas que nos regala la lógica), y eso es formalizar el razonamiento, es decir, reducir cualquier razonamiento (que, en su versión no formalizada, significa justificar racionalmente que la verdad de unas premisas en un modelo fuerza a la verdad de la conclusión) a una concatenación de un número finito (fijo) de reglas de inferencia semánticas previamente demostradas, sin hacer referencia alguna a modelos.

Terminamos esta sección introduciendo algunos conceptos semánticos adicionales. Notemos que una regla de inferencia de la forma $\vDash \alpha$ significa que podemos justificar que la fórmula α tiene que ser verdadera en todo modelo de su lenguaje. Vamos a dar nombre a este hecho:

Definición 2.2 Una fórmula α de un lenguaje formal $\mathcal{L}$ es *lógicamente válida* si es verdadera en todo modelo de $\mathcal{L}$, es decir, si $\vDash \alpha$. Diremos que α es *insatisfacible* si es falsa en todo modelo, α es *satisfacible* si es verdadera en algún modelo y es *falseable* si es falsa en algún modelo.

Nota La definición que acabamos de dar de fórmula lógicamente válida no es aceptable en los términos en que la hemos dado. En principio, $\models \alpha$ no significa que α es verdadera en todo modelo, sino que *podemos demostrar* (informalmente) que α tiene que ser verdadera en todo modelo.

La diferencia es esencial. ¿Qué sucedería si no encontráramos ningún argumento que demostrara que una fórmula es verdadera en todo modelo, y al mismo tiempo no fuéramos capaces de definir ningún modelo en el que la fórmula no fuera verdadera? ¿Podríamos afirmar que, pese a todo, habrá o no habrá un modelo en el que no sea verdadera y que, por consiguiente, la fórmula no será lógicamente válida o sí lo será, aunque no sepamos cuál es el caso? La realidad es que sí, según veremos, pero no es menos cierto que ahora no estamos en condiciones de justificarlo.

El concepto de modelo supone hablar de conjuntos, relaciones y funciones, "bien definidas", y no tenemos un concepto bien definido de lo que son conjuntos etc. "bien definidos", así que no podemos decir honestamente que sabemos lo que queremos decir cuando afirmamos que una fórmula es verdadera "en todo modelo" (de su lenguaje formal). Todos nuestros razonamientos sobre modelos son esquemáticos, en el sentido de que sólo adquieren un significado definido cuando se aplican a modelos concretos que podemos reconocer que están bien definidos.

Por ello, aunque hemos dado la definición anterior en los términos en que la hemos dado porque a la larga justificaremos que tiene sentido, de momento sólo podemos concebirla en el sentido más débil que ya habíamos considerado en la definición de regla de inferencia semántica: una fórmula es lógicamente válida si podemos razonar que tiene que ser verdadera en todo modelo, es insatisfacible si podemos razonar que tiene que ser falsa en todo modelo, es satisfacible si podemos definir un modelo en el que es verdadera y es falseable si podemos definir un modelo en el que es falsa. ∎

Claramente se cumple:

1. α es lógicamente válida syss $\neg\alpha$ es insatisfacible.
2. α es insatisfacible syss $\neg\alpha$ es lógicamente válida.
3. α no puede ser lógicamente válida e insatisfacible.
4. α es satisfacible syss $\neg\alpha$ es falseable.
5. α es falseable syss $\neg\alpha$ es satisfacible.

Todos estos hechos han de entenderse en los términos explicados en la nota anterior. Por ejemplo, la afirmación 1) expresa que si existe un argumento que nos convence de que una fórmula α es necesariamente verdadera en cualquier modelo dado, dicho argumento se prolonga inmediatamente hasta otro que nos convence de que $\neg\alpha$ es necesariamente falsa en cualquier modelo dado, etc.

La propiedad 3) hace uso de que todo lenguaje formal $\mathcal{L}$ tiene al menos un modelo. En efecto, es fácil definir un modelo cuyo universo tenga, por ejemplo, un único objeto y en el que las constantes, los relatores y los funtores de $\mathcal{L}$ se interpreten de forma obvia.

2.2 Sistemas deductivos formales

Empezamos a desarrollar aquí el proyecto que hemos trazado en la sección anterior, es decir, seleccionar algunas reglas de inferencia semánticas (concretamente seleccionaremos diez, de las cuales ocho serán de la forma $\vDash \alpha$) de tal forma que todo razonamiento informal sobre los objetos de un modelo de un determinado lenguaje formal pueda descomponerse en una concatenación de aplicaciones de dichas reglas de inferencia a las premisas. En primer lugar introducimos las definiciones oportunas para concretar estas ideas:

Definición 2.3 Un *sistema deductivo formal* (de primer orden) F sobre un lenguaje formal $\mathcal{L}$ viene determinado por un conjunto de fórmulas de $\mathcal{L}$, llamadas *axiomas* de F, y un conjunto de *reglas primitivas de inferencia* de F, que determinan cuándo una fórmula de $\mathcal{L}$ es *consecuencia inmediata* de otra u otras fórmulas de $\mathcal{L}$.

Es importante señalar que cuando hablamos de reglas primitivas de inferencia no hay que entender que sean reglas de inferencia semánticas, sino que nos referimos a cualquier criterio que estipule que una fórmula es consecuencia de otras. Por ejemplo, una regla primitiva de inferencia para un sistema deductivo formal podría ser: "De $\alpha \to \beta$ y $\neg\alpha$ es consecuencia inmediata $\neg\beta$". Permitimos este nivel de arbitrariedad para garantizar que la definición que acabamos de dar (al igual que las siguientes) sea puramente formal, es decir, que no haga alusión alguna a los posibles modelos del lenguaje considerado.

Una *deducción* en un sistema deductivo formal F a partir de un conjunto de fórmulas Γ es una sucesión finita $\alpha_1, \ldots, \alpha_n$ de fórmulas de $\mathcal{L}$ tales que cada α_i es un axioma de F, una fórmula de Γ o una consecuencia inmediata de fórmulas anteriores de la sucesión. Las fórmulas de Γ se llaman *premisas* de la deducción. Una *demostración* es una deducción sin premisas.

Una fórmula α es una *consecuencia* en F de un conjunto de fórmulas Γ si α es la última fórmula de una deducción en F a partir de Γ. Lo representaremos con la notación $\Gamma \underset{F}{\vdash} \alpha$.

Una fórmula α es un *teorema* de F si es la última fórmula de una demostración en F. Lo representaremos mediante $\underset{F}{\vdash} \alpha$.

Si el conjunto de premisas es finito, digamos $\gamma_1, \ldots, \gamma_n$, escribiremos también $\gamma_1, \ldots, \gamma_n \underset{F}{\vdash} \alpha$. La notación $\Gamma, \gamma_1, \ldots, \gamma_n \underset{F}{\vdash} \alpha$ significará, como es obvio, que α es consecuencia en F de las premisas de Γ más las indicadas explícitamente.

Como en una deducción sólo cabe un número finito de premisas, es claro que si $\Gamma \underset{F}{\vdash} \alpha$, existen $\gamma_1, \ldots, \gamma_n$ en Γ tales que $\gamma_1, \ldots, \gamma_n \underset{F}{\vdash} \alpha$.

De entre el amplio abanico de posibilidades que satisfacen la definición de sistema deductivo formal, nos interesan únicamente aquellas que cumplan un requisito semántico fundamental:

Definición 2.4 Un sistema deductivo formal es *correcto* si todos sus axiomas son fórmulas lógicamente válidas y todas sus reglas de inferencia primitivas son reglas de inferencia semánticas.

Esto tiene una consecuencia clara. Para enunciarla conviene generalizar ligeramente el concepto de regla de inferencia semántica para admitir infinitas premisas: Si Γ es un conjunto de fórmulas de un lenguaje formal $\mathcal{L}$, escribiremos $\Gamma \models \alpha$ si podemos razonar (informalmente) que la fórmula α tiene que ser verdadera en todo modelo de Γ.

Teorema 2.5 *Si F es un sistema deductivo formal correcto sobre un lenguaje formal $\mathcal{L}$, Γ es un conjunto de fórmulas de $\mathcal{L}$ y α es una fórmula tal que $\Gamma \underset{F}{\vdash} \alpha$, entonces $\Gamma \models \alpha$. En particular, todos los teoremas de F son lógicamente válidos.*

DEMOSTRACIÓN: Sea $\beta_1, \ldots, \beta_m$ una deducción de α en F a partir de las premisas. Supongamos que M es un modelo de $\mathcal{L}$ en el que son verdaderas las premisas y veamos que podemos razonar sucesivamente que cada β_i tiene que ser verdadera en M. En efecto, β_1 tiene que ser un axioma o una premisa. Si es un axioma, sabemos razonar que es verdadero en M porque es lógicamente válido, y si es una premisa es verdadera en M por hipótesis. Supuesto que hayamos razonado ya que $\beta_1, \ldots, \beta_i$ tienen que ser verdaderas en M, consideramos β_{i+1}. Si es un axioma o una premisa, razonamos como antes, y si es una consecuencia inmediata de fórmulas anteriores de la deducción, como las reglas de inferencia primitivas de F son reglas de inferencia semánticas, existe un razonamiento que justifica que β_{i+1} tiene que ser verdadera en M a partir del hecho de que las fórmulas precedentes lo son. Por lo tanto, podemos razonar que, en efecto, β_{i+1} es verdadera en M. De este modo llegamos a que $\beta_m \equiv \alpha$ es verdadera en M. ∎

De este modo, las deducciones en los sistemas deductivos formales correctos formalizan razonamientos informales válidos. Lo que buscamos en realidad es un sistema deductivo formal capaz de formalizar *todos* los razonamientos informales válidos (sobre un lenguaje formal fijo), lo cual no será fácil de justificar, pero la corrección es lo mínimo que podemos exigir para empezar. Nuestra propuesta es la siguiente:

El sistema deductivo formal $K_{\mathcal{L}}$ Dado un lenguaje formal $\mathcal{L}$, llamaremos $K_{\mathcal{L}}$ al sistema deductivo formal determinado por los siguientes axiomas y reglas de inferencia:

Axiomas de $K_{\mathcal{L}}$: Los axiomas de $K_{\mathcal{L}}$ son todas las fórmulas de los tipos siguientes, donde α, β, γ son fórmulas cualesquiera de $\mathcal{L}$ y t es un término cualquiera de $\mathcal{L}$.

K1	$\alpha \to (\beta \to \alpha)$	
K2	$(\alpha \to (\beta \to \gamma)) \to ((\alpha \to \beta) \to (\alpha \to \gamma))$	
K3	$(\neg\alpha \to \neg\beta) \to (\beta \to \alpha)$	
K4	$\bigwedge x_i \alpha \to \mathtt{S}^t_{x_i}\alpha$	
K5	$\bigwedge x_i(\alpha \to \beta) \to (\alpha \to \bigwedge x_i \beta)$	si x_i no está libre en α,
K6	$\bigwedge x_i(x_i = t \to \alpha) \leftrightarrow \mathtt{S}^t_{x_i}\alpha$	si x_i no está libre en t,

K7 $\bigvee^1 x_i\alpha \to \mathtt{S}_{x_i}^{x_i|\alpha}\alpha$ si $\mathcal{L}$ tiene descriptor,

K8 $\neg\bigvee^1 x_i\alpha \to x_i|\alpha = x_j|(x_j = x_j)$ si $\mathcal{L}$ tiene descriptor.

Según se indica, una fórmula de tipo K5 sólo es un axioma si se cumple la condición indicada sobre la variable x_i, y lo mismo es válido para las fórmulas de tipo K6.

Reglas de inferencia de $K_{\mathcal{L}}$:

Modus ponendo ponens (MP): de α y $\alpha \to \beta$ es consecuencia inmediata β.

Introducción del generalizador (IG): de α es consecuencia inmediata $\bigwedge x_i\alpha$.

■

Nota Observemos que $K_{\mathcal{L}}$ no tiene ocho axiomas, sino infinitos, pues, por ejemplo, $\alpha \to (\beta \to \alpha)$ no es un axioma de $K_{\mathcal{L}}$, sino un esquema de axioma tal que hay infinitas fórmulas de $\mathcal{L}$ que son axiomas de $K_{\mathcal{L}}$ por tener la forma indicada por este esquema.

Escribiremos $\Gamma \vdash \alpha$ y $\vdash \alpha$ en lugar de $\Gamma \underset{K_{\mathcal{L}}}{\vdash} \alpha$ y $\underset{K_{\mathcal{L}}}{\vdash} \alpha$.

Los axiomas y teoremas de $K_{\mathcal{L}}$ se llaman *axiomas y teoremas lógicos,* las consecuencias en $K_{\mathcal{L}}$ se llaman *consecuencias lógicas.*

Esta notación sugiere ya que $K_{\mathcal{L}}$ es el sistema deductivo formal que estamos buscando, aunque todavía no estamos en condiciones de probarlo. Lo que sí que podemos (y debemos) probar ya es su corrección:

Teorema 2.6 (Teorema de corrección) *El sistema deductivo formal $K_{\mathcal{L}}$ es correcto, es decir, si Γ es un conjunto de fórmulas de $\mathcal{L}$ y α es una fórmula de $\mathcal{L}$, si se cumple $\Gamma \vdash \alpha$, también $\Gamma \vDash \alpha$. En particular, todos los teoremas lógicos son fórmulas lógicamente válidas.*

Demostración: Por el teorema 2.5 sólo hay que probar que $K_{\mathcal{L}}$ es correcto, es decir, que los axiomas lógicos son lógicamente válidos y que las reglas de inferencia de $K_{\mathcal{L}}$ son reglas de inferencia semánticas. Lo segundo lo hemos probado ya en la sección precedente. En cuanto a lo primero, se trata de probar que disponemos de argumentos que nos convencen de que es imposible tener un modelo de un lenguaje $\mathcal{L}$ en el que alguno de los axiomas de $K_{\mathcal{L}}$ no sea verdadero.

Supongamos, pues, que M es un modelo de un lenguaje $\mathcal{L}$ y veamos que cualquier axioma ϕ de $K_{\mathcal{L}}$ ha de cumplir $M \vDash \phi$. A su vez, para ello fijamos una valoración v en M y trataremos de justificar que $M \vDash \phi[v]$.

Si ϕ es un axioma de tipo K1, K2 o K3 basta calcular sus tablas de verdad y comprobar que ninguno de ellos puede ser falso sean cuales sean los valores de verdad de las fórmulas α, β, γ que aparecen en ellos.

El axioma K4 está esencialmente comprobado en la sección anterior, donde vimos que

$$\bigwedge x\alpha \vDash \mathtt{S}_x^t\alpha$$

es una regla de inferencia semántica válida. En efecto, lo que tenemos que probar ahora es que

$$M \vDash (\bigwedge x\alpha \to \mathtt{S}^t_x\alpha)[v]$$

pero eso es tanto como comprobar que si $M \vDash \bigwedge x\alpha[v]$ entonces necesariamente $M \vDash \mathtt{S}^t_x\alpha[v]$, y eso es justo lo que comprobamos para justificar la regla de inferencia.

Si ϕ es de tipo K5 entonces $\phi \equiv \bigwedge x(\alpha \to \beta) \to (\alpha \to \bigwedge x\beta)$, y la variable x no está libre en α. Suponemos que $M \vDash \bigwedge x(\alpha \to \beta)[v]$ y hemos de probar que $M \vDash (\alpha \to \bigwedge x\beta)[v]$. A su vez, para ello suponemos que $M \vDash \alpha[v]$ y hemos de probar que $M \vDash \bigwedge x\beta[v]$.

Fijemos a en M. Tenemos que $M \vDash (\alpha \to \beta)[v^a_x]$ y, como x no está libre en α, por el teorema 1.9 también $M \vDash \alpha[v^a_x]$, luego $M \vDash \beta[v^a_x]$. Como esto se cumple para todo a de M, concluimos que $M \vDash \bigwedge x\beta[v]$, como queríamos probar.

Si ϕ es de tipo K6 su validez lógica está probada en la sección anterior.

Si ϕ es de tipo K7 entonces $\phi \equiv \overset{1}{\bigvee} x\alpha \to \mathtt{S}^{x|\alpha}_x\alpha$. Suponemos que $M \vDash \overset{1}{\bigvee} x\alpha[v]$, lo que significa que existe un único a en M tal que $M \vDash \alpha[v^a_x]$. Por consiguiente $a \equiv M(x|\alpha)[v]$ y así $M \vDash \alpha[v_x^{M(x|\alpha)[v]}]$. Por el teorema 1.12 $M \vDash \mathtt{S}^{x|\alpha}_x\alpha[v]$, como queríamos probar.

Si ϕ es de tipo K8 entonces $\phi \equiv \neg\overset{1}{\bigvee} x\alpha \to x|\alpha = z|(z = z)$. Suponemos que $M \vDash \neg\overset{1}{\bigvee} x\alpha[v]$, con lo que no existe un único a en M tal que $M \vDash \alpha[v^a_x]$. Por lo tanto $M(x|\alpha)[v] \equiv d \equiv M(z|(z = z))[v]$, luego $M \vDash (x|\alpha = z|(z = z))[v]$, como teníamos que probar. ∎

Para demostrar que $K_{\mathcal{L}}$ no sólo es correcto, sino que de hecho es capaz de formalizar todos los razonamientos informales, necesitamos primero "explorar" la capacidad de razonamiento de $K_{\mathcal{L}}$, es decir, hacernos una idea de qué cosas podemos demostrar en $K_{\mathcal{L}}$, para más adelante justificar (ya veremos cómo) que esa capacidad de $K_{\mathcal{L}}$ es suficiente, de hecho, para demostrar cualquier cosa que "debería" ser demostrable.

Veamos un primer ejemplo de demostración en $K_{\mathcal{L}}$:

Teorema 2.7 *Si α es una fórmula de $\mathcal{L}$, entonces $\vdash \alpha \to \alpha$.*

DEMOSTRACIÓN:

(1)	$(\alpha \to ((\alpha \to \alpha) \to \alpha)) \to ((\alpha \to (\alpha \to \alpha)) \to (\alpha \to \alpha))$	K2
(2)	$\alpha \to ((\alpha \to \alpha) \to \alpha)$	K1
(3)	$(\alpha \to (\alpha \to \alpha)) \to (\alpha \to \alpha)$	MP 1,2
(4)	$\alpha \to (\alpha \to \alpha)$	K1
(5)	$\alpha \to \alpha$	MP 3,4

∎

Observaciones En este punto hemos de señalar varios hechos:

— Nótese la numeración a la izquierda de las líneas y las anotaciones a la derecha que indican cómo se obtiene cada línea. Esto no lo exige la definición de demostración pero ayuda a leerla y entenderla, por lo que en adelante lo tomaremos por costumbre.

— El teorema anterior no es un teorema en el sentido que acabamos de definir, así como su demostración tampoco es una demostración en el sentido de 2.3. Lo que sigue a la palabra "teorema" no es una fórmula de $\mathcal{L}$, sino una afirmación *metamatemática* sobre $K_{\mathcal{L}}$, a saber, que todas las fórmulas del tipo $\alpha \to \alpha$ son teoremas de $K_{\mathcal{L}}$. Es lo que también se llama un *metateorema.* La diferencia entre un teorema formal y un metateorema es la misma que hay entre una partida de ajedrez y un resultado como "con un rey y un alfil no se puede dar mate a un rey solo". Esto último no es algo que se hace siguiendo reglas fijas, como una partida de ajedrez, sino que es algo que uno puede concluir analizando adecuadamente dichas reglas. Igualmente, analizando el concepto de demostración en $K_{\mathcal{L}}$ que hemos dado, podemos concluir que tiene la propiedad indicada. Esto no lo concluimos a partir de ciertos axiomas o reglas, sino que nos damos cuenta de ello informalmente, igual que podemos juzgar sobre qué podemos hacer al jugar a las cartas, o a cualquier otra cosa.[1]

En concreto, lo que hemos hecho es dar un *esquema de demostración,* de manera que si cogemos cualquier fórmula, como $x = y$, por ejemplo, y la ponemos allí donde pone α, obtenemos ciertamente una demostración de $x = y \to x = y$ en el sentido de 2.3.

Notemos que en el capítulo anterior hemos probado ya muchos metateoremas. Cualquier afirmación metamatemática que requiere una justificación es un metateorema, como, por ejemplo, el hecho de que una expresión no puede ser a la vez un término y una fórmula.

— Lo siguiente que debemos entender es que esto no es una demostración de que si α entonces α. Mejor dicho, técnicamente sí que es una demostración, pues $K_{\mathcal{L}}$ es correcto, por lo que la prueba anterior codifica un argumento informal que justifica que $\alpha \to \alpha$ es lógicamente válida, pero no es una demostración en el sentido "psicológico" de un argumento que realmente sirve de ayuda a alguien para entender por qué algo tiene que ser cierto. A nadie se le ocurriría recurrir a semejante "argumento" para justificar que $\alpha \to \alpha$.

Si tratamos de precisar esta primera impresión tratando de expresarla de forma más objetiva (menos "psicológica"), lo que podemos decir es que la conclusión es mucho más evidente que los axiomas que presuntamente la demuestran. Más concretamente, lo que a uno le convence de que $\alpha \to \alpha$ "es verdad" (más técnicamente, que es una fórmula lógicamente válida, verdadera en todo

[1]Sin perjuicio de que, como ya hemos comentado, si consideramos que los signos de un lenguaje formal son números naturales, entonces todas las afirmaciones formales sobre $K_{\mathcal{L}}$ (es decir, afirmaciones que no involucran modelos, como es el caso del metateorema anterior) se pueden interpretar como afirmaciones sobre números naturales, y sucede que, visto así, el metateorema anterior puede deducirse formalmente a partir de los axiomas de Peano, como veremos más adelante.

modelo) es considerar esta modesta tabla de verdad:

α	$\alpha \to \alpha$
V	V
F	V

En cambio, para probar que el primer axioma de la prueba es lógicamente válido, y por lo tanto un axioma legítimo, hay que construir esta tabla de verdad:

α	$\alpha \to \alpha$	$(\alpha \to \alpha) \to \alpha$	$(\alpha \to ((\alpha \to \alpha) \to \alpha))$	$\alpha \to (\alpha \to \alpha)$
V	V	V	V	V
F	V	F	V	V

$((\alpha\to(\alpha\to\alpha))\to(\alpha\to\alpha))$	$\alpha\to((\alpha\to\alpha)\to\alpha))\to((\alpha\to(\alpha\to\alpha))\to(\alpha\to\alpha))$
V	V
V	V

En este sentido podemos decir objetivamente que los axiomas de esta deducción son mucho más complicados que la conclusión. Y la idea que inevitablemente debería venirnos a la mente es si no estaremos haciendo el ridículo con $K_{\mathcal{L}}$. La respuesta es negativa, pero conviene entender por qué.

Una axiomática "tradicional" (por ejemplo una axiomática para la geometría euclídea) pretende reducir lógicamente afirmaciones complejas a otras lo más simples posibles que se toman como axiomas. De este modo, una deducción a partir de unos axiomas resulta explicativa. Normalmente los axiomas de una teoría axiomática se toman lo más simples posibles para que alguien que los vea pueda formarse una idea clara de lo que supone admitirlos como tales axiomas. Ya que los axiomas no se pueden demostrar, se procura al menos que se puedan juzgar fácilmente y sea igualmente fácil determinar bajo qué condiciones podemos esperar que se cumplan.

En cambio, $K_{\mathcal{L}}$ se ha definido siguiendo unos criterios que no tienen nada que ver con éstos. No hay necesidad de tomar axiomas simples porque los axiomas *se demuestran* (ya hemos demostrado que son lógicamente válidos), es decir, nadie necesita sopesarlos para decidir si "se los cree o no se los cree". Por ello, en lugar de la simplicidad, lo que se ha buscado en ellos es la potencia: que contengan la máxima información en el mínimo espacio.

Podríamos haber elegido otro conjunto de axiomas para $K_{\mathcal{L}}$ que fueran mucho más simples y "naturales", como $\alpha \to \alpha \vee \beta$, $\alpha \wedge \beta \to \alpha$, etc., pero el precio sería que en lugar de bastarnos con ocho esquemas de axioma habríamos necesitado unas dos docenas de ellos, más o menos. ¿Merece la pena comprimir el número de axiomas necesarios (de esquemas de axioma, en realidad) a costa de hacerlos "poco naturales"? La respuesta es afirmativa. Como ya hemos comentado, la naturalidad es poco importante, pues al fin y al cabo podemos demostrar que son lógicamente válidos, y eso los legitima como axiomas. Como contrapartida, hay resultados de la lógica matemática (del tipo "si se puede demostrar tal cosa también se puede demostrar tal otra cosa") que requieren razonar sobre los esquemas de axioma de $K_{\mathcal{L}}$ uno por uno, y entonces es muy

de agradecer que esa parte de la prueba pueda despacharse distinguiendo sólo ocho casos en lugar de veintiocho.

Con este criterio para elegir los axiomas no podemos ver las deducciones en $K_{\mathcal{L}}$ como explicativas, es decir, la prueba anterior no es una explicación de por qué es cierto (es lógicamente válido) que $\alpha \to \alpha$, sino que es una deducción a partir de axiomas lógicamente válidos (pero no obviamente lógicamente válidos) de una conclusión obviamente lógicamente válida. ¿Qué aporta entonces? De la deducción anterior podemos aprender algo interesante y no trivial, que no es que alfa implica alfa (eso es trivial y, por eso mismo, de interés dudoso) sino que alfa implica alfa es deducible en $K_{\mathcal{L}}$. Prueba de que esto no es trivial es que a pocos se les habría ocurrido cómo hacer la deducción. Y este hecho no trivial es sólo el primer paso en la prueba de otro hecho menos trivial aún, y es que todas las fórmulas lógicamente válidas son deducibles en $K_{\mathcal{L}}$.

Hemos presentado un conjunto extraño (pero comprimido) de axiomas, y queremos demostrar que esos "bichos raros" que hemos tomado como axiomas de entre todas las fórmulas lógicamente válidas son suficientes para deducir todas las demás. Pero esto no significa que todas las deducciones en $K_{\mathcal{L}}$ vayan a ser "monstruos" como el ejemplo anterior. Por el contrario, un uso "sensato" de $K_{\mathcal{L}}$ pasa por dos fases:

1) En una primera fase, lo que procede es "desembalar" la lógica que está ingeniosamente "embalada" en los axiomas de $K_{\mathcal{L}}$. Es como si alguien hubiera logrado embalar una mesa desmontada en piezas en un espacio tan pequeño y tan bien aprovechado que no deja un hueco libre y parece mentira que toda una mesa de ese tamaño quepa en una caja tan diminuta. Lo primero que procede hacer es abrir la caja, sacar las piezas y ensamblarlas para tener la mesa lista para ser usada (aunque luego nadie supiera volver a meterla en la caja sin dejar cuarenta piezas fuera). En la práctica, esto significa demostrar las dos docenas (o más) de resultados que habríamos tomado como axiomas o reglas de inferencia de $K_{\mathcal{L}}$ si no nos hubiera obsesionado reducir los axiomas a la mínima expresión. Tales demostraciones serán tremendamente "monstruosas" o, si se prefiere, tremendamente ingeniosas, pues son parte de la demostración de un teorema nada trivial, el que dice "toda la lógica cabe en esta reducidísima caja".

2) En segundo lugar, una vez demostrados esos resultados básicos que bien podrían haberse tomado como axiomas y reglas de inferencia de un sistema deductivo formal, lo que procede es olvidarse de los axiomas de $K_{\mathcal{L}}$ y de todas las demostraciones monstruosas, y no usarlos nunca más salvo para propósitos teóricos (es como tirar el embalaje de la mesa), y a partir de ese momento realizar las deducciones prácticas en $K_{\mathcal{L}}$ apoyándonos, no en los axiomas, sino en los resultados básicos deducidos de ellos que habrían sido axiomas si hubiéramos apostado por la naturalidad en vez de por la compacidad. Veremos que las posibilidades de razonamiento que estarán a nuestro alcance en esta "fase 2" no sólo serán "naturales", sino que se corresponderán exactamente con la forma en que los matemáticos razonan en su trabajo cotidiano.

Esto significa en particular, que si uno estudia lógica con la pretensión de familiarizarse con las técnicas de razonamiento matemático, no debe preocuparse

por ser capaz de generar demostraciones "monstruosas" del estilo de las que requiere la "fase 1", pues ésas sólo sirven para "desembalar la lógica", y eso basta hacerlo una vez siguiendo las "instrucciones de montaje". Las técnicas de razonamiento con las que uno debe familiarizarse son las que tendremos disponibles en la "fase 2", que son las que realmente utilizan los matemáticos.

Aunque, obviamente, todavía no estamos en condiciones de justificarla, vamos a ver a título ilustrativo (aunque sólo sea por contrarrestar la desagradable impresión que causa la demostración "fase 1" que hemos dado) qué aspecto tiene una deducción en $K_{\mathcal{L}}$ al estilo de la "fase 2".

Consideramos el lenguaje de la aritmética y consideramos en él la abreviatura (que leeremos "x divide a y") dada por $x \mid y \equiv \bigvee z\ y = xz$ y vamos a probar lo siguiente:

$$\bigwedge xyz((xy)z = x(yz)) \vdash \bigwedge xyz(x \mid y \wedge y \mid z \to x \mid z).$$

Mejor dicho, no vamos a dar una prueba, sino que vamos a dar lo que será una prueba cuando hayamos "desembalado" la lógica comprimida en $K_{\mathcal{L}}$. Para facilitar la comparación con la forma de razonar usual de los matemáticos incluimos una primera columna que contiene lo que diría un matemático al desarrollar la prueba, en lugar de las indicaciones técnicas de la última columna. Técnicamente podemos suprimir la primera columna sin dejar de tener por ello un argumento completo.

Suponemos que	(1)	$\bigwedge xyz((xy)z = x(yz))$	Premisa
Sean x, y, z arbitrarios tales que	(2)	$x \mid y \wedge y \mid z$	Hipótesis
En particular	(3)	$x \mid y$	EC 2
Por definición	(4)	$\bigvee u\ y = xu$	R 3
Fijamos, pues, un u tal que	(5)	$y = xu$	EP 4
De (2) se sigue también	(6)	$y \mid z$	EC 2
y por definición	(7)	$\bigvee u\ z = yu$	R 6
Fijamos v tal que	(8)	$z = yv$	EP 7
Sustituyendo (5) en (8) queda	(9)	$z = (xu)v$	ETI 5, 8
De (1) se sigue que	(10)	$(xu)v = x(uv)$	EG 1
luego de (9) y (10) se sigue	(11)	$z = x(uv)$	TI 9, 10
llamando u a uv queda	(12)	$\bigvee u\ z = xu$	IP 11
que por definición es	(13)	$x \mid z$	R 12
Con esto hemos probado que	(14)	$x \mid y \wedge y \mid z \to x \mid z$	
y como x, y, z eran arbitrarios	(15)	$\bigwedge xyz(x \mid y \wedge y \mid z \to x \mid z)$	IG 14

Los detalles que justifican que esto es realmente una deducción en $K_{\mathcal{L}}$ (o, mejor dicho, que esto justifica que la conclusión puede deducirse de la premisa en $K_{\mathcal{L}}$) los veremos en las secciones siguientes, pero de momento conviene hacer algunas observaciones sobre cómo argumenta el matemático, pues nuestro propósito es justificar que podemos imitar sus técnicas trabajando en $K_{\mathcal{L}}$.

Como decíamos, las dos primeras columnas constituyen lo que un matemático reconocería como una demostración válida (muy detallada) de la conclusión

a partir de la premisa. Aunque parece que sus explicaciones son suficientes para justificar la validez del razonamiento, lo cierto es que hay una serie de reglas que el matemático respeta implícitamente, incluso subconscientemente, y que, aunque no estén indicadas de forma explícita, son indispensables para que la prueba sea correcta.

Por ejemplo, el matemático distingue entre las variables x, y, z, que en la línea (2) usa para referirse a tres números *arbitrarios,* y las variables u y v, que introduce en (5) y (8) para referirse a dos números *particulares.* El matemático usa subconscientemente que no puede hacer lo mismo con una variable que represente a un objeto arbitrario que con otra que represente a un objeto particular. Por ejemplo, en la línea (12) escribe $\bigvee u$ y jamás habría pensado en escribir $\bigwedge u$, porque esa u representa al número particular uv, que es particular porque u y v lo eran. En cierto sentido, la posibilidad de escribir $\bigwedge u$ o sólo $\bigvee u$ depende para él de "la historia" de las variables, y no sólo de la línea (11), a partir de la cual se introduce el cuantificador.

Sin embargo, aunque, según decimos, el hecho de que u y v sean variables particulares prohíbe al matemático ligarlas con un generalizador, el hecho de que x, y, z sean variables generales no significa que pueda ligarlas en cualquier momento por un generalizador. Por ejemplo, al matemático jamás se le habría pasado por la cabeza escribir (3) $\bigwedge xyz(x \mid y \wedge y \mid z)$. Él lo razonaría así: "$x, y, z$ son tres números arbitrarios de los que supongo que $x \mid y \wedge y \mid z$, pero eso no significa que tres números cualesquiera deban cumplir esto, por lo que sería absurdo introducir ahí un generalizador. Sólo cuando llego a la línea (14) y tengo que si $x \mid y \wedge y \mid z$ también se cumple que $x \mid z$, sólo entonces puedo decir que esto vale para tres números arbitrarios, y eso justifica el paso a (15).

Más en general: cuando queremos probar que todo x que cumple A también cumple B, puedo tomar un x genérico que cumpla A, pero dicha generalidad no me permite afirmar que todo x cumple A. Sólo cuando pruebo la implicación $A \to B$ es cuando puedo decir que dicha implicación vale para todo x. Puedo generalizar respecto de x después, pero no antes de llegar a la implicación."

Otro hecho que el matemático tiene en cuenta instintivamente es que, aunque en la línea (7) tiene un $\bigvee u$, al eliminar el cuantificador está obligado a sustituir la variable u por otra variable nueva v, porque ya está usando la variable u para referirse al número que cumple (5) y no tiene por qué ser el mismo que cumpla (8). En cambio, cuando elimina los cuantificadores de (1) para escribir (10), no le importa que ya esté usando la variable x ni tampoco tiene inconveniente en sustituir las variables y, z por las variables u, v que ya está usando para nombrar otros objetos. La diferencia la marca que los cuantificadores de (1) son universales, por lo que las variables ligadas por ellos pueden sustituirse por cualquier objeto, aunque sea uno del que ya estemos hablando.

En general, podemos decir que el matemático aplica reglas "globales", es decir, reglas que tienen en cuenta la demostración en su conjunto y el punto de ella en el que estemos: no es lo mismo generalizar antes de haber probado la implicación que después, no es lo mismo quitar un $\bigvee u$ si antes ya hemos hablado de una u que si no, etc. Por el contrario, las reglas de inferencia de $K_{\mathcal{L}}$ son locales: dadas una o dos fórmulas, o de ellas se deduce algo o no se deduce, no

importa qué otras líneas haya en la deducción. Otro ejemplo de "globalidad" es que el matemático tiene claro que, en caso de prolongar la deducción, no podría usar ya que $x \mid y$, porque esto no es una premisa de la deducción, sino una hipótesis local que ha usado para probar la implicación (15), en caso de seguir usando la línea (3) estaría convirtiendo en premisa de su deducción lo que no era más que una premisa local, y no sería cierto que sus consecuencias fueran realmente consecuencias de la línea (1), que es la única premisa declarada.

Esto podría hacernos sospechar que $K_{\mathcal{L}}$ no es suficientemente potente como para formalizar razonamientos como el anterior, puesto que le falta la "sensibilidad al contexto" en la que constante y casi inconscientemente se apoya el matemático. Pese a las apariencias, veremos que no es así. Eso sí, para formalizar este tipo de razonamientos en $K_{\mathcal{L}}$ tendremos que explicitar todos los detalles como los que hemos comentado y que un matemático tiene en cuenta casi sin decirlo explícitamente, pues para que algo sea una auténtica deducción en $K_{\mathcal{L}}$ tiene que cumplir una serie de requisitos muy concretos totalmente explícitos, y eso es incompatible con asumir condiciones tácitas que el "sentido común" dicte en cada momento improvisadamente.

El segundo paso para "desembalar" la lógica de $K_{\mathcal{L}}$ (el primero ha sido demostrar que $\alpha \to \alpha$) es demostrar un potente teorema que vuelve trivial el primer paso (pero que lo requiere en su prueba):

Teorema 2.8 (Teorema de deducción) *Sean α y β fórmulas de $\mathcal{L}$ y sea Γ un conjunto de fórmulas de $\mathcal{L}$. Si $\Gamma, \alpha \vdash \beta$ y existe una deducción de β en la que no se generalice respecto a variables libres en α, entonces $\Gamma \vdash \alpha \to \beta$.*

En pocas palabras, lo que dice el teorema de deducción es que para deducir $\alpha \to \beta$ a partir de unas premisas, podemos añadir α como premisa y deducir β. Lo que obtendremos así no será una deducción de $\alpha \to \beta$ en $K_{\mathcal{L}}$, pero la demostración del teorema de deducción muestra cómo a partir de esta deducción es posible obtener una auténtica deducción de $\alpha \to \beta$ a partir de las premisas indicadas.

Notemos la restricción (que no puede ser pasada por alto) de que en la deducción de β no puede generalizarse respecto de variables libres en α. Esta limitación puede parecer extraña a un matemático, pero, como explicábamos más arriba, en realidad es una limitación que todos los matemáticos se autoimponen inconscientemente cuando se ven en una situación concreta de las descritas por el teorema anterior.

Por ejemplo, en la deducción que poníamos como ejemplo, desde el momento en que suponemos $x \mid y \wedge y \mid z$ queda prohibido generalizar respecto de x, y, z. Desde el momento en que introducimos esta hipótesis debemos romper nuestro convenio semántico de que una variable libre representa a un objeto arbitrario, y a partir de aquí x, y, z representan números arbitrarios de entre los que cumplan la hipótesis (y si generalizáramos como si fueran números totalmente arbitrarios estaríamos afirmando que todos los números cumplen la hipótesis, lo cual no tiene por qué ser cierto). Una vez hemos aplicado el teorema de deducción para escribir (14) el teorema nos asegura que esta línea es consecuencia de la premisa

(1), y a partir de aquí podemos "olvidarnos" de que hemos usado el teorema de deducción para probarlo (pues ya sabemos que existe una deducción "normal" de (14) a partir de (1)), y podemos generalizar para pasar a (15).

DEMOSTRACIÓN: Por hipótesis existe una deducción $\delta_1, \ldots, \delta_m$ con premisas en Γ y α tal que $\delta_m \equiv \beta$ y en la que no se generaliza respecto de variables libres en α.

Vamos a construir una deducción con premisas en Γ que contenga las fórmulas $\alpha \to \delta_i$ (con otras posibles fórmulas intercaladas). Como la última de estas fórmulas es $\alpha \to \beta$, con esto tendremos que la implicación es consecuencia de Γ.

Si δ_i es un axioma lógico o una premisa de Γ, la forma de incorporar $\alpha \to \delta_i$ a la deducción es la siguiente:

(1)	δ_i	axioma o premisa
(2)	$\delta_i \to (\alpha \to \delta_i)$	K1
(3)	$\alpha \to \delta_i$	MP 1, 2

Si $\delta_i \equiv \alpha$, entonces debemos incorporar a la deducción que estamos construyendo la fórmula $\alpha \to \alpha$, y esto es correcto porque ya hemos visto anteriormente cómo deducir esta fórmula en $K_{\mathcal{L}}$.

Si δ_i se deduce por MP, entonces hay dos fórmulas anteriores de la forma δ_j y $\delta_j \to \delta_i$. Puesto que en nuestra deducción vamos incorporando las implicaciones de forma sucesiva, cuando lleguemos a δ_i ya habremos incorporado las fórmulas $\alpha \to \delta_j$ y $\alpha \to (\delta_j \to \delta_i)$. Éstas son, pues, líneas anteriores de nuestra deducción y podemos usarlas para deducir $\alpha \to \delta_i$. Lo hacemos de este modo:

(1)	$\alpha \to \delta_j$	fórmula anterior
(2)	$\alpha \to (\delta_j \to \delta_i)$	fórmula anterior
(3)	$(\alpha \to (\delta_j \to \delta_i)) \to ((\alpha \to \delta_j) \to (\alpha \to \delta_i))$	K2
(4)	$(\alpha \to \delta_j) \to (\alpha \to \delta_i)$	MP 2, 3
(5)	$\alpha \to \delta_i$	MP 1, 4

Supongamos, por último que δ_i se deduce por generalización de otra fórmula anterior δ_j. Esto significa que $\delta_i \equiv \bigwedge x\delta_j$, y por la hipótesis del teorema sabemos que x no está libre en α. Entonces incorporamos $\alpha \to \bigwedge x\delta_j$ de este modo:

(1)	$\alpha \to \delta_j$	fórmula anterior
(2)	$\bigwedge x(\alpha \to \delta_j)$	IG 1
(3)	$\bigwedge x(\alpha \to \delta_j) \to (\alpha \to \bigwedge x\delta_j)$	K5 (porque x no está libre en α)
(4)	$\alpha \to \bigwedge x\delta_j$	MP 2, 3

Esto completa la prueba. ∎

De la demostración del teorema de deducción se sigue que en la deducción que se construye de $\alpha \to \beta$ se generaliza exactamente respecto de las mismas variables que en la deducción dada de β.

Ahora podemos ver que los axiomas K1, K2 y K5, que son los más técnicos de todos los axiomas de $K_{\mathcal{L}}$, son precisamente los necesarios para demostrar el teorema de deducción.

En la práctica podemos usar el teorema de deducción en una versión ligeramente más general. Imaginemos que estamos construyendo una deducción, digamos $\delta_1, \ldots, \delta_m$, a partir de unas premisas Γ, y a continuación queremos deducir una fórmula de tipo $\alpha \to \beta$. Entonces escribimos α y la usamos como una premisa más hasta obtener β (sin generalizar respecto de variables libres en α). Al llegar a β, lo que hemos probado es que $\delta_1, \ldots, \delta_m, \alpha \vdash \beta$, con una deducción en la que no se generalizan variables libres en α, luego el teorema de deducción nos da que $\delta_1, \ldots, \delta_m \vdash \alpha \to \beta$, es decir, que existe una deducción de $\alpha \to \beta$ con premisas en $\delta_1, \ldots, \delta_m$. Pero por otra parte sabemos que $\delta_1, \ldots, \delta_m$ se deducen de Γ, luego también $\Gamma \vdash \alpha \to \beta$. (Para tener una deducción que prueba esto deducimos cada δ_i de Γ y luego deducimos $\alpha \to \beta$ de las δ_i).

En la práctica marcaremos todas las líneas desde que suponemos α hasta que llegamos a β con una línea vertical (tal y como se ve en el ejemplo de la página 54). Esta línea advierte de que las fórmulas abarcadas por ella no son consecuencia de las premisas de la deducción principal, sino de las premisas más una hipótesis auxiliar (la fórmula α, que marcaremos con la etiqueta de "hipótesis") que sólo hemos aceptado provisionalmente para aplicar el teorema de deducción. Si una vez hemos añadido $\alpha \to \beta$ a la deducción prosiguiéramos haciendo uso de las líneas marcadas, la deducción sería inválida, pues estaríamos haciendo uso de una premisa α que no forma parte de las premisas de la deducción.

La observación de que para obtener la deducción cuya existencia afirma el teorema de deducción sólo se generaliza respecto de variables generalizadas en la prueba dada es esencial para que estemos seguros de que no estamos generalizando en un momento dado respecto de una variable "prohibida". El uso del teorema de deducción oculta el uso de algunos axiomas y reglas de inferencia (las que aparecen en la demostración al construir la deducción de $\alpha \to \beta$), pero no oculta ningún uso de IG.

Ejemplo Aunque hemos mostrado que la restricción sobre el uso de IG en el teorema de deducción es "razonable", podemos probar que es, de hecho, necesaria. Para ello veamos un ejemplo de falsa deducción en la que se usa el teorema de deducción sin respetar la restricción sobre el uso de IG. Concretamente, en la línea (2) generalizamos respecto de una variable libre en la hipótesis:

(1)	$x = y$	Hipótesis
(2)	$\bigwedge xy\ x = y$	IG 1 (dos veces)
(3)	$x = y \to \bigwedge xy\ x = y$	?
(4)	$\bigwedge y(x = y \to \bigwedge xy\ x = y)$	IG 3
(5)	$\bigwedge y(x = y \to \bigwedge xy\ x = y) \to S^x_y(x = y \to \bigwedge xy\ x = y)$	K4
(6)	$x = x \to \bigwedge xy\ x = y$	MP 4,5
(7)	$x = x$	
(8)	$\bigwedge xy\ x = y$	MP 6,7

La línea (7) es demostrable en $K_{\mathcal{L}}$, aunque posponemos la prueba hasta la sección siguiente (donde "desembalaremos" la lógica del igualador). Aceptando este hecho, podemos incorporarla en nuestra deducción.

Si el teorema de deducción fuera válido sin la restricción sobre el uso de IG, la deducción anterior sería correcta y podríamos concluir que $\vdash \bigwedge xy\ x = y$, luego por el teorema de corrección $\vDash \bigwedge xy\ x = y$, pero esta fórmula no es lógicamente válida. Al contrario, es falsa en todos los modelos del lenguaje formal considerado cuyo universo tenga más de un objeto. Esta contradicción prueba que la hipótesis sobre IG es necesaria en el teorema de deducción. ∎

El recíproco del teorema de deducción es inmediato:

Ejercicio: Probar que si $\Gamma \vdash \alpha \to \beta$, entonces $\Gamma, \alpha \vdash \beta$.

2.3 Reglas derivadas de inferencia

Llamaremos *reglas derivadas de inferencia* a los resultados de la forma

$$\alpha_1, \ldots, \alpha_n \vdash \alpha,$$

y las llamamos así porque una vez hemos comprobado que esto es cierto (es decir, hemos encontrado una deducción de α a partir de las premisas correspondientes) podemos usarlas en las deducciones como reglas de inferencia en pie de igualdad con las dos reglas primitivas MP e IG. En efecto, si en una deducción hemos escrito ya (entre otras) las líneas $\alpha_1, \ldots, \alpha_n$, podemos escribir α en cualquier momento que nos interese, aunque no se deduzca ni por MP ni por IG. Con esto nuestra deducción tendrá un "agujero", pero sabemos que ese agujero se puede "llenar" intercalando las líneas de la deducción de α a partir de $\alpha_1, \ldots, \alpha_n$ que ya conocemos.

Por ejemplo, en la deducción de la página 54, para pasar de la línea (2) a la (3) usamos la regla de inferencia derivada de eliminación del conjuntor, que en realidad son dos reglas de inferencia a las que no merece la pena dar nombres distintos:

$$\alpha \wedge \beta \vdash \alpha, \qquad \alpha \wedge \beta \vdash \beta.$$

Para que la deducción indicada estuviera completa haría falta (entre otras cosas) insertar la deducción en $K_{\mathcal{L}}$ de $x \mid y$ a partir de $x \mid y \wedge y \mid z$, pero si demostramos que las reglas anteriores valen en general para todas las fórmulas α y β podemos omitir las líneas necesarias para ello en nuestras deducciones con la garantía de que si quisiéramos una deducción completa, sin que falte una sola línea, sabríamos cómo añadir lo que hemos omitido.

Esto es lo que hacen los matemáticos cada vez que citan un teorema ya demostrado. Una demostración que en un momento dado diga "por el teorema del valor medio podemos afirmar que..." está incompleta, y sólo estaría completa si antes de decir "por el teorema del valor medio" insertáramos una demostración del teorema del valor medio, pero no ganaríamos nada con ello, al contrario, sólo volveríamos ilegibles todas las demostraciones.

La regla de eliminación del conjuntor, como todas las reglas de inferencia que vamos a demostrar aquí, son trivialmente reglas de inferencia semánticas, pero no es eso lo que tenemos que demostrar, sino que son deducibles en $K_{\mathcal{L}}$.

Cuando hayamos probado que todos los razonamientos informales son formalizables en $K_{\mathcal{L}}$ tendremos la garantía de que toda regla de inferencia semántica es deducible en $K_{\mathcal{L}}$, pero de momento tenemos que encontrar pruebas explícitas, las cuales pueden ser muy retorcidas y antinaturales. Ya hemos explicado cómo hay que entender esto.

Regla de repetición (R): $\alpha \vdash \alpha$.

Es inmediato que de α se deduce α, pues la propia α es una deducción de α con premisa α. Sin embargo, aquí queremos algo ligeramente más fuerte, y es que en una deducción podemos repetir una línea anterior si queremos. Esto ya no es evidente, pues α no es consecuencia de α por ninguna de las dos reglas de inferencia. No obstante, si tenemos α en una deducción, podemos escribir $\alpha \to \alpha$ por ser un teorema lógico y a continuación escribir α de nuevo por MP.

Esta regla es útil porque a menudo conviene cambiar el nombre con el que nos estamos refiriendo a una misma fórmula. Así lo hemos hecho, por ejemplo en la prueba de la página 54 al pasar de la línea (3) a la (4) que son la misma fórmula, pues (3) es por definición una abreviatura taquigráfica de (4).

Modus Barbara (MB): $\alpha \to \beta,\ \beta \to \gamma \vdash \alpha \to \gamma$.

Demostración: Esta regla admite una prueba natural gracias al teorema de deducción:

(1)	$\alpha \to \beta$	Premisa
(2)	$\beta \to \gamma$	Premisa
(3)	α	Hipótesis
(4)	β	MP 1, 3
(5)	γ	MP 2, 4
(6)	$\alpha \to \gamma$	

Las reglas siguientes nos proporcionan herramientas naturales para manipular fórmulas en función de los signos lógicos que aparecen en ellas. Empezamos con las relacionadas con el negador:

2.3.1 Reglas relacionadas con el negador y la implicación

Reglas de la doble negación (DN): $\neg\neg\alpha \vdash \alpha, \qquad \alpha \vdash \neg\neg\alpha$.

Demostración:

(1)	$\neg\neg\alpha$	Premisa
(2)	$\neg\neg\alpha \to (\neg\neg\neg\neg\alpha \to \neg\neg\alpha)$	K1
(3)	$\neg\neg\neg\neg\alpha \to \neg\neg\alpha$	MP 1, 2
(4)	$(\neg\neg\neg\neg\alpha \to \neg\neg\alpha) \to (\neg\alpha \to \neg\neg\neg\alpha)$	K3
(5)	$\neg\alpha \to \neg\neg\neg\alpha$	MP 3, 4
(6)	$(\neg\alpha \to \neg\neg\neg\alpha) \to (\neg\neg\alpha \to \alpha)$	K3
(7)	$\neg\neg\alpha \to \alpha$	MP 5, 6
(8)	α	MP 1, 7

Así pues, $\neg\neg\alpha \vdash \alpha$. Por el teorema de deducción $\vdash \neg\neg\alpha \to \alpha$. Esto vale para toda fórmula α. Aplicándolo a $\neg\alpha$ obtenemos que $\vdash \neg\neg\neg\alpha \to \neg\alpha$.

(1)	α	Premisa
(2)	$\neg\neg\neg\alpha \to \neg\alpha$	Teorema lógico
(3)	$(\neg\neg\neg\alpha \to \neg\alpha) \to (\alpha \to \neg\neg\alpha)$	K3
(4)	$\alpha \to \neg\neg\alpha$	MP 2, 3
(5)	$\neg\neg\alpha$	MP 1, 4

Por el teorema de deducción $\vdash \alpha \to \neg\neg\alpha$. También llamaremos DN a los teoremas $\vdash \alpha \to \neg\neg\alpha$ y $\vdash \neg\neg\alpha \to \alpha$.

Reglas de la negación de la implicación (NI):

$$\alpha \to \beta \vdash \neg\beta \to \neg\alpha \qquad \neg\beta \to \neg\alpha \vdash \alpha \to \beta$$

$$\alpha \to \neg\beta \vdash \beta \to \neg\alpha \qquad \neg\alpha \to \beta \vdash \neg\beta \to \alpha$$

Demostración:

(1)	$\neg\neg\alpha \to \alpha$	DN
(2)	$\alpha \to \beta$	Premisa
(3)	$\neg\neg\alpha \to \beta$	MB 1, 2
(4)	$\beta \to \neg\neg\beta$	DN
(5)	$\neg\neg\alpha \to \neg\neg\beta$	MB 3, 4
(6)	$(\neg\neg\alpha \to \neg\neg\beta) \to (\neg\beta \to \neg\alpha)$	K3
(7)	$\neg\beta \to \neg\alpha$	MP 5, 6

Las otras variantes se prueban de forma similar.

Modus tollendo tollens (MT): $\alpha \to \beta,\ \neg\beta \vdash \neg\alpha, \qquad \alpha \to \neg\beta,\ \beta \vdash \neg\alpha.$

(Por NI y el recíproco del teorema de deducción.)

Una regla muy especial es la siguiente:

Regla de la contradicción (C): $\alpha,\ \neg\alpha \vdash \beta$.

En palabras: a partir de una contradicción (es decir, si contamos con una fórmula y su negación como premisas) podemos deducir cualquier otra fórmula. La versión semántica de esta regla es peculiar, porque consiste en que β tiene que ser verdadera en todo modelo donde sean verdaderas α y $\neg\alpha$ sencillamente porque no puede haber modelos en los que esto suceda.

Demostración:

(1)	α	Premisa
(2)	$\alpha \to (\neg\beta \to \alpha)$	K1
(3)	$\neg\beta \to \alpha$	MP 1, 2
(4)	$\neg\alpha$	Premisa
(5)	$\neg\neg\beta$	MT 3, 4
(6)	β	DN 5

2.3.2 Reglas relacionadas con el disyuntor

Reglas de equivalencia entre disyunción e implicación (EDI):

$$\alpha \vee \beta \vdash \neg\alpha \to \beta \qquad \neg\alpha \to \beta \vdash \alpha \vee \beta.$$

Estas reglas son un caso particular de la regla de repetición, pues las dos fórmulas que relacionan son de hecho la misma, por la definición que hemos dado del disyuntor. Notemos que nos proporcionan una técnica que es útil a menudo para demostrar una disyunción: suponemos que no se cumple una de las fórmulas y usamos eso para demostrar la otra. El teorema de deducción nos da entonces $\neg\alpha \to \beta$ y esto equivale a $\alpha \vee \beta$. Recíprocamente, una forma de "aprovechar" una premisa que sea una disyunción (no es la única posible) consiste en llegar a que no se cumple una de las fórmulas y concluir que se tiene que cumplir la otra. Esto lo expresa de forma más explícita la regla siguiente:

Modus tollendo ponens (MTP): $\alpha \vee \beta,\ \neg\alpha \vdash \beta, \qquad \alpha \vee \beta,\ \neg\beta \vdash \alpha.$

Demostración:

(1)	$\alpha \vee \beta$	Premisa
(2)	$\neg\alpha \to \beta$	EDI 1
(3)	$\neg\alpha$	Premisa
(4)	β	MP 2, 3

La otra es similar.

Regla del tertium non datur (TND): $\vdash \alpha \vee \neg\alpha$.

Es un caso particular del teorema $\vdash \alpha \to \alpha$, pues $\alpha \vee \neg\alpha \equiv \neg\alpha \to \neg\alpha$.

Reglas de introducción del disyuntor (ID): $\alpha \vdash \alpha \vee \beta \qquad \alpha \vdash \beta \vee \alpha.$

Demostración: Por el teorema de deducción aplicado a (C) obtenemos que $\alpha \vdash \neg\alpha \to \beta$, o sea, $\alpha \vdash \alpha \vee \beta$.

(1)	α	Premisa
(2)	$\alpha \to (\neg\beta \to \alpha)$	K1
(3)	$\neg\beta \to \alpha$	MP 1, 2
(4)	$\beta \vee \alpha$	EDI 3

Regla de eliminación del disyuntor (ED): $\alpha \vee \alpha \vdash \alpha$.

Esta regla parece inofensiva, pero una demostración directa a partir de los resultados que tenemos disponibles es excesivamente "monstruosa". La presentamos aquí para tener agrupadas las reglas de forma coherente, pero posponemos la prueba hasta la subsección siguiente. El lector debe observar que en la subsección siguiente no usaremos ni (ED) ni la regla (Dil) que probamos a continuación.

También llamaremos (ED) al teorema $\vdash \alpha \vee \alpha \to \alpha$.

Con esta regla podemos probar otra técnica útil para sacarle partido a una disyunción en una deducción: si contamos con $\alpha \vee \beta$, una forma de concluir de ahí que se cumple una fórmula γ es probar que cada una de las dos fórmulas por separado implica γ, esto es lo que conoce como "distinguir casos":

Regla del dilema (Dil): $\alpha \to \gamma,\ \beta \to \gamma \vdash \alpha \vee \beta \to \gamma$.
DEMOSTRACIÓN:

(1)	$\alpha \to \gamma$	Premisa
(2)	$\beta \to \gamma$	Premisa
(3)	$\alpha \vee \beta$	Hipótesis
(4)	$\neg\alpha \to \beta$	EDI 3
(5)	$\neg\alpha \to \gamma$	MB 2,4
(6)	$\neg\gamma \to \neg\alpha$	NI 1
(7)	$\neg\gamma \to \gamma$	MB 5, 6
(8)	$\gamma \vee \gamma$	EDI 7
(9)	γ	ED 8
(10)	$(\alpha \vee \beta) \to \gamma$	

Un caso particular de argumento "distinguiendo casos" consiste en tomar cualquier fórmula α que sea relevante en la discusión y probar que tanto $\alpha \to \beta$ como $\neg\alpha \to \beta$. Así, como tenemos $\alpha \vee \neg\alpha$ por TND, podemos concluir β por la regla anterior.

2.3.3 Reglas relacionadas con el conjuntor

Leyes de De Morgan (DM):

$$\begin{array}{ll} \alpha \wedge \beta \vdash \neg(\neg\alpha \vee \neg\beta) & \neg(\neg\alpha \vee \neg\beta) \vdash \alpha \wedge \beta \\ \alpha \vee \beta \vdash \neg(\neg\alpha \wedge \neg\beta) & \neg(\neg\alpha \wedge \neg\beta) \vdash \alpha \vee \beta \\ \neg(\alpha \wedge \beta) \vdash \neg\alpha \vee \neg\beta & \neg\alpha \vee \neg\beta \vdash \neg(\alpha \wedge \beta) \\ \neg(\alpha \vee \beta) \vdash \neg\alpha \wedge \neg\beta & \neg\alpha \wedge \neg\beta \vdash \neg(\alpha \vee \beta) \end{array}$$

DEMOSTRACIÓN: Las dos primeras son casos particulares de (R), pues por definición $\alpha \wedge \beta \equiv \neg(\neg\alpha \vee \neg\beta)$.

(1)	$\alpha \vee \beta$	Premisa
(2)	$\neg\alpha \to \beta$	EDI 1
(3)	$\neg\beta \to \neg\neg\alpha$	NI 2
(4)	$\neg\neg\neg\alpha \to \neg\neg\beta$	NI 3
(5)	$\neg\neg\alpha \vee \neg\neg\beta$	EDI 4
(6)	$\neg\neg(\neg\neg\alpha \vee \neg\neg\beta)$	DN 5
(7)	$\neg(\neg\alpha \wedge \neg\beta)$	R 6

(1)	$\neg(\neg\alpha \wedge \neg\beta)$	Premisa
(2)	$\neg\neg(\neg\neg\alpha \vee \neg\neg\beta)$	R 1
(3)	$\neg\neg\alpha \vee \neg\neg\beta$	DN 2
(4)	$\neg\neg\neg\alpha \to \neg\neg\beta$	EDI 3
(5)	$\neg\beta \to \neg\neg\alpha$	NI 4
(6)	$\neg\alpha \to \beta$	NI 5
(7)	$\alpha \vee \beta$	EDI 6

Las restantes se siguen fácilmente de éstas.

Regla de no contradicción (NC): $\vdash \neg(\alpha \wedge \neg\alpha)$.

Demostración:

(1)	$\neg\alpha \vee \neg\neg\alpha$	TND
(2)	$\neg(\alpha \wedge \neg\alpha)$	DM 1

Regla de introducción del conjuntor (IC): $\alpha,\ \beta \vdash \alpha \wedge \beta$.

Demostración: Por el teorema de deducción sobre (MTP) se cumple

$$(*): \qquad \neg\neg\beta \vdash \neg\alpha \vee \neg\beta \to \neg\alpha.$$

(1)	β	Premisa
(2)	$\neg\neg\beta$	DN 1
(3)	$\neg\alpha \vee \neg\beta \to \neg\alpha$	$(*)$
(4)	$\neg\neg\alpha \to \neg(\neg\alpha \vee \neg\beta)$	NI 3
(5)	α	Premisa
(6)	$\neg\neg\alpha$	DN 5
(7)	$\neg(\neg\alpha \vee \neg\beta)$	MP 4, 6
(8)	$\alpha \wedge \beta$	DM 7

Reglas de eliminación del conjuntor (EC): $\alpha \wedge \beta \vdash \alpha, \qquad \alpha \wedge \beta \vdash \beta$.

Demostración: Partimos de la aplicación a ID del teorema de deducción:

(1)	$\neg\alpha \to \neg\alpha \vee \neg\beta$	ID
(2)	$\neg(\neg\alpha \vee \neg\beta) \to \alpha$	NI 1
(3)	$\alpha \wedge \beta$	Premisa
(4)	$\neg(\neg\alpha \vee \neg\beta)$	DM 3
(5)	α	MP 2, 4

Análogamente se prueba la otra. También llamaremos (EC) a los teoremas

$$\vdash \alpha \wedge \beta \to \alpha \qquad \vdash \alpha \wedge \beta \to \beta.$$

Ahora estamos en condiciones de demostrar fácilmente la regla (ED) que habíamos dejado pendiente en la subsección anterior:

Demostración:

(1)	$\neg\alpha \to \neg\alpha \wedge \neg\alpha$	IC
(2)	$\neg(\neg\alpha \wedge \neg\alpha) \to \alpha$	NI 1
(3)	$\alpha \vee \alpha$	Premisa
(4)	$\neg(\neg\alpha \wedge \neg\alpha)$	DM 3
(5)	α	MP 2, 4

Razonamiento por reducción al absurdo Con las reglas que tenemos justificadas hasta aquí es fácil justificar una técnica habitual de razonamiento: Si en una deducción tomamos $\neg\alpha$ como premisa adicional y, sin generalizar respecto de variables libres en α, llegamos a una contradicción $\beta \wedge \neg\beta$, podemos concluir α. Más precisamente, lo que estamos afirmando es que en esta situación:

(1)	γ_1	
$\vdots$		deducción a partir de unas premisas $\alpha_1, \ldots, \alpha_n$
(m)	γ_m	
$(m+1)$	$\neg\alpha$	Hipótesis
$\vdots$		deducción a partir de las premisas, las líneas precedentes y $\neg\alpha$
(k)	$\beta \wedge \neg\beta$	
$(k+1)$	α	Reducción al absurdo
$\vdots$		

si a partir del momento en que suponemos $\neg\alpha$ no generalizamos respecto a variables libres en α, la fórmula α escrita en la línea $k+1$ es deducible en $K_{\mathcal{L}}$ a partir de las premisas, pero las líneas marcadas con la raya vertical a la izquierda no lo son, porque suponen la hipótesis adicional $\neg\alpha$.

En efecto, lo que nos dice el teorema de deducción al llegar a la línea k es que $\neg\alpha \to (\beta \wedge \neg\beta)$ es deducible de las premisas, luego podríamos haber escrito esta fórmula en la línea $k+1$, y la deducción podría haber continuado así:

$(k+1)$	$\neg\alpha \to (\beta \wedge \neg\beta)$	
$(k+2)$	$\neg(\beta \wedge \neg\beta)$	NC
$(k+3)$	$\neg\neg\alpha$	MT $k+1, k+2$
$(k+4)$	α	DN $k+3$

En la práctica suprimiremos estas líneas y escribiremos directamente α tras haber llegado a una contradicción. Tampoco es necesario reunir en una única fórmula la contradicción $\beta \wedge \neg\beta$, sino que basta haber obtenido dos líneas contradictorias (una la negación de la otra). Podemos omitir la aplicación de IC para reunirlas en una. Igualmente, usando DN vemos que si suponemos α y llegamos a una contradicción podemos concluir $\neg\alpha$.

Nuevamente, la restricción sobre la generalización es algo que todo matemático respetará de forma instintiva en cada caso concreto. Por ejemplo, supongamos que, en el curso de una prueba, un matemático quiere probar que un cierto número p que está considerando es primo y se propone hacerlo por reducción al absurdo. Jamás razonaría así:

(1)	$\neg p$ es primo	Reducción al absurdo
(2)	$\bigwedge p \neg p$ es primo	Aplicación ilícita de IG
(3)	$\neg\neg\bigwedge p \neg p$ es primo	DN 2
(4)	$\neg\bigvee p\ p$ es primo	R 3
(5)	$\bigvee p\, p$ es primo	resultado probado anteriormente.
(6)	p es primo	Por la contradicción de 4 y 5.

Esto en realidad no prueba nada. Podemos suponer para llegar a un absurdo que p no es primo, pero no podemos pasar de ahí a que ningún p es primo (que es lo que obtenemos si generalizamos ilegalmente respecto de p).

2.3.4 Reglas relacionadas con el bicondicionador

Reglas de introducción y eliminación del bicondicionador

$$\begin{array}{ll}(\mathrm{IB}) & \alpha \to \beta,\ \beta \to \alpha \vdash \alpha \leftrightarrow \beta.\\ (\mathrm{EB}) & \alpha \leftrightarrow \beta \vdash \alpha \to \beta, \quad \alpha \leftrightarrow \beta \vdash \beta \to \alpha\end{array}$$

Son casos particulares de (IC), (EC), pues $\alpha \leftrightarrow \beta \equiv (\alpha \to \beta) \wedge (\beta \to \alpha)$.

2.3.5 Reglas relacionadas con los cuantificadores

Regla de eliminación del generalizador (EG): $\bigwedge x\alpha \vdash \mathtt{S}^t_x\alpha$.

Por el axioma (K4) y el recíproco del teorema de deducción.

Reglas de negación del generalizador (NG):

$$\begin{array}{ll}\neg\bigwedge x\neg\alpha \vdash \bigvee x\alpha & \bigvee x\alpha \vdash \neg\bigwedge x\neg\alpha\\ \neg\bigwedge x\alpha \vdash \bigvee x\neg\alpha & \bigvee x\neg\alpha \vdash \neg\bigwedge x\alpha\end{array}$$

Demostración: Las dos primeras son casos particulares de (R), pues $\bigvee x\alpha \equiv \neg\bigwedge x\neg\alpha$. Las demás se siguen de la aplicación oportuna de DN:

(1)	$\neg\bigwedge x\alpha$	Premisa
(2)	$\neg\bigvee x\neg\alpha$	Hipótesis (reducción al absurdo)
(3)	$\neg\neg\bigwedge x\neg\neg\alpha$	R 2
(4)	$\bigwedge x\neg\neg\alpha$	DN 3
(5)	$\neg\neg\alpha$	EG 4
(6)	α	DN 5
(7)	$\bigwedge x\alpha$	IG 6 (contradicción con 1)
(8)	$\bigvee x\neg\alpha$	

Observemos que desde el momento en que suponemos (2) está prohibido generalizar respecto de variables libres en (2), pero luego sólo generalizamos respecto de x, que no está libre en (2).

(1)	$\bigvee x\neg\alpha$	Premisa
(2)	$\bigwedge x\alpha$	Hipótesis (reducción al absurdo)
(3)	α	EG 2
(4)	$\neg\neg\alpha$	DN 3
(5)	$\bigwedge x\neg\neg\alpha$	IG 4
(6)	$\neg\neg\bigwedge x\neg\neg\alpha$	DN 5
(7)	$\neg\bigvee x\neg\alpha$	R 6 (contradicción con 1)
(8)	$\neg\bigwedge x\alpha$	

Nuevamente observamos que el uso de IG es legítimo.

Reglas de negación del particularizador (NP):

$$\neg\bigvee x\alpha \vdash \bigwedge x\neg\alpha \qquad \bigwedge x\neg\alpha \vdash \neg\bigvee x\alpha$$
$$\neg\bigvee x\neg\alpha \vdash \bigwedge x\alpha \qquad \bigwedge x\alpha \vdash \neg\bigvee x\neg\alpha$$

DEMOSTRACIÓN:

(1)	$\neg\bigvee x\alpha$	Premisa	(1)	$\bigwedge x\neg\alpha$	Premisa
(2)	$\neg\neg\bigwedge x\neg\alpha$	R 1	(2)	$\neg\neg\bigwedge x\neg\alpha$	DN 1
(3)	$\bigwedge x\neg\alpha$	DN 2	(3)	$\neg\bigvee x\alpha$	R 2

(1)	$\neg\bigvee x\neg\alpha$	Premisa	(1)	$\bigwedge x\alpha$	Premisa
(2)	$\neg\neg\bigwedge x\neg\neg\alpha$	R1	(2)	α	EG 1
(3)	$\bigwedge x\neg\neg\alpha$	DN 2	(3)	$\neg\neg\alpha$	DN 2
(4)	$\neg\neg\alpha$	EG 3	(4)	$\bigwedge x\neg\neg\alpha$	IG 3
(5)	α	DN 4	(5)	$\neg\neg\bigwedge x\neg\neg\alpha$	DN 4
(6)	$\bigwedge x\alpha$	IG 5	(6)	$\neg\bigvee x\neg\alpha$	R5

Regla de introducción del particularizador (IP): $\mathtt{S}^t_x\alpha \vdash \bigvee x\alpha$.

Esta regla afirma que si hemos probado que un cierto término t cumple lo que dice α, entonces podemos afirmar que existe un x que cumple α. Tenemos un ejemplo de su uso en la línea 12 de la deducción de la página 54.

DEMOSTRACIÓN:

(1)	$\mathtt{S}^t_x\alpha$	Premisa
(2)	$\neg\bigvee x\alpha$	Hipótesis (reducción al absurdo)
(3)	$\bigwedge x\neg\alpha$	NP 2
(4)	$\mathtt{S}^t_x\neg\alpha$	EG 3
(5)	$\neg\mathtt{S}^t_x\alpha$	R4 (contradicción con 1)
(6)	$\bigvee x\alpha$	

Es importante destacar que la prueba de esta regla que liga la variable x no usa IG. Así, si queremos ligar una variable en un contexto en el que no está permitido usar IG, siempre podemos usar IP.

Por último nos falta una "regla de eliminación del particularizador", que es la que aplica un matemático cuando tiene que $\bigvee x\alpha$ y dice "tomemos un x que cumpla α". Tenemos dos ejemplos en las líneas 5 y 8 de la deducción de la página 54. Ahora bien, esta "regla" no es realmente una regla derivada de inferencia, sino un (meta)teorema análogo al teorema de deducción. Para probarlo necesitamos un hecho previo que tiene interés en sí mismo:

Teorema 2.9 *Se cumplen los siguientes hechos:*

1. *Si y no está en α, entonces* $\vdash \bigwedge x\alpha \leftrightarrow \bigwedge y\ \mathtt{S}^y_x\alpha$.
2. *Si y no está en α, entonces* $\vdash \bigvee x\alpha \leftrightarrow \bigvee y\ \mathtt{S}^y_x\alpha$.
3. *Si $y_1, \ldots, y_n$ son variables distintas que no están en α y son distintas de $x_1, \ldots, x_n$, entonces*
   $\vdash \bigvee x_1 \cdots x_n\alpha \leftrightarrow \bigvee y_1 \cdots y_n \bigwedge x_1 \cdots x_n(\alpha \leftrightarrow y_1 = x_1 \wedge \cdots \wedge y_n = x_n)$.

DEMOSTRACIÓN: Para probar 1), suponemos $\bigwedge x\alpha$, pasamos a $\mathtt{S}^y_x\alpha$ por (EG) y pasamos a $\bigwedge y\mathtt{S}^y_x\alpha$ por (IG). Esto nos da una implicación. La otra se prueba igualmente usando que $\mathtt{S}^x_y\mathtt{S}^y_x\alpha \equiv \alpha$ (teorema 1.13).

2) sale de 1) aplicado a $\neg\alpha$.

Para probar 3) observamos que, por definición,

$$\overset{1}{\bigvee}x_1\cdots x_n\,\alpha \equiv \bigvee z_1\cdots z_n\bigwedge x_1\cdots x_n(\alpha \leftrightarrow z_1 = x_1 \wedge\cdots\wedge z_n = x_n),$$

donde $z_1,\ldots,z_n$ son variables que no estén en α y sean distintas de $x_1,\ldots,x_n$. Lo que queremos probar es que si cambiamos z_i por y_i obtenemos una fórmula equivalente, y esto es un caso particular de 2). ∎

Regla de eliminación del particularizador (EP): En esta situación:

(1)	γ_1	
⋮		deducción a partir de unas premisas $\alpha_1,\ldots,\alpha_n$
(k)	$\bigvee x\alpha$	
⋮		
(m)	γ_m	
$(m+1)$	$\mathtt{S}^y_x\alpha$	EP k
⋮		deducción a partir de las premisas, las líneas precedentes y $\mathtt{S}^y_x\alpha$

Si la variable y no está en α o bien $y \equiv x$ (en cuyo caso $\mathtt{S}^y_x\alpha \equiv \alpha$) y a partir de la línea $m+1$ no se generaliza respecto de variables libres en $\mathtt{S}^y_x\alpha$, entonces toda línea posterior β que no tenga libre la variable y es una consecuencia de las premisas.

DEMOSTRACIÓN: En principio, en la situación descrita tenemos que

$$\alpha_1,\ldots,\alpha_n,\mathtt{S}^y_x\alpha \vdash \beta.$$

Ahora bien, como para obtener β no se ha generalizado respecto de ninguna variable libre en $\mathtt{S}^y_x\alpha$, podemos aplicar el teorema de deducción y concluir que

$$\alpha_1,\ldots,\alpha_n \vdash \mathtt{S}^y_x\alpha \to \beta.$$

Pero también sabemos que $\alpha_1,\ldots,\alpha_n \vdash \bigvee x\alpha$, y por el teorema anterior

$$\alpha_1,\ldots,\alpha_n \vdash \bigvee y\mathtt{S}^y_x\alpha.$$

Por lo tanto, sólo necesitamos probar que $\mathtt{S}^y_x\alpha \to \beta, \bigvee yS^y_x\alpha \vdash \beta$. En efecto:

(1)	$\mathtt{S}^y_x\alpha \to \beta$	Premisa
(2)	$\bigvee y\mathtt{S}^y_x\alpha$	Premisa
(3)	$\neg\beta$	Hipótesis (reducción al absurdo)
(4)	$\neg\mathtt{S}^y_x\alpha$	MT 1,3
(5)	$\bigwedge y\neg\mathtt{S}^y_x\alpha$	IG 4
(6)	$\neg\bigvee y\mathtt{S}^y_x\alpha$	NP 5 (contradicción con 2)
(7)	β	

Es muy importante observar que en la deducción de β a partir de las premisas se generaliza respecto de la variable y. Esto significa que si aplicamos EP en un contexto en el que tenemos prohibido generalizar respecto de ciertas variables, debemos elegir la variable y como una nueva variable sobre la que no exista prohibición de generalizar. Una vez más, esto es algo que el matemático hace instintivamente. Consideremos de nuevo el ejemplo de la página 54:

Desde la línea 2, tenemos prohibido generalizar respecto de x, y, z. En la línea 5 eliminamos un particularizador $\bigvee u$, y no cambiamos de variable porque no hay problema en generalizar respecto de u. A partir de esa línea tenemos prohibido generalizar respecto de u, por lo que, cuando queremos eliminar el $\bigvee u$ de la línea 7 nos vemos obligados a sustituir la variable u por una nueva variable v que tiene que ser distinta de x, y, z, u, que son las variables respecto a las que tenemos prohibido generalizar. El matemático hace esto instintivamente cuando piensa que no puede tomar u tal que $z = yu$ porque ya está llamando u a otro número, y mucho menos se le ocurriría llamarlo x, y, z por el mismo motivo, porque sabe que el u que existe por 7 no tiene por qué ser el mismo que cualquiera de los números que está considerando hasta entonces.

Al introducir el particularizador en 12 dejamos de tener libres las variables u, v, por lo que todas las fórmulas que siguen son ya auténticas consecuencias de las premisas (y de la hipótesis 2), y podemos aplicar el teorema de deducción para pasar a 14 (hubiera sido incorrecto hacerlo con una fórmula que tuviera libre la variable u o la variable v).

Nótese la sutileza: a partir del momento en que suponemos $y = xu$ queda prohibido generalizar respecto de u, pero la deducción de $x \mid z$ a partir de las premisas y la hipótesis (2) usa la regla IG respecto de u, es decir, tenemos prohibido generalizar, pero a la vez la prueba completa que elimina la premisa adicional $y = xu$ generaliza respecto de u.

Como al eliminar un particularizador dejamos libre una variable que no puede quedar libre en la conclusión, una forma de volver a ligarla es mediante la regla de introducción del particularizador (IP), porque la regla de introducción del generalizador la tenemos prohibida. Aquí es fundamental recordar que en la demostración de IP no se generaliza respecto a la variable que particularizamos.

Notemos que es "de sentido común": si una variable procede de eliminar un $\bigvee u$, luego no podemos ligarla con un $\bigwedge u$, sino que tendremos que volver a introducir un particularizador.

Veamos también un ejemplo de la importancia de no generalizar respecto de variables libres en $\mathtt{S}^y_x\alpha$. Por ejemplo, partamos de la fórmula $\bigwedge x \bigvee y\, x \leq y$, que es verdadera si la interpretamos en el modelo que tiene por universo los números naturales y en la que el relator $\leq$ se interpreta como la relación de orden usual. Si no tenemos en cuenta la restricción indicada podríamos "demostrar" lo siguiente:

(1)	$\bigwedge x \bigvee y\, x \leq y$	Premisa
(2)	$\bigvee y\, x \leq y$	EG 1
(3)	$x \leq y$	EP 2
(4)	$\bigwedge x\, x \leq y$	IG 3
(5)	$\bigvee y \bigwedge x\, x \leq y$	IP 4

Así, a partir de una afirmación verdadera sobre los números naturales (todo número natural tiene otro posterior), hemos "demostrado" una afirmación falsa: (hay un número natural mayor que todos los demás). La falacia está en la línea (4), pues desde el momento en que hemos eliminado el particularizador en (3) ya no podemos generalizar respecto de x, y. ∎

2.3.6 Reglas relacionadas con el igualador

Reglas de introducción y eliminación del igualador

(II) $\mathtt{S}_x^t\alpha \vdash \bigwedge x(x = t \to \alpha)$, si x no está libre en t.
(EI) $\bigwedge x(x = t \to \alpha) \vdash \mathtt{S}_x^t\alpha$, si x no está libre en t.

Por (K6) y (EB).

Regla de la identidad (I) $\vdash t = t$.

Sea x una variable que no esté libre en t.

(1)	$x = t \to x = t$	Teorema lógico
(2)	$\bigwedge x(x = t \to x = t)$	IG 1
(3)	$\mathtt{S}_x^t(x = t)$	EI 2
(4)	$t = t$	R 3

Regla de la simetría de la identidad (SI): $t_1 = t_2 \vdash t_2 = t_1$.

DEMOSTRACIÓN: Sea x una variable que no esté en t_1 ni en t_2.

(1)	$t_2 = t_2$	I
(2)	$\mathtt{S}_x^{t_2}(t_2 = x)$	R 1
(3)	$\bigwedge x(x = t_2 \to t_2 = x)$	II 2
(4)	$t_1 = t_2 \to t_2 = t_1$	EG 3
(5)	$t_1 = t_2$	Premisa
(6)	$t_2 = t_1$	MP 4, 5

Regla de la transitividad de la identidad (TI): $t_1 = t_2,\ t_2 = t_3 \vdash t_1 = t_3$.

DEMOSTRACIÓN: Sea x una variable que no esté libre en t_1, t_2, t_3.

(1)	$t_2 = t_3$	Premisa
(2)	$\bigwedge x(x = t_2 \to x = t_3)$	II 1
(3)	$t_1 = t_2 \to t_1 = t_3$	EG 2
(4)	$t_1 = t_2$	Premisa
(5)	$t_1 = t_3$	MP 3, 4

Regla de equivalencia entre términos idénticos (ETI):

$$t_1 = t_2,\ \mathtt{S}_x^{t_2}\alpha \vdash \mathtt{S}_x^{t_1}\alpha, \qquad t_1 = t_2 \vdash \mathtt{S}_x^{t_1}t = \mathtt{S}_x^{t_2}t.$$

DEMOSTRACIÓN: La prueba es muy simple si suponemos que x no está libre en t_1, t_2, pero para tratar el caso general tomamos una variable y que no esté en α, t, t_1, t_2 y probamos lo siguiente:

⎪	(1)	$x = y$	Hipótesis
⎪⎪	(2)	$\alpha \equiv \mathtt{S}_y^x\mathtt{S}_x^y\alpha$	Hipótesis
⎪⎪	(3)	$\bigwedge y(y = x \to \mathtt{S}_x^y\alpha)$	II 2
⎪⎪	(4)	$y = x \to \mathtt{S}_x^y\alpha$	EG 3
⎪⎪	(5)	$y = x$	SI 1
⎪⎪	(6)	$\mathtt{S}_x^y\alpha$	MP 4, 5
⎪	(7)	$\alpha \to \mathtt{S}_x^y\alpha$	
⎪⎪	(8)	$\mathtt{S}_x^y\alpha$	Hipótesis
⎪⎪	(9)	$\bigwedge x(x = y \to \alpha)$	II 8
⎪⎪	(10)	$x = y \to \alpha$	EG 9
⎪⎪	(11)	α	MP 1, 10
⎪	(12)	$\mathtt{S}_x^y\alpha \to \alpha$	
⎪	(13)	$\alpha \leftrightarrow \mathtt{S}_x^y\alpha$	IB 7, 12
	(14)	$x = y \to (\alpha \leftrightarrow \mathtt{S}_x^y\alpha)$	
	(15)	$\bigwedge xy(x = y \to (\alpha \leftrightarrow \mathtt{S}_x^y\alpha))$	IG 14
	(16)	$\bigwedge y(t = y \to (\mathtt{S}_x^t\alpha \leftrightarrow \mathtt{S}_x^y\alpha))$	EG 15
	(17)	$t = t \to (\mathtt{S}_x^t\alpha \leftrightarrow \mathtt{S}_y^t\mathtt{S}_x^y\alpha)$	EG 16
	(18)	$t = t$	I
	(19)	$\mathtt{S}_x^t\alpha \leftrightarrow \mathtt{S}_y^t\mathtt{S}_x^y\alpha$	MP 17,18

Aplicando esto a $\alpha \equiv z = t'$ obtenemos: $\vdash z = \mathtt{S}_x^t t' \to z = \mathtt{S}_y^t\mathtt{S}_x^y t'$, donde z es cualquier variable que no esté en t, t', x, y. A su vez:

(1)	$z = \mathtt{S}_x^t t' \to z = \mathtt{S}_y^t\mathtt{S}_x^y t'$	Teorema
(2)	$\bigwedge z(z = \mathtt{S}_x^t t' \to z = \mathtt{S}_y^t\mathtt{S}_x^y t')$	IG 1
(3)	$\mathtt{S}_x^t t' = S_x^t t' \to \mathtt{S}_x^t t' = \mathtt{S}_y^t\mathtt{S}_x^y t'$	EG 2
(4)	$\mathtt{S}_x^t t' = \mathtt{S}_x^t t'$	I
(5)	$\mathtt{S}_x^t t' = \mathtt{S}_y^t\mathtt{S}_x^y t'$	MP 3, 4

Con esto ya podemos probar ETI:

(1)	$\mathtt{S}_x^{t_2}\alpha$	Premisa	(1)	$S_x^{t_2}t = \mathtt{S}_y^{t_2}\mathtt{S}_x^y t$	Teorema
(2)	$\mathtt{S}_x^{t_2}\alpha \to \mathtt{S}_y^{t_2}\mathtt{S}_x^y\alpha$	Teorema	(2)	$\mathtt{S}_y^{t_2}\mathtt{S}_x^y t = S_x^{t_2}t$	SI 1
(3)	$\mathtt{S}_y^{t_2}\mathtt{S}_x^y\alpha$	MP 1, 2	(3)	$\mathtt{S}_y^{t_2}(\mathtt{S}_x^y t = \mathtt{S}_x^{t_2}t)$	R 2
(4)	$\bigwedge y(y = t_2 \to \mathtt{S}_x^y\alpha)$	II 3	(4)	$\bigwedge y(y = t_2 \to \mathtt{S}_x^y t = \mathtt{S}_x^{t_2}t)$	II 3
(5)	$t_1 = t_2 \to \mathtt{S}_y^{t_1}\mathtt{S}_x^y\alpha$	EG 4	(5)	$t_1 = t_2 \to \mathtt{S}_y^{t_1}\mathtt{S}_x^y t = \mathtt{S}_x^{t_2}t$	EG 4
(6)	$t_1 = t_2$	Premisa	(6)	$t_1 = t_2$	Premisa
(7)	$\mathtt{S}_y^{t_1}\mathtt{S}_x^y\alpha$	MP 5, 6	(7)	$\mathtt{S}_y^{t_1}\mathtt{S}_x^y t = \mathtt{S}_x^{t_2}t$	MP 5,6
(8)	$\mathtt{S}_y^{t_1}\mathtt{S}_x^y\alpha \to \mathtt{S}_x^{t_1}\alpha$	Teorema	(8)	$\mathtt{S}_x^{t_1}t = \mathtt{S}_y^{t_1}\mathtt{S}_x^y t$	Teorema
(9)	$\mathtt{S}_x^{t_1}\alpha$	MP 7, 8	(9)	$\mathtt{S}_x^{t_1}t = \mathtt{S}_x^{t_2}t$	TI 6, 7

Lo que afirma esta regla es que si tenemos que $t_1 = t_2$ y que t_2 cumple lo que afirma α, entonces lo mismo vale para t_1 o, más en general, que si tenemos $t_1 = t_2$ entonces todo lo que sepamos de t_2 vale también para t_1 (y viceversa, por SI).

2.3.7 Reglas relacionadas con el descriptor

En el capítulo siguiente estudiaremos con detalle el uso del descriptor. De momento nos limitamos a enunciar en forma de reglas de inferencia los axiomas K7 y K8:

Regla de las descripciones propias (DP): $\bigvee^1 x\alpha \vdash \mathtt{S}_x^{x|\alpha}\alpha$.

Regla de las descripciones impropias (DI): $\neg\bigvee^1 x\alpha \vdash x|\alpha = y|(y = y)$.

Se siguen de (K7), (K8) y el recíproco del teorema de deducción.

Notemos que todos los casos de la regla (IG) que aparecen en las pruebas de las reglas de inferencia se aplican a variables que no están libres en las premisas. Esto quiere decir que todas ellas pueden ser usadas incluso en contextos en los que no sea lícito generalizar respecto de ciertas variables, como cuando se aplica el teorema de deducción.

2.4 Algunos teoremas lógicos

Aunque con las reglas de inferencia que hemos deducido ya estamos en condiciones de deducir cualquier cosa de forma "natural", no está de más contar con algunos resultados ya probados con los que podemos contar sin tener que deducirlos cada vez que sea necesario recurrir a ellos. En esta sección presentaremos unos cuantos, además de dar algunos ejemplos ilustrativos de deducciones formales.

El álgebra del cálculo proposicional Empezamos demostrando las propiedades "booleanas" del cálculo deductivo. Las primeras de ellas son las propiedades asociativas de la conjunción y la disyunción, que justifican que de aquí en adelante no pongamos paréntesis en expresiones de la forma

$$\alpha_1 \wedge \cdots \wedge \alpha_n \qquad \text{o} \qquad \alpha_1 \vee \cdots \vee \alpha_n,$$

pues distintas disposiciones de paréntesis dan lugar a fórmulas equivalentes.

Teorema 2.10 *Las fórmulas siguientes son teoremas lógicos:*

1)	$\alpha \wedge \beta \leftrightarrow \beta \wedge \alpha$,	$\alpha \vee \beta \leftrightarrow \beta \vee \alpha$,
2)	$(\alpha \wedge \beta) \wedge \gamma \leftrightarrow \alpha \wedge (\beta \wedge \gamma)$,	$(\alpha \vee \beta) \vee \gamma \leftrightarrow \alpha \vee (\beta \vee \gamma)$,
3)	$\alpha \wedge \alpha \leftrightarrow \alpha$,	$\alpha \vee \alpha \leftrightarrow \alpha$,
4)	$\alpha \wedge (\beta \vee \gamma) \leftrightarrow (\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$,	$\alpha \vee (\beta \wedge \gamma) \leftrightarrow (\alpha \vee \beta) \wedge (\alpha \vee \gamma)$.

DEMOSTRACIÓN: Veamos únicamente un par de ellas, y dejamos el resto como ejercicio para el lector:

|	(1)	$(\alpha \vee \beta) \vee \gamma$	Hipótesis
||	(2)	$\neg\alpha$	Hipótesis
||	(3)	$\neg\beta$	Hipótesis
||	(4)	$\neg\alpha \wedge \neg\beta$	IC 2, 3
||	(5)	$\neg(\alpha \vee \beta)$	DM 4
||	(6)	γ	MTP 1, 5
|	(7)	$\neg\beta \to \gamma$	
|	(8)	$\beta \vee \gamma$	EDI 7
	(9)	$\neg\alpha \to (\beta \vee \gamma)$	
|	(10)	$\alpha \vee (\beta \vee \gamma)$	EDI 9
	(11)	$(\alpha \vee \beta) \vee \gamma \to \alpha \vee (\beta \vee \gamma)$	

La otra implicación es análoga.

|	(1)	$\alpha \wedge (\beta \vee \gamma)$	Hipótesis
||	(2)	$\neg(\alpha \wedge \beta)$	Hipótesis
||	(3)	$\neg\alpha \vee \neg\beta$	DM 2
||	(4)	α	EC 1
||	(5)	$\neg\beta$	MTP 3, 4 (omitiendo DN)
||	(6)	$\beta \vee \gamma$	EC 1
||	(7)	γ	MTP 5, 6
||	(8)	$\alpha \wedge \gamma$	IC 4, 7
|	(9)	$\neg(\alpha \wedge \beta) \to (\alpha \wedge \gamma)$	
|	(10)	$(\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$	EDI 9
	(11)	$\alpha \wedge (\beta \vee \gamma) \to ((\alpha \wedge \beta) \vee (\alpha \wedge \gamma))$	
|	(12)	$(\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$	Hipótesis
||	(13)	$\neg\alpha$	Hipótesis
||	(14)	$\neg\alpha \vee \neg\beta$	ID 13
||	(15)	$\neg(\alpha \wedge \beta)$	DM 14
||	(16)	$\alpha \wedge \gamma$	MTP 12, 15
||	(17)	α	EC 16 (contradicción con 13)
|	(18)	α	
||	(19)	$\neg\beta$	Hipótesis
||	(20)	$\neg\alpha \vee \neg\beta$	ID 19
||	(21)	$\neg(\alpha \wedge \beta)$	DM 20
||	(22)	$\alpha \wedge \gamma$	MTP 12, 21
||	(23)	γ	EC 22
|	(24)	$\neg\beta \to \gamma$	
|	(25)	$\beta \vee \gamma$	EDI 24
|	(26)	$\alpha \wedge (\beta \vee \gamma)$	IC 18, 25
	(27)	$(\alpha \wedge \beta) \vee (\alpha \wedge \gamma) \to \alpha \wedge (\beta \vee \gamma)$	
	(28)	$\alpha \wedge (\beta \vee \gamma) \leftrightarrow (\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$	IB 11, 27

■

Quizá el lector se pregunte si no sería preferible demostrar las equivalencias del teorema anterior usando tablas de verdad. Las tablas de verdad prueban que las fórmulas son lógicamente válidas. Más adelante demostraremos que todas las fórmulas lógicamente válidas son teoremas lógicos, y eso justificará tal alternativa.

Fórmulas equivalentes Las equivalencias como las del teorema anterior pueden ser usadas para reemplazar cualquier subfórmula de una fórmula por otra equivalente, en virtud del teorema siguiente:

Teorema 2.11 *Las fórmulas siguientes son teoremas lógicos:*

1. $(\alpha \leftrightarrow \alpha') \leftrightarrow (\neg\alpha \leftrightarrow \neg\alpha')$,
2. $\big((\alpha \leftrightarrow \alpha') \wedge (\beta \leftrightarrow \beta')\big) \rightarrow \big((\alpha \rightarrow \beta) \leftrightarrow (\alpha' \rightarrow \beta')\big)$,
3. $\big((\alpha \leftrightarrow \alpha') \wedge (\beta \leftrightarrow \beta')\big) \rightarrow \big((\alpha \vee \beta) \leftrightarrow (\alpha' \vee \beta')\big)$,
4. $\big((\alpha \leftrightarrow \alpha') \wedge (\beta \leftrightarrow \beta')\big) \rightarrow \big((\alpha \wedge \beta) \leftrightarrow (\alpha' \wedge \beta')\big)$,
5. $\big((\alpha \leftrightarrow \alpha') \wedge (\beta \leftrightarrow \beta')\big) \rightarrow \big((\alpha \leftrightarrow \beta) \leftrightarrow (\alpha' \leftrightarrow \beta')\big)$,
6. $\bigwedge x(\alpha \leftrightarrow \beta) \rightarrow (\bigwedge x\ \alpha \leftrightarrow \bigwedge x\ \beta)$,
7. $\bigwedge x(\alpha \leftrightarrow \beta) \rightarrow (\bigvee^1 x\ \alpha \leftrightarrow \bigvee^1 x\ \beta)$,
8. $\bigwedge x(\alpha \leftrightarrow \beta) \rightarrow (\bigvee x\ \alpha \leftrightarrow \bigvee x\ \beta)$.

DEMOSTRACIÓN: Veamos, por ejemplo, 6):

(1)	$\bigwedge x(\alpha \leftrightarrow \beta)$	Hipótesis
(2)	$\alpha \leftrightarrow \beta$	EG 1
(3)	$\bigwedge x\alpha$	Hipótesis
(4)	α	EG 3
(5)	β	MP 2, 5 (omitiendo EB)
(6)	$\bigwedge x\beta$	IG 5 (generalización legal)
(7)	$\bigwedge x\alpha \rightarrow \bigwedge x\beta$	
(8)	$\bigwedge x\beta \rightarrow \bigwedge x\alpha$	Se prueba análogamente
(9)	$\bigwedge x\alpha \leftrightarrow \bigwedge x\beta$	IB 7, 8

■

Existencia con unicidad Hemos definido

$$\bigvee^1 x_1 \cdots x_n \alpha \equiv \bigvee y_1 \cdots y_n \bigwedge x_1 \cdots x_n (\alpha \leftrightarrow y_1 = x_1 \wedge \cdots \wedge y_n = x_n).$$

Sin embargo, hay una fórmula equivalente que se usa con más frecuencia porque permite tratar por separado la existencia y la unicidad:

Teorema 2.12 *Si las variables $y_1, \ldots, y_n$ no están en la fórmula $\alpha(x_1, \ldots, x_n)$ y son distintas de $x_1, \ldots, x_n$, entonces*

$$\vdash \bigvee^1 x_1 \cdots x_n \alpha(x_1, \cdots, x_n) \leftrightarrow \bigvee x_1 \cdots x_n\, \alpha(x_1, \ldots, x_n) \wedge$$

$$\bigwedge x_1 \cdots x_n y_1 \cdots y_n (\alpha(x_1, \ldots x_n) \wedge \alpha(y_1, \ldots, y_n) \rightarrow y_1 = x_1 \wedge \cdots \wedge y_n = x_n).$$

Nota Con la notación que estamos empleando habitualmente para las sustituciones la última parte del teorema se escribe

$$\bigwedge x_1 \cdots x_n y_1 \cdots y_n (\alpha \wedge S^{y_1 \cdots y_n}_{x_1 \cdots x_n} \alpha \to y_1 = x_1 \wedge \cdots \wedge y_n = x_n).$$

Hemos empleado la notación alternativa en el enunciado porque así es como este resultado suele aparecer en la práctica. Por simplicidad probamos la equivalencia en el caso de una única variable.

DEMOSTRACIÓN: Veamos primero una implicación (por abreviar aplicaremos varias reglas de inferencia simultáneamente).

(1)	$\bigvee x\alpha \wedge \bigwedge xy(\alpha \wedge S^y_x\alpha \to y = x)$	Hipótesis
(2)	$\mathtt{S}^y_x\alpha$	EC, EP 1
(3)	α	Hipótesis
(4)	$\alpha \wedge \mathtt{S}^y_x\alpha \to y = x$	EC, EG, 1
(5)	$y = x$	IC 3, 2; MP 4
(6)	$\alpha \to y = x$	
(7)	$y = x$	Hipótesis
(8)	α	ETI 2, 7
(9)	$y = x \to \alpha$	
(10)	$\alpha \leftrightarrow y = x$	IB 6, 9
(11)	$\bigwedge x(\alpha \leftrightarrow y = x)$	IG 10
(12)	$\bigvee y \bigwedge x(\alpha \leftrightarrow y = x)$	IP 11
(13)	$\overset{1}{\bigvee} x\alpha$	Teorema 2.9
(14)	$\bigvee x\alpha \wedge \bigwedge xy(\alpha \wedge S^y_x\alpha \to y = x) \to \overset{1}{\bigvee} x\alpha$	

La otra implicación es similar:

(1)	$\overset{1}{\bigvee} x\alpha$	Hipótesis
(2)	$\bigvee y \bigwedge x(\alpha \leftrightarrow y = x)$	Teorema 2.9
(3)	$\bigwedge x(\alpha \leftrightarrow z = x)$	EP 2
(4)	$\mathtt{S}^z_x\alpha \leftrightarrow z = z$	EG 3
(5)	$\mathtt{S}^z_x\alpha$	EB, I, MP, 4
(6)	$\bigvee x\, \alpha$	IP 5
(7)	$\alpha \wedge \mathtt{S}^y_x\alpha$	Hipótesis
(8)	$\alpha \to z = x$	EG, EB 3
(9)	$\mathtt{S}^y_x\alpha \to z = y$	EG, EB 3
(10)	$y = x$	EC 7, MP 8, MP 9, SI, TI
(11)	$\alpha \wedge \mathtt{S}^y_x\alpha \to y = x$	
(12)	$\bigwedge xy(\alpha \wedge \mathtt{S}^y_x\alpha \to y = x)$	IG 11
(13)	$\overset{1}{\bigvee} x\alpha \to \bigvee x\alpha \wedge \bigwedge xy(\alpha \wedge S^y_x\alpha \to y = x)$	

■

Ejercicio: Comprobar que todas las generalizaciones en la prueba anterior son correctas.

Forma prenexa Los lenguajes formales permiten definir una noción de "complejidad" de una afirmación que resulta útil en contextos muy variados. Es frecuente que a los estudiantes les cueste asimilar la noción de límite de una función en un punto más de lo que les cuesta comprender otros conceptos del mismo nivel. Uno de los factores que influyen en ello es que empieza más o menos así: *"Para todo* $\epsilon > 0$ *existe un* $\delta > 0$ *tal que para todo* $x \in \mathbb{R}$, ...". La dificultad no está en que haya tres cuantificadores, pues una definición que empiece con *"Para todo* ϵ*, para todo* δ *y para todo* x *se cumple ..."* resulta mucho más sencilla. La complejidad de la definición de límite se debe a que los tres cuantificadores se alternan: *"para todo... existe... para todo ..."*

Vamos a definir la complejidad de una fórmula en términos de la alternancia de sus cuantificadores. Para ello introducimos la noción de forma prenexa:

Definición 2.13 Se dice que una fórmula sin descriptores α de un lenguaje formal $\mathcal{L}$ está en *forma prenexa* si $\alpha \equiv \pi\alpha_0$, donde α_0 es una fórmula sin cuantificadores y π es una sucesión finita de cuantificadores universales $\bigwedge x$ y/o existenciales $\bigvee x$. A π se le llama *prefijo* de α. En tal caso, se dice que α es de tipo Σ_n (Π_n) si su prefijo consta de n bloques de cuantificadores alternados empezando por un cuantificador existencial (universal). Las fórmulas sin cuantificadores[2] se llaman fórmulas Δ_0.

Por ejemplo, una fórmula de tipo Σ_3 es

$$\bigvee xy \bigwedge uvw \bigvee z(x + u = y \wedge yvv = z).$$

Esta clasificación tiene interés porque, como veremos a continuación, toda fórmula sin descriptores es lógicamente equivalente a una fórmula en forma prenexa. La prueba se basa en el teorema siguiente, que dejamos como ejercicio:

Teorema 2.14 *Se cumple:*

$$\begin{aligned}
&\vdash (\alpha \to \textstyle\bigwedge x\beta) \leftrightarrow \bigwedge x(\alpha \to \beta) && \text{si } x \text{ no está libre en } \alpha,\\
&\vdash (\alpha \to \textstyle\bigvee x\beta) \leftrightarrow \bigvee x(\alpha \to \beta) && \text{si } x \text{ no está libre en } \alpha,\\
&\vdash (\textstyle\bigwedge x\alpha \to \beta) \leftrightarrow \bigvee x(\alpha \to \beta) && \text{si } x \text{ no está libre en } \beta,\\
&\vdash (\textstyle\bigvee x\alpha \to \beta) \leftrightarrow \bigwedge x(\alpha \to \beta) && \text{si } x \text{ no está libre en } \beta,\\
&\vdash (\alpha \vee \textstyle\bigwedge x\beta) \leftrightarrow \bigwedge x(\alpha \vee \beta) && \text{si } x \text{ no está libre en } \alpha,\\
&\vdash (\alpha \vee \textstyle\bigvee x\beta) \leftrightarrow \bigvee x(\alpha \vee \beta) && \text{si } x \text{ no está libre en } \alpha,\\
&\vdash (\textstyle\bigwedge x\alpha \wedge \beta) \leftrightarrow \bigwedge x(\alpha \wedge \beta) && \text{si } x \text{ no está libre en } \beta,\\
&\vdash (\textstyle\bigvee x\alpha \wedge \beta) \leftrightarrow \bigvee x(\alpha \wedge \beta) && \text{si } x \text{ no está libre en } \beta.
\end{aligned}$$

Ahora es fácil probar:

[2]En realidad, el concepto de fórmula Δ_0 depende del contexto, aunque se trata siempre de una restricción sobre los cuantificadores que pueden aparecer en la fórmula. Aquí hemos considerado el caso extremo de no admitir cuantificadores, pero, por ejemplo, en aritmética se admite que contengan cuantificadores de la forma $\bigwedge x \le y$, $\bigvee x \le y$, mientras que en teoría de conjuntos se admiten cuantificadores $\bigwedge x \in y$, $\bigvee x \in y$ en las fórmulas Δ_0.

Teorema 2.15 *Si α es una fórmula sin descriptores, existe otra fórmula β en forma prenexa con las mismas variables libres que α tal que $\vdash \alpha \leftrightarrow \beta$.*

DEMOSTRACIÓN: Lo probamos por inducción sobre la longitud de α.

Si $\alpha \equiv R_i^n t_1 \cdots t_n$, entonces ya está en forma prenexa (porque al no haber descriptores los términos no tienen cuantificadores). Tomamos $\beta \equiv \alpha$.

Si $\alpha \equiv \neg\gamma$, por hipótesis de inducción sabemos que $\vdash \gamma \leftrightarrow \pi\delta$, para cierta fórmula $\pi\delta$ en forma prenexa, luego por el teorema 2.11 tenemos $\vdash \neg\gamma \leftrightarrow \neg\pi\delta$. Aplicando (NG) y (NP) podemos "meter" el negador, y así $\vdash \neg\gamma \leftrightarrow \pi'\neg\delta$, donde π' es la sucesión de cuantificadores que resulta de cambiar cada cuantificador universal de π por uno existencial y viceversa.

Si $\alpha \equiv \gamma \to \delta$, por hipótesis de inducción $\vdash \gamma \leftrightarrow \pi\epsilon$ y $\vdash \delta \leftrightarrow \pi'\eta$. Aplicando el teorema 2.9 si es preciso, podemos suponer que las variables que liga π no están en $\pi'\eta$ y viceversa. Por el teorema 2.11 tenemos que $\vdash \alpha \leftrightarrow (\pi\epsilon \to \pi'\eta)$. Por el teorema anterior, $\vdash \alpha \leftrightarrow \pi'(\pi\epsilon \to \eta)$, y así mismo, $\vdash (\pi\epsilon \to \eta) \leftrightarrow \pi''(\epsilon \to \eta)$.

Por (IG) y 2.11 tenemos que $\vdash \pi'(\pi\epsilon \to \eta) \leftrightarrow \pi'\pi''(\epsilon \to \eta)$ y, por lo tanto, $\vdash \alpha \leftrightarrow \pi'\pi''(\epsilon \to \eta)$.

Si $\alpha \equiv \bigwedge x\gamma$, por hipótesis de inducción $\vdash \gamma \leftrightarrow \pi\delta$. Por (IG) tenemos que $\vdash \bigwedge x(\gamma \leftrightarrow \pi\delta)$ y por 2.11 queda $\vdash \alpha \leftrightarrow \bigwedge x\pi\delta$.

Es fácil comprobar que en cada caso las variables libres de la fórmula construida son las mismas que las de la fórmula dada. ∎

En la práctica es fácil extraer los cuantificadores de cualquier fórmula dada usando el teorema 2.14.

Sustitución de variables ligadas Conviene observar que si en una expresión cambiamos sus variables ligadas por otras "nuevas" obtenemos una expresión equivalente. Para formular este hecho con precisión consideramos la definición siguiente:

Dada una expresión θ, pongamos que las variables que aparecen ligadas en θ se encuentran entre $x_0, \ldots, x_k$, y llamamos $u_0, \ldots, u_k$ a otras variables (distintas de las x_i) que no aparezcan en θ ni libres ni ligadas. Definimos θ' como la expresión dada por las condiciones siguientes:

1. Si $\theta \equiv x_i$, entonces $\theta' \equiv x_i$.
2. Si $\theta \equiv c_i$, entonces $\theta' \equiv c_i$.
3. Si $\theta \equiv R_i^n t_1 \cdots t_n$, entonces $\theta' \equiv R_i^n t'_1 \cdots t'_n$.
4. Si $\theta \equiv f_i^n t_1 \cdots t_n$, entonces $\theta' \equiv f_i^n t'_1 \cdots t'_n$.
5. Si $\theta \equiv \neg\alpha$, entonces $\theta' \equiv \neg\alpha'$.
6. Si $\theta \equiv \alpha \to \beta$, entonces $\theta' \equiv \alpha' \to \beta'$.
7. Si $\theta \equiv \bigwedge x_i\alpha$, entonces $\theta' \equiv \bigwedge u_i \mathtt{S}_{x_i}^{u_i}\alpha'$.
8. Si $\theta \equiv x_i|\alpha$, entonces $\theta' \equiv u_i|\mathtt{S}_{x_i}^{u_i}\alpha'$.

Es claro que θ' es la expresión que resulta de sustituir cada variable ligada x_i por la variable nueva u_i. Ahora, para subexpresiones de una expresión prefijada que determina las variables x_i, u_i, razonamos como sigue:

1. *Las variables ligadas de cualquier sustitución* $\mathtt{S}_{x_{i_1}}^{u_{i_1}} \cdots \mathtt{S}_{x_{i_r}}^{u_{i_r}} \theta'$ *están entre* $u_0, \ldots, u_k$*, sin que la definición de sustitución obligue a introducir otras nuevas.*

   En efecto, razonando por inducción sobre la longitud de θ, esto es inmediato si θ es una variable o una constante y se sigue inmediatamente de la hipótesis de inducción salvo si $\theta \equiv \bigwedge x_i \alpha$ o $\theta \equiv x_i | \alpha$. Consideramos el primer caso, pues el segundo es análogo.

   Tenemos que $\theta' \equiv \bigwedge u_i \mathtt{S}_{x_i}^{u_i} \alpha'$. Si x_{i_r} está libre en θ', ha de ser necesariamente $x_{i_r} \not\equiv x_i$, luego $u_{i_r} \not\equiv u_i$, luego

   $$\mathtt{S}_{x_{i_r}}^{u_{i_r}} \theta' \equiv \bigwedge u_i \mathtt{S}_{x_{i_r}}^{u_{i_r}} \mathtt{S}_{x_i}^{u_i} \alpha',$$

   y esto es trivialmente cierto si x_{i_r} no está libre en θ'. Razonando del mismo modo llegamos a que

   $$\mathtt{S}_{x_{i_1}}^{u_{i_1}} \cdots \mathtt{S}_{x_{i_r}}^{u_{i_r}} \theta' \equiv \bigwedge u_i \mathtt{S}_{x_{i_1}}^{u_{i_1}} \cdots \mathtt{S}_{x_{i_r}}^{u_{i_r}} \mathtt{S}_{x_i}^{u_i} \alpha'$$

   y, por hipótesis de inducción, las variables ligadas de $\mathtt{S}_{x_{i_1}}^{u_{i_1}} \cdots \mathtt{S}_{x_{i_r}}^{u_{i_r}} \mathtt{S}_{x_i}^{u_i} \alpha'$ están entre $u_0, \ldots, u_k$, luego lo mismo vale para $\mathtt{S}_{x_{i_1}}^{u_{i_1}} \cdots \mathtt{S}_{x_{i_r}}^{u_{i_r}} \theta'$.

2. θ' *tiene las mismas variables libres que* θ *y sus variables ligadas están entre* $u_0, \ldots, u_k$.

   Se prueba trivialmente por inducción sobre la longitud de θ. Los únicos casos no triviales se dan cuando $\theta \equiv \bigwedge x_i \alpha$ o $\theta \equiv x_i | \alpha$, en cuyo caso basta tener en cuenta el apartado precedente.

Con esto ya podemos probar:

Teorema 2.16 *Si* θ *es una expresión cuyas variables ligadas estén entre las variables* $x_0, \ldots, x_k$ *y* $u_0, \ldots . u_k$ *son otras variables (distintas de las* x_i*) que no aparezcan en* θ *ni libres ni ligadas, llamando* θ' *a la expresión que resulta de reemplazar cada ocurrencia ligada de* x_i *en* θ *por otra variable* u_i*, si* θ *es un término,* $\vdash (\theta = \theta')$*, y si* θ *es una fórmula,* $\vdash (\theta \leftrightarrow \theta')$*.*

Demostración: Por inducción sobre la longitud de θ.

Si θ es una variable o una constante es inmediato, pues $\theta' \equiv \theta$.

Si $\theta \equiv R_i^n t_1 \cdots t_n$, por hipótesis de inducción $\vdash t_i = t_i'$, de donde se sigue fácilmente la equivalencia $\vdash (\theta \leftrightarrow \theta')$. El caso en que $\theta \equiv f_i^n t_1 \cdots t_n$ es similar.

Los casos en que $\theta \equiv \neg\alpha$ o $\theta \equiv \alpha \to \beta$ también son sencillos.

Si $\theta \equiv \bigwedge x_i \alpha$, por hipótesis de inducción $\vdash \alpha \leftrightarrow \alpha'$, de donde se sigue $\vdash \bigwedge x_i (\alpha \leftrightarrow \alpha')$, y a su vez $\vdash \mathtt{S}_{x_i}^{u_i} \alpha \leftrightarrow \mathtt{S}_{x_i}^{u_i} \alpha'$. Si suponemos θ, tenemos $\mathtt{S}_{x_i}^{u_i} \alpha$ (por EG), luego $\mathtt{S}_{x_i}^{u_i} \alpha'$, luego θ' (por IG). Esto nos da la implicación $\theta \to \theta'$, e igualmente se prueba la opuesta.

Si $\theta \equiv x_i|\alpha$, por hipótesis de inducción $\vdash \alpha \leftrightarrow \alpha'$, luego $\vdash \bigwedge x_i(\alpha \leftrightarrow \alpha')$. Por 2.11 tenemos que $\vdash \overset{1}{\bigvee} x_i\alpha \leftrightarrow \overset{1}{\bigvee} x_i\alpha'$. Por otra parte, se cumple también que $\vdash \overset{1}{\bigvee} x_i\alpha' \leftrightarrow \overset{1}{\bigvee} u_i \mathtt{S}^{u_i}_{x_i}\alpha'$.

En efecto, si suponemos $\overset{1}{\bigvee} x_i\alpha'$, esto significa $\bigvee y \bigwedge x_i(\alpha' \leftrightarrow y = x_i)$, de donde podemos pasar a $\bigwedge x_i(\alpha' \leftrightarrow y = x_i)$ por EP, de ahí a su vez, por EG, pasamos a $\mathtt{S}^{u_i}_{x_i}\alpha' \leftrightarrow y = u_i$, de donde $\bigwedge u_i(\mathtt{S}^{u_i}_{x_i}\alpha' \leftrightarrow y = u_i)$, por IG, y por último $\bigvee y \bigwedge u_i(\mathtt{S}^{u_i}_{x_i}\alpha' \leftrightarrow y = u_i)$, que es $\overset{1}{\bigvee} u_i \mathtt{S}^{u_i}_{x_i}\alpha'$ y esto nos da la implicación $\vdash \overset{1}{\bigvee} x_i\alpha' \to \overset{1}{\bigvee} u_i \mathtt{S}^{u_i}_{x_i}\alpha'$. La implicación opuesta se prueba igualmente.

En total tenemos que $\vdash \overset{1}{\bigvee} x_i\alpha \leftrightarrow \overset{1}{\bigvee} u_i \mathtt{S}^{u_i}_{x_i}\alpha'$.

Ahora, si suponemos $\overset{1}{\bigvee} x_i\alpha$, tenemos también $\overset{1}{\bigvee} u_i \mathtt{S}^{u_i}_{x_i}\alpha'$, de donde por DP, $\mathtt{S}^{x_i|\alpha}_{x_i}\alpha$ y $\mathtt{S}^{u_i|\mathtt{S}^{u_i}_{x_i}\alpha'}_{u_i}\mathtt{S}^{u_i}_{x_i}\alpha'$, pero de $\bigwedge x_i(\alpha \leftrightarrow \alpha')$ se sigue que

$$\mathtt{S}^{x_i|\alpha}_{x_i}\alpha \leftrightarrow \mathtt{S}^{x_i|\alpha}_{x_i}\alpha',$$

luego tenemos $\mathtt{S}^{x_i|\alpha}_{x_i}\alpha' \equiv \mathtt{S}^{x_i|\alpha}_{u_i}\mathtt{S}^{u_i}_{x_i}\alpha'$. Así pues,

$$\mathtt{S}^{x_i|\alpha}_{u_i}\mathtt{S}^{u_i}_{x_i}\alpha' \wedge \mathtt{S}^{u_i|\mathtt{S}^{u_i}_{x_i}\alpha'}_{u_i}\mathtt{S}^{u_i}_{x_i}\alpha',$$

luego la unicidad implica que $x_i|\alpha = u_i|\mathtt{S}^{u_i}_{x_i}\alpha'$, es decir, que $\theta = \theta'$.

Si, por el contrario, $\neg\overset{1}{\bigvee} x_i\alpha$, tenemos también $\neg\overset{1}{\bigvee} u_i \mathtt{S}^{u_i}_{x_i}\alpha'$, y DI implica entonces que $\theta = (x|x = x) = \theta'$ y en ambos casos tenemos la igualdad requerida. ∎

2.5 Consideraciones finales

En este momento ya tenemos una idea precisa de lo que supone formalizar el razonamiento matemático: A partir de los axiomas de Peano se puede probar, por ejemplo, (informalmente) que la suma de números naturales es conmutativa. Formalizar este razonamiento significa, no sólo expresar mediante fórmulas del lenguaje de la aritmética tanto los axiomas de Peano como la conmutatividad de la suma, sino obtener esta fórmula a partir de los axiomas formalizados mediante una aplicación sistemática de los axiomas y reglas de inferencia de $K_{\mathcal{L}}$ (o de las técnicas de deducción prácticas que hemos justificado). No es evidente a priori que esto pueda hacerse. Y, aunque el lector lo intente y tenga éxito, eso no garantiza que, en general, siempre que podamos convencernos racionalmente que si unos objetos cumplen los axiomas de Peano deben necesariamente cumplir tal otra propiedad (la conmutatividad de la suma o la que sea), encontraremos una combinación de reglas de inferencia en $K_{\mathcal{L}}$ que nos permita llegar de los axiomas a la propiedad en cuestión. Si suprimiéramos algún axioma en $K_{\mathcal{L}}$ sería fácil

encontrar fórmulas lógicamente válidas no deducibles de los axiomas restantes, pero, ¿cómo sabemos que no nos falta ningún axioma en $K_{\mathcal{L}}$? Podría faltar algún axioma que no fuera necesario para extraer consecuencias sencillas de unos axiomas dados, pero que fuera imprescindible para obtener otras consecuencias lógicas más sofisticadas que ni se nos ocurren ahora. Demostrar que no es así, probar que $K_{\mathcal{L}}$ captura fiel y plenamente el concepto de "razonamiento matemático", es el reto principal que tenemos planteado a medio plazo.

Capítulo III

Teorías axiomáticas

Hasta ahora nuestro interés se ha centrado en $K_{\mathcal{L}}$, porque estamos interesados en obtener una caracterización precisa del razonamiento matemático y en el capítulo siguiente confirmaremos que $K_{\mathcal{L}}$ nos proporciona lo que buscamos, pero en realidad no nos interesa el razonamiento lógico, sino lo que sucede cuando aplicamos el razonamiento lógico a determinados conjuntos de axiomas con contenido matemático, como los axiomas de Peano o los axiomas de la teoría de conjuntos. En este capítulo presentaremos algunas generalidades sobre las teorías axiomáticas que resultan de fijar un conjunto "interesante" de axiomas y discutiremos algunos casos particulares de teorías de interés. De paso tendremos ocasión de examinar cómo trabajan los matemáticos con estas teorías.

3.1 Consistencia y completitud

Definición 3.1 Una *teoría axiomática* (de primer orden) sobre un lenguaje formal $\mathcal{L}$ es un sistema deductivo formal T sobre $\mathcal{L}$ cuyos axiomas contengan a los de $K_{\mathcal{L}}$ y cuyas reglas de inferencia sean las de $K_{\mathcal{L}}$.

En estas condiciones, los axiomas de $K_{\mathcal{L}}$ se llaman *axiomas lógicos* de T, mientras que los axiomas de T que no sean axiomas de $K_{\mathcal{L}}$ se llaman *axiomas propios* de T. En la práctica, cuando hablemos de los axiomas de una teoría axiomática se sobrentenderá —salvo que se indique lo contrario— que nos referimos a sus axiomas propios.

Observemos que si Γ es el conjunto de los axiomas (propios) de una teoría T y α es cualquier fórmula de $\mathcal{L}$, entonces

$$\underset{T}{\vdash} \alpha \qquad \text{syss} \qquad \Gamma \vdash \alpha.$$

En efecto, una sucesión de fórmulas de $\mathcal{L}$ es una demostración en T si y sólo si es una deducción en $K_{\mathcal{L}}$ a partir de Γ. En un caso las fórmulas de Γ se consideran como axiomas y en otro como premisas.

Por ello, todos los resultados que conocemos sobre deducciones en $K_{\mathcal{L}}$ son válidos inmediatamente para cualquier teoría axiomática.

Uno de nuestros objetivos a largo plazo será encontrar una teoría axiomática cuyos teoremas sean precisamente los teoremas que aceptan como tales los matemáticos. Veremos que no hay una sola. Dichas teorías se conocen como *teorías de conjuntos,* porque pretenden formalizar la noción informal (e imprecisa) de conjunto.[1]

Un *modelo* de una teoría axiomática T es un modelo M de sus axiomas, es decir, un modelo de su lenguaje formal en el que son verdaderos todos sus axiomas. Lo representaremos por $M \models T$.

Del teorema de corrección 2.6 se sigue que si $M \models T$, entonces todos los teoremas de T son verdaderos en M (aunque puede haber fórmulas verdaderas en M que no sean teoremas de T). Recíprocamente, ninguna fórmula falsa en M puede ser un teorema de T.

Observemos que el concepto de teoría axiomática es puramente formal, es decir, que para definir una teoría axiomática y trabajar con total rigor en ella basta definir un lenguaje formal y seleccionar un conjunto de axiomas entre sus fórmulas, sin que exista ninguna obligación de explicar qué pretenden significar los signos de la teoría, es decir, sin necesidad de dar un modelo de su lenguaje y mucho menos de sus axiomas.

Ejemplo La *Aritmética de Peano* (de primer orden) es la teoría axiomática AP sobre el lenguaje $\mathcal{L}_a$ de la aritmética cuyos axiomas son los *axiomas de Peano:*

$$\begin{array}{ll}
\text{(AP1)} & \bigwedge x\ x' \neq 0 \\
\text{(AP2)} & \bigwedge xy(x' = y' \to x = y) \\
\text{(AP3)} & \bigwedge x\ x + 0 = x \\
\text{(AP4)} & \bigwedge xy(x + y' = (x + y)') \\
\text{(AP5)} & \bigwedge x\ x \cdot 0 = 0 \\
\text{(AP6)} & \bigwedge xy(xy' = xy + x) \\
\text{(AP7)} & \phi(0) \wedge \bigwedge x(\phi(x) \to \phi(x')) \to \bigwedge x\phi(x),
\end{array}$$

donde ϕ es cualquier fórmula, tal vez con más variables libres aparte de x.

Observemos que el modelo natural del lenguaje $\mathcal{L}_a$, es decir, el modelo cuyo universo es el conjunto $\mathbb{N}$ de los números naturales y en el que los signos de $\mathcal{L}_a$ se interpretan de forma obvia, cumple $\mathbb{N} \models \mathrm{AP}$, por lo que ya no nos referiremos más a $\mathbb{N}$ como el modelo natural *del lenguaje* de la aritmética, sino como el *modelo natural de la aritmética* (de Peano).

[1] Notemos que el hecho de que $K_{\mathcal{L}}$ capture plenamente la capacidad de razonamiento lógico no significa que capture la totalidad del razonamiento matemático. Por ejemplo, si $\mathcal{L}_a$ es el lenguaje de la aritmética, en $K_{\mathcal{L}_a}$ no puede demostrarse que $0'' + 0'' = 0''''$. Esto no es un teorema lógico, pues es fácil dar un modelo de $\mathcal{L}_a$ (interpretando la suma como cualquier operación que se nos antoje) en el que la sentencia sea falsa. Dicha sentencia puede probarse, por ejemplo, a partir de los axiomas de Peano, pero éstos no son suficientes ni de lejos para que a partir de ellos se pueda demostrar cualquier teorema matemático. Para ello hacen falta teorías mucho más potentes.

En efecto, la prueba consiste en constatar meramente que si tomamos cualquiera de los axiomas y aplicamos la definición de $\mathbb{N} \vDash \alpha$, obtenemos una afirmación (informal) sobre los números naturales, que no es sino la versión informal del correspondiente axioma de Peano y que, por lo tanto, es verdadera. Por ejemplo, si consideramos una instancia del principio de inducción, vemos que

$$\mathbb{N} \vDash \phi(0) \wedge \bigwedge x(\phi(x) \to \phi(x')) \to \bigwedge x\phi(x).$$

equivale a que (fijada una valoración v), si

$$\mathbb{N} \vDash \phi(0)[v] \qquad \text{y} \qquad \mathbb{N} \vDash \big(\bigwedge x(\phi(x) \to \phi(x'))\big)[v],$$

entonces $\mathbb{N} \vDash \bigwedge x\phi(x)[v]$.

La primera parte de la hipótesis es que $\mathbb{N} \vDash \mathtt{S}^0_x\phi[v]$, que por 1.12 equivale a que $\mathbb{N} \vDash \phi[v^0_x]$, donde el 0 que aparece junto a v no es la constante 0, sino el número natural 0.

La segunda parte de la hipótesis equivale a que, para cada número a, se cumple $\mathbb{N} \vDash (\phi \to \mathtt{S}^{x'}_x\phi)[v^a_x]$, es decir, a que si $\mathbb{N} \vDash \phi[v^a_x]$ entonces también $\mathbb{N} \vDash \phi[v^{a+1}_x]$ (donde hemos usado de nuevo 1.12).

Por su parte la conclusión a la que debemos llegar equivale a que todo número a cumple $\mathbb{N} \vDash \phi[v^a_x]$.

En definitiva, que la fórmula que estamos considerando sea verdadera equivale a que sea verdadero el principio (informal) de inducción para la propiedad Pa que se cumple si y sólo si $\mathbb{N} \vDash \phi[v^a_x]$. Notemos que, sea cual sea ϕ, se trata en última instancia de una propiedad del número natural a expresable exclusivamente en términos del cero, la operación sucesor, la suma y el producto.[2]

∎

El riesgo principal que corremos cuando definimos una teoría axiomática es que ésta resulte ser contradictoria:

Definición 3.2 Un conjunto de fórmulas Γ de un lenguaje formal $\mathcal{L}$ es *contradictorio* si existe una fórmula α de $\mathcal{L}$ tal que $\Gamma \vdash \alpha$ y $\Gamma \vdash \neg\alpha$. En caso contrario se dice que es *consistente.* Equivalentemente, una teoría axiomática T es *contradictoria* si existe una fórmula α tal que $\underset{T}{\vdash} \alpha$ y $\underset{T}{\vdash} \neg\alpha$. En caso contrario es *consistente.*

La equivalencia consiste en que, obviamente, una teoría axiomática es consistente o contradictoria si y sólo si lo son sus axiomas: Es equivalente decir

[2]En definitiva, la prueba se reduce en última instancia a que si se cumple $P0$ y cuando se cumple Pn también se cumple $P(n+1)$, entonces todos los números naturales tienen que cumplir Pn, y es inconcebible que esto no sea cierto. Si tomamos cualquier número natural, por ejemplo el 1000, podremos razonar en 1000 pasos que, o bien se cumple $P1000$, o bien es falso uno de los supuestos que estamos haciendo. Esto no es demostrable racionalmente a partir de principios más elementales, sino que se cumple porque sabemos que la afirmación "todo número natural cumple P" significa precisamente que se cumple $P0$ y $P1$ y $P2$ y que esta sucesión de afirmaciones se puede prolongar sin fin.

"la aritmética de Peano es consistente" que decir "los axiomas de la aritmética de Peano son consistentes". En general, todos los resultados sobre consistencia pueden enunciarse equivalentemente en términos de conjuntos de fórmulas o de teorías axiomáticas. Nosotros usaremos arbitraria e indistintamente una u otra formulación. El teorema siguiente muestra la importancia en la lógica matemática de la noción de consistencia.

Teorema 3.3 *Una teoría axiomática T sobre un lenguaje $\mathcal{L}$ es contradictoria si y sólo si todas las fórmulas de $\mathcal{L}$ son teoremas de T.*

DEMOSTRACIÓN: Si en T puede probarse una contradicción, la regla de inferencia (C) de contradicción nos permite prolongar la prueba hasta una demostración de cualquier fórmula. El recíproco es todavía más evidente. ∎

Así pues, la consistencia es el requisito mínimo que ha de tener una teoría axiomática para que tenga interés trabajar en ella.

Nota A menudo conviene tener presente la versión recíproca del teorema anterior: *Una teoría axiomática es consistente si y sólo si existe una fórmula que no es un teorema.* Por ejemplo, sabemos que $K_{\mathcal{L}}$ es consistente, pues sus teoremas son necesariamente fórmulas lógicamente válidas, y hay fórmulas que no son lógicamente válidas, luego no son teoremas lógicos. Este argumento se puede generalizar:

Teorema 3.4 *Si una teoría axiomática tiene un modelo, entonces es consistente.*

DEMOSTRACIÓN: Si una teoría T tiene un modelo M, entonces todos los teoremas de T son verdaderos en M, y como en todo modelo hay afirmaciones falsas (las negaciones de las verdaderas), concluimos que no todo es demostrable en T. ∎

Aunque no es trivial en absoluto, en el capítulo siguiente veremos que el recíproco es cierto: si una teoría es consistente, entonces tiene un modelo, y ésta es la interpretación semántica de la consistencia (que es un concepto puramente sintáctico): una teoría axiomática es consistente si y sólo si habla de algo, si y sólo si existen objetos que (con las relaciones y funciones oportunas) cumplen lo que requieren sus axiomas.

En particular podemos afirmar que la aritmética de Peano es consistente, ya que tiene su modelo natural.[3]

Cuantos más axiomas le ponemos a una teoría, más riesgo hay de que sea contradictoria:

[3]El lector queda advertido de que hay quienes cuestionan este resultado, pues (al contrario de lo que sucede con la consistencia de $K_{\mathcal{L}}$, que puede probarse considerando un modelo con un solo elemento) no puede considerarse una prueba finitista en sentido estricto, ya que involucra de forma esencial al conjunto (infinito) de los números naturales como un todo, y no es reducible a términos que involucren únicamente procesos finitos.

Teorema 3.5 *Sean Δ y Γ conjuntos de fórmulas de un lenguaje formal $\mathcal{L}$ tales que toda fórmula de Δ sea consecuencia de Γ (esto sucede en particular si Δ está contenido en Γ). Entonces, si Γ es consistente, Δ también lo es y si Δ es contradictoria, Γ también lo es.*

DEMOSTRACIÓN: Una contradicción deducida de las premisas de Δ puede probarse a partir de Γ deduciendo primero todas las premisas empleadas (lo cual es posible por hipótesis) y después continuando con el mismo razonamiento. ∎

Observemos que la consistencia es una propiedad negativa: una teoría es consistente si ciertas cosas no se pueden demostrar en ella. Esto hace que en general no sea fácil determinar si una teoría dada es consistente o no. De hecho, veremos que en los casos más importantes es imposible. Por ello tienen interés los resultados de consistencia relativa, es decir, pruebas de que si una teoría es consistente sigue siéndolo al añadirle algún axioma más. En esta línea es útil tener presente la siguiente equivalencia:

Teorema 3.6 *Sea Γ un conjunto de fórmulas y α una sentencia. Entonces $\Gamma \cup \{\alpha\}$ es consistente*[4] *si y sólo si no $\Gamma \vdash \neg\alpha$.*

DEMOSTRACIÓN: Si $\Gamma \vdash \neg\alpha$ entonces $\Gamma \cup \{\alpha\} \vdash \alpha$ y $\Gamma \cup \{\alpha\} \vdash \neg\alpha$, luego $\Gamma \cup \{\alpha\}$ es contradictorio.

Si no $\Gamma \vdash \neg\alpha$, para ver que $\Gamma \cup \{\alpha\}$ es consistente basta probar que no $\Gamma \cup \{\alpha\} \vdash \neg\alpha$, pero en caso contrario, por el teorema de deducción (y aquí usamos que α es una sentencia) tendríamos $\Gamma \vdash \alpha \to \neg\alpha$, es decir, $\Gamma \vdash \neg\alpha \vee \neg\alpha$. Por consiguiente $\Gamma \vdash \neg\alpha$, en contra de lo supuesto. ∎

Por ejemplo, es equivalente decir que el quinto postulado de Euclides es independiente de los demás axiomas de la geometría (es decir, que no se puede demostrar a partir de ellos) que decir que la negación del quinto postulado es consistente con los demás axiomas de la geometría.

Ejercicio: Mostrar que el teorema anterior es falso si α no es una sentencia. (Considerar, por ejemplo, $\Gamma = \{\bigvee xy\, \neg x = y\}$, $\alpha \equiv x = y$.)

Clausuras universales Nada impide que los axiomas de una teoría formal tengan variables libres, pero a veces conviene exigir que sean sentencias. Ello no supone ninguna pérdida de generalidad por lo siguiente:

Dada una fórmula α, su *clausura universal* es la sentencia α^c que resulta de ligar todas sus variables libres con cuantificadores universales. Aplicando repetidamente IG y EG se ve que

$$\alpha \vdash \alpha^c \qquad \text{y} \qquad \alpha^c \vdash \alpha.$$

[4]Evidentemente, $\Gamma \cup \{\alpha\}$ representa el conjunto que resulta de añadir la sentencia α a Γ. La notación conjuntista es mera taquigrafía. El enunciado no corresponde a un teorema de la teoría de conjuntos, sino que es un metateorema no formalizado, como todos los resultados que estamos probando.

Si llamamos Γ^c al conjunto de las clausuras universales de las fórmulas de Γ, resulta que toda fórmula de Γ se deduce de Γ^c y viceversa, por lo que ambas tienen las mismas consecuencias lógicas, y además el teorema 3.5 nos da que Γ es consistente si y sólo si lo es Γ^c. Más aún, si M es un modelo del lenguaje formal correspondiente, se cumple $M \models \Gamma$ si y sólo si $M \models \Gamma^c$.

Ejercicio: Probar que en general no es cierto que $\vdash \alpha \leftrightarrow \alpha^c$, por ejemplo si $\alpha \equiv x = y$.

Definición 3.7 Una teoría axiomática T es *completa* si toda sentencia[5] α de su lenguaje formal cumple $\underset{T}{\vdash} \alpha$ o $\underset{T}{\vdash} \neg\alpha$.

Hay un caso en el que una teoría dada es indudablemente completa: si es contradictoria, pero obviamente este caso carece de interés. Si la consistencia es lo mínimo que ha de cumplir una teoría axiomática para que tenga interés, la completitud es lo máximo que le podemos pedir. Una teoría completa es una teoría capaz de resolvernos cualquier duda sobre su objeto de estudio.

Sin embargo, veremos que ninguna teoría matemática razonable y suficientemente potente (lo justo para poder demostrar que los números naturales cumplen los axiomas de Peano) puede ser completa.

3.2 La teoría básica de conjuntos

Presentamos aquí una versión simplificada de la teoría axiomática que usan habitualmente los matemáticos, es decir, una teoría de conjuntos.

3.2.1 Axiomas y primeras consecuencias

Definición 3.8 El *lenguaje de la teoría de conjuntos* $\mathcal{L}_{\mathrm{tc}}$ es el lenguaje formal (con descriptor) cuyo único signo eventual es un relator diádico que representaremos por $\in$ y lo llamaremos *relator de pertenencia.* Escribiremos $t_1 \notin t_2 \equiv \neg t_1 \in t_2$. También es frecuente abreviar

$$\bigwedge u \in x\, \alpha \equiv \bigwedge u(u \in x \to \alpha), \qquad \bigvee u \in x\, \alpha \equiv \bigvee u(u \in x \wedge \alpha).$$

La *teoría básica de conjuntos* es la teoría axiomática B cuyos axiomas son las sentencias siguientes:

Extensionalidad	$\bigwedge xy(\bigwedge u(u \in x \leftrightarrow u \in y) \to x = y)$
Par	$\bigwedge xy\bigvee z\bigwedge u(u \in z \leftrightarrow u = x \vee u = y)$
Unión	$\bigwedge x\bigvee y\bigwedge u(u \in y \leftrightarrow \bigvee v(u \in v \wedge v \in x))$
Diferencia	$\bigwedge xy\bigvee z\bigwedge u(u \in z \leftrightarrow u \in x \wedge u \notin y)$

[5] La restricción de la definición a sentencias es fundamental. Si definiéramos la completitud con fórmulas cualesquiera, como la fórmula $x \neq y$ es contradictoria, toda teoría consistente y completa debería cumplir $\underset{T}{\vdash}(x = y)$ y admitiría a lo sumo modelos con un elemento.

Vamos a analizar esta definición.[6] Ante todo observemos que todo lo dicho tiene sentido. La definición anterior determina exactamente qué es una fórmula del lenguaje $\mathcal{L}_{\mathrm{tc}}$, qué es lo que significa "ser un axioma" de B y, por consiguiente, está perfectamente determinado qué significa "ser un teorema" de B. En breve nos pondremos a demostrar formalmente teoremas de B y todo esto podemos hacerlo sin responder en ningún momento a la pregunta de "qué significa $\in$".

Más aún, vamos a establecer, no sin cierta dosis de cinismo, que cuando leamos sentencias de $\mathcal{L}_{\mathrm{tc}}$, cada vez que nos encontremos con un $\bigwedge x$ leeremos "para todo conjunto x", cada vez que nos encontremos con un $\bigvee x$ leeremos "existe un conjunto x tal que", y cuando nos encontremos con una fórmula $x \in y$ leeremos que "el conjunto x pertenece a (o es un elemento de) el conjunto y", y si alguien nos pregunta qué queremos decir cuando hablamos de "conjuntos" y de "pertenencia", le responderemos que nada, que es sólo una forma cómoda de leer las fórmulas de $\mathcal{L}_{\mathrm{tc}}$. Y lo bueno es que nadie podrá acusarnos de falta de rigor.

Más concretamente, podemos leer el axioma de extensionalidad diciendo que "si dos conjuntos x e y tienen los mismos elementos, entonces son iguales", y podemos trabajar con esta afirmación, y deducir consecuencias lógicas de ella y los demás axiomas, sin necesidad de dar explicaciones sobre qué son esos conjuntos y esa relación de pertenencia de la que se supone que hablamos.

Si nos cansamos de ser cínicos y preferimos ser algo más conciliadores, podemos decir que suponemos que existen unos objetos llamados conjuntos, entre los cuales está definida una relación de pertenencia, y de forma que, cuando consideramos el modelo cuyo universo es la colección[7] de todos esos conjuntos e interpretamos el relator $\in$ como la relación de pertenencia, todos los axiomas de B resultan ser verdaderos.

Pero si nos preguntan si tenemos garantías de que existen realmente tales objetos y tal relación, o si podemos poner algún ejemplo de objetos que realmente cumplan esos axiomas, podemos responder sin vacilar que no necesitamos responder a ninguna de esas preguntas para trabajar rigurosamente con la teoría B, porque *para razonar formalmente no es necesario conocer los objetos sobre los que presuntamente estamos razonando,* sino que basta con respetar las reglas sintácticas de $\mathcal{L}_{\mathrm{tc}}$ y las reglas deductivas de $K_{\mathcal{L}_{\mathrm{tc}}}$.

En resumen: todo el aparato lógico que hemos montado hasta aquí nos sirve ahora para que podamos hablar de conjuntos y de pertenencia sin necesidad de responder a ninguna pregunta embarazosa sobre qué son los conjuntos y qué es la pertenencia. Nos basta con *suponer* que los conjuntos y la pertenencia

[6] El lector no debe preocuparse de momento por el contenido concreto de los axiomas que acabamos de dar. Lo discutiremos unas páginas más abajo, tras algunas reflexiones generales sobre la teoría en general.

[7] Hasta ahora hemos venido empleando la palabra "conjunto" para referirnos a colecciones de objetos bien definidas. Sin embargo, en este contexto no podemos seguir empleando esa palabra sin caer en equívocos, por lo que aquí usamos la palabra "colección" para lo que siempre hemos llamado "conjunto". Sucede que ahora "conjunto" es un término técnico, el nombre que damos a los objetos de un modelo de la teoría B, y así, la colección de todos esos objetos (que es un conjunto en el sentido general que hasta ahora dábamos a la palabra), ya no es un conjunto en el sentido técnico, pues no es (o, al menos, no tiene por qué ser) ninguno de los objetos de dicho universo.

cumplen los axiomas de B y respetar en todo momento las reglas de deducción formal que hemos establecido.

Conceptos primitivos En la práctica, cuando uno ya está habituado a estas situaciones y no necesita dar tantas vueltas a estos hechos, abrevia diciendo simplemente que los conceptos de "conjunto" y "pertenencia" son los *conceptos primitivos* (o no definidos) de la teoría B.

La lógica formal tendrá mil virtudes, pero si hoy tiene el grado de desarrollo que tiene, es porque sirve precisamente para esto, para justificar que uno se ponga a hablar de cosas sin decir de qué cosas está hablando sin que nadie le pueda reprochar falta de rigor. La fundamentación de las matemáticas consiste esencialmente en eso: en poder hablar de conjuntos y de pertenencia sin verse en la necesidad (o en el aprieto) de explicar qué significan exactamente estas palabras.

Colecciones y conjuntos Una vez eximidos de la necesidad de explicar qué es un conjunto o qué es la relación de pertenencia, nada nos impide tratar de formarnos una idea sobre qué características tendrían que tener unos objetos para cumplir los axiomas de B. Si x es un conjunto, entonces podemos hablar de la colección (en el sentido no técnico de la palabra) de todos los conjuntos que cumplen $u \in x$. En principio, esa colección recibe el nombre de la *extensión* de x, y lo que dice el axioma de extensionalidad es que si dos conjuntos tienen la misma extensión, entonces son iguales.

De aquí obtenemos dos cosas: por una parte, cada conjunto (en el sentido técnico, no definido, de la palabra) tiene asociada una colección de objetos (su extensión) y en segundo lugar está completamente determinado por ella, de modo que lo único que diferencia a un conjunto de otro es que haya otros conjuntos que pertenezcan a uno y no al otro. Por lo tanto, podemos identificar a un conjunto con su extensión. Podemos considerar que cuando hablamos de un conjunto, en realidad estamos hablando de su extensión. Según esto, los conjuntos *son* colecciones de objetos (colecciones de conjuntos). Eso sí, sería ingenuo y catastrófico identificar "conjunto" y "colección de conjuntos". Si resulta que B tiene un modelo M y considero unos cuantos objetos de M, nadie me asegura que exista un conjunto (es decir, un elemento de M) cuya extensión sea precisamente la colección de objetos de M que he tomado. Todo conjunto (elemento de M) puede identificarse con una colección de conjuntos (de elementos de M), pero no toda colección de conjuntos (elementos de M) es necesariamente identificable con un conjunto (elemento de M). De hecho, luego veremos que siempre hay colecciones de conjuntos que no son (extensiones de) conjuntos.

Definiciones formales Aunque la teoría tiene sólo dos conceptos primitivos (el de conjunto y el de pertenencia) podemos introducir otros nuevos. Por ejemplo, podemos decir:

Definición $x \subset y \equiv \bigwedge u(u \in x \to u \in y)$.

y establecer que este nuevo signo se lee "x es un subconjunto de y".

Analicemos esto. Se trata de una definición formal en una teoría matemática. No es la primera que nos encontramos, pero tal vez sí la primera que nos encontramos en el contexto en que los matemáticos definen cosas cotidianamente. Nosotros estamos adoptando el convenio de que las definiciones de este tipo son meras abreviaturas, es decir, los dos miembros de la definición son dos nombres distintos para *la misma* fórmula. Vistas así, las definiciones no son "oficialmente" nada. No hemos añadido nada a la teoría, si en un razonamiento pasamos de $x \subset y$ a $\bigwedge u(u \in x \to u \in y)$ simplemente estamos aplicando la regla R de repetición.

Hay otra forma alternativa de concebir las definiciones, que consiste en considerar que $\subset$ es un nuevo relator diádico que añadimos al lenguaje $\mathcal{L}_{\mathrm{tc}}$ y que la definición es en realidad un nuevo axioma:

$$x \subset y \leftrightarrow \bigwedge u(u \in x \to u \in y),$$

que a partir de este momento podemos utilizar en las demostraciones que hagamos en B. Este planteamiento es más complicado y no lo vamos a adoptar. Lo citamos meramente para dejar claro que *no* es esto lo que hacemos. "Oficialmente" en $\mathcal{L}_{\mathrm{tc}}$ no hay ningún relator diádico más que $=$ y $\in$, cualquier fórmula en la que aparezca $\subset$ es simplemente una fórmula que hemos decidido nombrar con un cierto convenio, pero podríamos nombrarla también sin considerar el signo $\subset$.

Ahora podemos demostrar tres teoremas de B:

$$\underset{B}{\vdash} \bigwedge x\ x \subset x, \quad \underset{B}{\vdash} \bigwedge xy(x \subset y \wedge y \subset x \to x = y),$$

$$\underset{B}{\vdash} \bigwedge xyz(x \subset y \wedge y \subset z \to x \subset z).$$

Veamos, por ejemplo, el segundo:

(1)	$x \subset y \wedge y \subset x$	Hipótesis
(2)	$x \subset y$	EC 1
(3)	$y \subset x$	EC 1
(4)	$\bigwedge u(u \in x \to u \in y)$	R 2
(5)	$\bigwedge u(u \in y \to u \in x)$	R 3
(6)	$u \in x \to u \in y$	EG 4
(7)	$u \in y \to u \in x$	EG 5
(8)	$u \in x \leftrightarrow u \in y$	IB 6, 7
(9)	$\bigwedge u(u \in x \leftrightarrow u \in y)$	IG 8
(10)	$\bigwedge xy(\bigwedge u(u \in x \leftrightarrow u \in y) \to x = y)$	Extensionalidad
(11)	$\bigwedge u(u \in x \leftrightarrow u \in y) \to x = y$	EG 10
(12)	$x = y$	MP 9, 11
(13)	$x \subset y \wedge y \subset x \to x = y$	
(14)	$\bigwedge xy(x \subset y \wedge y \subset x \to x = y)$	IG 13

Esto sería una demostración en B según lo visto en los capítulos precedentes. Sin embargo, un matemático nunca detalla tanto los argumentos, sino que se

limita a dar las ideas que considera suficientes para que cualquiera pueda desarrollarlas si quiere hasta este nivel. En este caso (si no se limita a decir que es evidente), el matemático diría algo así como

> Como $x \subset y$, tenemos que todo $u \in x$ cumple también $u \in y$, y como $y \subset x$, tenemos también la implicación opuesta, luego tenemos que $u \in x \leftrightarrow u \in y$. El axioma de extensionalidad nos da que $x = y$.

La cuestión es que esto es un argumento riguroso porque es formalizable, porque si alguien pusiera alguna objeción siempre podría detallarse más y más hasta llegar si fuera preciso a la prueba completa anterior, donde se puede comprobar sin lugar a dudas que cada paso sigue las reglas establecidas para la deducción en B (la generalización en 9 es legítima, etc.)

En lo sucesivo (salvo en contextos muy específicos) nunca volveremos a razonar numerando líneas y citando reglas de inferencia, sino que esbozaremos las demostraciones siguiendo el uso habitual en matemáticas, destacando los aspectos lógicos subyacentes sólo cuando sean relevantes por algún motivo. Pero el lector debe tener claro que lo que distingue un razonamiento válido de una falacia es precisamente que el primero puede detallarse todo lo necesario hasta explicitar qué regla de inferencia se aplica a cada paso y la segunda no.

Definiciones de términos Analicemos ahora el segundo axioma de B. Afirma que, dados dos conjuntos, existe un tercer conjunto que los tiene por elementos. Combinando esto con el axioma de extensionalidad podemos concluir (razonando en B, por supuesto) que:

$$\bigwedge xy \overset{1}{\bigvee} z \bigwedge u(u \in z \leftrightarrow u = x \vee u = y)$$

En efecto, la existencia nos la da el axioma del par, y la unicidad se debe a que si z_1 y z_2 son dos conjuntos cuyos elementos sean x e y, entonces ambos tienen los mismos elementos, luego $z_1 = z_2$ por el axioma de extensionalidad.

El matemático dice entonces "llamaremos $\{x, y\}$ al conjunto cuyos elementos son x e y", pero ¿cómo debe entenderse esto? No es trivial. Cuando un matemático define algo como $x \subset y$, que aparentemente es un relator, siempre podemos decir que no hay relator alguno, sino que se trata de una mera abreviatura, pero ¿qué clase de abreviatura es $\{x, y\}$? ¿a qué término se supone que abrevia?

Hay básicamente tres formas de entender una definición como ésta:

1. Considerar que la definición del conjunto par de dos conjuntos supone añadir un nuevo funtor diádico al lenguaje $\mathcal{L}_{\mathrm{tc}}$, junto con el axioma

$$\bigwedge xyu\ (u \in \{x, y\} \leftrightarrow u = x \vee u = y).$$

2. Considerar que $\{x, y\}$ no significa nada en realidad, pero que cada fórmula que contenga a este signo puede interpretarse como el nombre de otra fórmula equivalente que no lo contiene.

Por ejemplo, $\{x, y\} \in w$ puede verse como una forma de referirnos a la fórmula

$$\bigvee z \in w \bigwedge u\ (u \in z \leftrightarrow u = x \vee u = y).$$

3. Podemos usar el descriptor para definir (como abreviatura)

$$\{x, y\} \equiv z|\bigwedge u\ (u \in z \leftrightarrow u = x \vee u = y).$$

La posibilidad 1) supone que el lenguaje $\mathcal{L}_{\mathrm{tc}}$ va cambiando a medida que vamos definiendo cosas, y obliga a demostrar metateoremas que vengan a decir (y justificar) que a efectos teóricos es como si no añadiéramos nada. La posibilidad 2) exige especificar qué condiciones se han de cumplir para que sea legítimo añadir un signo y considerar que en el fondo es como si no estuviera. Todo ello puede hacerse y, en el fondo, se adopte la opción que se adopte, hay que hacerlo de un modo u otro, pero a nuestro juicio la opción 3) permite hacerlo de la forma más clara y elegante posible (con el coste que supone que el descriptor complica ligeramente la sintaxis de los lenguajes formales, pero consideramos que dicho coste es asumible a cambio de la claridad que proporciona en la gestión de las definiciones).

Veamos con más detalle lo que supone la opción 3). Hemos probado que $\bigwedge xy \overset{1}{\bigvee} z \bigwedge u\ (u \in z \leftrightarrow u = x \vee u = y)$, y tenemos la regla de las descripciones propias, que dice, en general, que

$$\overset{1}{\bigvee} z\alpha \vdash \mathsf{s}_z^{z|\alpha}\alpha.$$

En nuestro caso concreto, con $\alpha \equiv \bigwedge u\ (u \in z \leftrightarrow u = x \vee u = y)$, teniendo en cuenta que hemos definido $z|\alpha \equiv \{x, y\}$, lo que dice la regla es que

$$\overset{1}{\bigvee} z \bigwedge u\,(u \in z \leftrightarrow u = x \vee u = y) \vdash \bigwedge u\ (u \in \{x, y\} \leftrightarrow u = x \vee u = y).$$

Es decir: si existe un único conjunto con la propiedad de que sus elementos son x e y, y llamamos $\{x, y\}$ al z tal que z tiene por elementos a x e y, entonces la regla DP nos dice que $\{x, y\}$ tiene la propiedad que lo define, a saber, que $\{x, y\}$ es *el* conjunto de elementos x e y, tal y como queríamos.

Como nada exige que los conjuntos x e y sean distintos, podemos definir $\{x\} \equiv \{x, x\}$ y entonces es claro que

$$\bigwedge xu(u \in \{x\} \leftrightarrow u = x).$$

Podemos repetir el argumento con los otros dos axiomas. El axioma de la unión combinado con el de extensionalidad nos permite probar que

$$\bigwedge x \overset{1}{\bigvee} y \bigwedge u(u \in y \leftrightarrow \bigvee v(u \in v \wedge v \in x)).$$

(Nuevamente, dos conjuntos que cumplieran lo que se afirma de y tendrían los mismos elementos, a saber, los elementos de los elementos de x.) Por eso, si definimos

$$\bigcup x \equiv y|\bigwedge u(u \in y \leftrightarrow \bigvee v(u \in v \wedge v \in x)),$$

la regla DP nos permite concluir que

$$\bigwedge xu(u \in \bigcup x \leftrightarrow \bigvee v(u \in v \wedge v \in x)).$$

Por último, del axioma de la diferencia y el de extensionalidad deducimos

$$\bigwedge xy\overset{1}{\bigvee} z\bigwedge u(u \in z \leftrightarrow u \in x \wedge u \notin y),$$

lo que nos permite definir

$$x \setminus y \equiv z|\bigwedge u(u \in z \leftrightarrow u \in x \wedge u \notin y),$$

y demostrar que

$$\bigwedge xyu(u \in x \setminus y \leftrightarrow u \in x \wedge u \notin y).$$

En particular vemos que $\bigwedge u\, u \notin x \setminus x$, luego $\bigvee y\bigwedge u\, u \notin y$, es decir, existe un conjunto sin elementos, y por el axioma de extensionalidad $\overset{1}{\bigvee} y\bigwedge u\, u \notin y$, ya que si hubiera dos conjuntos sin elementos, ambos tendrían los mismos elementos (ninguno), luego por el axioma de extensionalidad serían el mismo. Esto nos permite definir el *conjunto vacío* como

$$\varnothing \equiv y \mid \bigwedge u\, u \notin y,$$

y así DP nos da que $\bigwedge u\, u \notin \varnothing$.

Por otra parte, en B podemos demostrar que

$$\bigwedge xy\overset{1}{\bigvee} z\bigwedge u(u \in z \leftrightarrow u \in x \vee u \in y).$$

En efecto, para la existencia tomamos $z = \bigcup\{x, y\}$, y la unicidad la da el axioma de extensionalidad. Consecuentemente, podemos definir la *unión* de conjuntos mediante

$$x \cup y \equiv z|\bigwedge u(u \in z \leftrightarrow u \in x \vee u \in y),$$

y demostrar mediante DP que

$$\bigwedge xyu(u \in x \cup y \leftrightarrow u \in x \vee u \in y).$$

Por último, demostramos que $\bigwedge xy\overset{1}{\bigvee} z\bigwedge u(u \in z \leftrightarrow u \in x \wedge u \in y)$ tomando $z = (x \cup y) \setminus ((x \setminus y) \cup (y \setminus x))$ para la existencia y usando el axioma de extensionalidad para la unicidad. Esto nos lleva a definir la *intersección* de conjuntos como

$$x \cap y \equiv z|\bigwedge u(u \in z \leftrightarrow u \in x \wedge u \in y),$$

y demostrar que

$$\bigwedge xyu(u \in x \cap y \leftrightarrow u \in x \wedge u \in y).$$

Otro concepto conjuntista básico que podemos definir en B es el de *par ordenado:*

$$(x, y) \equiv \{\{x\}, \{x, y\}\}.$$

Con esta definición es pura rutina comprobar que

$$\bigwedge xyx'y'((x, y) = (x', y') \leftrightarrow x = x' \wedge y = y').$$

3.2.2 Clases y conjuntos

Podemos preguntarnos si existe un conjunto que contiene a todos los conjuntos, es decir, si $\bigvee y \bigwedge x\ x \in y$. La respuesta es que la teoría B dista mucho de ser completa, y en ella no puede demostrarse ni esta sentencia ni su negación, pero, a pesar de ello, nada nos impide hablar en B de la *clase* de todos los conjuntos. Más en general, si $\phi(x)$ es cualquier fórmula de $\mathcal{L}_{\mathrm{tc}}$ (tal vez con más variables libres), escribiremos

$$A \equiv \{x \mid \phi(x)\}$$

para referirnos a la *clase* de todos los conjuntos que cumplen $\phi(x)$, con independencia de que exista o no un conjunto cuyos elementos sean los conjuntos que cumplen $\phi(x)$. En estos términos, la clase de todos los conjuntos es

$$V \equiv \{x \mid x = x\}.$$

En este punto el lector hará bien en rebelarse, porque esto que acabamos de decir no tiene rigor alguno. El lector tiene derecho a preguntar qué es exactamente una clase y se trata de una pregunta embarazosa, pero, como es habitual, ante las preguntas embarazosas sobre qué sentido tiene lo que hacemos podemos dar las respuestas "oficiales" (formales) y luego, cuando el formalismo nos saca del apuro, podemos abordar la cuestión semánticamente ya sin compromiso alguno. Empecemos con la parte formal:

Si $\phi(x)$ es una fórmula de $\mathcal{L}_{\mathrm{tc}}$ (tal vez con más variables libres), usaremos la notación

$$A \equiv \{x \mid \phi(x)\}$$

(léase "A es la clase de todos los conjuntos que cumplen $\phi(x)$") para expresar que cuando escribamos $x \in A$ no querremos decir ni más ni menos que $\phi(x)$. Más precisamente, si $B = \{x \mid \psi(x)\}$ es otra clase, entenderemos que:

1. $x \in A \equiv \phi(x)$.
2. $y = A \equiv \bigwedge x(x \in y \leftrightarrow x \in A) \equiv \bigwedge x(x \in y \leftrightarrow \phi(x))$.
3. $A = B \equiv \bigwedge x(x \in A \leftrightarrow x \in B) \equiv \bigwedge x(\phi(x) \leftrightarrow \psi(x))$.
4. $A \in z \equiv \bigvee y \in z\ y = A \equiv \bigvee y \in z \bigwedge x(x \in y \leftrightarrow \phi(x))$.
5. $A \in B \equiv \bigvee y(y = A \wedge y \in B) \equiv \bigvee y(\bigwedge x(x \in y \leftrightarrow \phi(x)) \wedge \psi(y))$.

Notemos que si una fórmula contiene una variable, ésta aparece necesariamente en subfórmulas de tipo $x = y$ o $x \in y$, por lo que si en cualquier fórmula de $\mathcal{L}_{\mathrm{tc}}$ sustituimos algunas de sus variables libres por una o varias clases, la expresión resultante nombra a una fórmula concreta de $\mathcal{L}_{\mathrm{tc}}$, la que resulta de sustituir las subfórmulas atómicas que contienen clases por las definiciones que acabamos de dar.

Por ejemplo, si escribimos $A \subset B$, no hay ninguna duda sobre a qué fórmula de $\mathcal{L}_{\mathrm{tc}}$ nos estamos refiriendo: según la definición de la inclusión se trata de la fórmula

$$\bigwedge x(x \in A \to x \in B) \equiv \bigwedge x(\phi(x) \to \psi(x)).$$

Para interpretar un término t definido en términos de clases aplicaremos el criterio anterior a una fórmula equivalente a $x \in t$ que no tenga descriptores[8]. Por ejemplo, para interpretar qué es $A \cup B$ observamos que

$$x \in y \cup z \leftrightarrow x \in y \vee x \in z,$$

y la fórmula de la derecha ya no tiene descriptores, por lo que interpretamos

$$x \in A \cup B \leftrightarrow x \in A \vee x \in B \leftrightarrow \phi(x) \vee \psi(x).$$

Podemos resumir esto diciendo que

$$A \cup B \equiv \{x \mid x \in A \vee x \in B\}.$$

En resumen: cada clase está definida a partir de una fórmula, y una fórmula que contiene clases se puede interpretar inequívocamente como una fórmula "normal" de $\mathcal{L}_{\mathrm{tc}}$ sin más que sustituir cada aparición de una clase en una subfórmula atómica por la fórmula que hemos indicado. En particular, cuando decimos que dos clases son iguales ($A = B$) esto simplemente significa que las fórmulas que las definen son equivalentes.

Esto zanja completamente el problema desde un punto de vista formal: mencionar clases es riguroso porque cualquier enunciado que mencione clases hace referencia a una fórmula de $\mathcal{L}_{\mathrm{tc}}$ perfectamente determinada.

Pero todavía no hemos terminado: vamos a introducir una notación adicional al respecto de las clases. Si tenemos una clase $A = \{x \mid \phi(x)\}$, diremos que *"A es un conjunto"* para referirnos a la fórmula de $\mathcal{L}_{\mathrm{tc}}$

$$\bigvee y \bigwedge x(x \in y \leftrightarrow x \in A) \equiv \bigvee y \bigwedge x(x \in y \leftrightarrow \phi(x)),$$

y en tal caso identificaremos la clase A con el conjunto y, que será único por el axioma de extensionalidad. En otras palabras, cuando decimos que una clase (definida por una fórmula $\phi(x)$) es un conjunto, queremos decir que existe un conjunto cuyos elementos son justamente los conjuntos que cumplen $\phi(x)$.

Si una clase A no es un conjunto diremos que es una *clase propia.* Concretamente, esto significa que no existe ningún conjunto cuyos elementos sean todos los conjuntos que satisfacen la fórmula que define la clase. Puesto que simplemente estamos adoptando ciertos convenios para nombrar ciertas fórmulas de $\mathcal{L}_{\mathrm{tc}}$, lo que hacemos es completamente riguroso y no tenemos obligación de dar más explicaciones.

Y ahora, una vez libres de toda obligación, vamos a considerar las clases desde un punto de vista semántico por pura curiosidad. Para ello supongamos que conocemos un modelo M de B (más adelante mostraremos uno explícitamente). Eso significa que M nos da una interpretación posible de los términos "conjunto" y "pertenencia". Respecto a esta interpretación, los conjuntos son los objetos del universo de M.

[8] Veremos más adelante que siempre es posible encontrar una fórmula así (teorema 3.32 y observaciones posteriores).

Ya hemos dicho que cada conjunto a puede verse como una colección de objetos, a saber, su extensión, la colección de los objetos de M que pertenecen a a (con respecto a la relación de pertenencia fijada por M), y ahora añadimos que a cada clase $A = \{x \mid \phi(x)\}$ también podemos asociarle una interpretación en M, a saber, la colección de todos los objetos a de M que cumplen $M \vDash \phi[v_x^a]$ (para cierta valoración que será relevante si ϕ tiene otros parámetros). Dicha colección (a la que podemos llamar también la *extensión* de A) estará perfectamente definida (supuesto que el modelo M lo esté), pero puede, o no, ser la extensión de un conjunto, y eso es lo que significa que la clase A sea o no un conjunto respecto del modelo M.

Así pues, desde un punto de vista semántico (respecto de un modelo prefijado de B) tanto las clases como los conjuntos pueden verse como colecciones de conjuntos. Un conjunto es la extensión de uno de los objetos del modelo y una clase propia es una colección de conjuntos que tienen en común satisfacer una determinada fórmula (con unos parámetros prefijados), pero que no es la extensión de un conjunto del modelo.

Esto tiene un correlato puramente formal: notemos que, de acuerdo con las definiciones que hemos dado, las fórmulas $A \in z$ o $A \in B$ (donde A y B son clases), implican que la clase A es un conjunto. Como, por otra parte, todo conjunto x pertenece al menos a otro conjunto ($x \in \{x\}$), podemos decir que una clase es un conjunto si y sólo si pertenece a otra clase o conjunto.

Ejemplo Ahora podemos ver en qué queda la paradoja de Russell en la teoría B. Si en B (o en otra teoría de conjuntos cualquiera) pudiéramos demostrar que $\bigvee y \bigwedge u(u \in y \leftrightarrow u \notin u)$ (es decir, que existe un conjunto cuyos elementos son los conjuntos que no se pertenecen a sí mismos) tendríamos una contradicción, pues dicho conjunto y debería cumplir $y \in y \leftrightarrow y \notin y$. Pero esta observación no nos lleva a una contradicción, sino que simplemente constituye la demostración en B de la fórmula

$$\neg \bigvee y \bigwedge u(u \in y \leftrightarrow u \notin u).$$

Equivalentemente, no existe ningún conjunto cuyos elementos sean los conjuntos que no se pertenecen a sí mismos. Podemos expresar este hecho en términos de clases definiendo $R = \{x \mid x \notin x\}$. Lo que acabamos de probar es que no hay ningún conjunto cuyos elementos sean los de R o, lo que es lo mismo, la clase R no es un conjunto. En particular $R \notin R$, pero esto no implica que $R \in R$, pues, de acuerdo con las definiciones que hemos dado, $R \in R$ exigiría que R fuera un conjunto, y ya hemos visto que no lo es.

El hecho de que R no sea un conjunto se interpreta como que no hay ningún objeto en M cuyos elementos sean precisamente los elementos de R. No hay contradicción alguna en ello: la extensión de la clase R en un modelo de B no puede ser la extensión de un conjunto. ∎

Según acabamos de ver, hay fórmulas (como $x \notin x$, aunque más adelante veremos ejemplos más sustanciosos) que definen clases propias en cualquier modelo, es decir, fórmulas que son satisfechas en un modelo por colecciones de conjuntos (bien definidas) que no son la extensión de ningún conjunto. De aquí

se sigue inevitablemente que es del todo inviable identificar "conjunto" con "colección de objetos". El lector debe entender que el término "conjunto" es un puro tecnicismo. No podemos acabar ninguna frase que empiece por "un conjunto es una colección de objetos tal que..." Sencillamente, los conjuntos, en el sentido en que los matemáticos usan esa palabra, son "ciertas" colecciones de elementos, pero no unas en concreto, sino que cada modelo de una teoría de conjuntos contiene unos objetos que son susceptibles de ser llamados "conjuntos" y, dentro de cada modelo, ciertas colecciones de conjuntos serán conjuntos y otras no.

En realidad la situación puede ser "peor": si definimos de algún modo una colección A de algunos de los objetos de su universo, podemos encontrarnos con tres situaciones: 1) que A sea la extensión de un conjunto de M, 2) que A no sea la extensión de un conjunto de M pero sí la de una clase, es decir, que sea la colección de los conjuntos de M que satisfacen una determinada fórmula, o 3) que A no sea ni la extensión de un conjunto ni la de una clase.[9]

Nota En este punto el lector puede revisar la sección A.1 del apéndice A en la que se exponen los principales conceptos conjuntistas, todos los cuales pueden ser formalizados en B en términos de clases.

3.2.3 Números naturales

A pesar de su simplicidad, la teoría B permite definir los números naturales. La idea básica es que podemos definir

$$x' \equiv x \cup \{x\}$$

Con esta definición de "sucesor" podemos ir definiendo

$$0 \equiv \varnothing, \quad 1 \equiv 0' = 0 \cup \{0\} = \{0\}, \quad 2 \equiv 1' = 1 \cup \{1\} = \{0, 1\}, \quad 3 \equiv 2' = \{0, 1, 2\}$$

y así sucesivamente. El problema es que "y así sucesivamente" no sirve. Las teorías formales nos permiten escaquearnos de definir sus conceptos primitivos, pero eso es a cambio del compromiso de razonar únicamente deduciendo fórmulas a partir de los axiomas, y si queremos demostrar algo así como que "el siguiente de un número natural es un número natural", necesitamos haber definido una fórmula $\mathrm{Nat}\, x$, con la que podamos enunciar y demostrar $\bigwedge x(\mathrm{Nat}\, x \to \mathrm{Nat}\, x')$, y el "así sucesivamente" no nos da una fórmula de $\mathcal{L}_{\mathrm{tc}}$. Para encontrar esa fórmula empezamos definiendo algunos conceptos:

Definición 3.9 Diremos que un conjunto x es

1. *transitivo* si $\bigwedge u \in x\ u \subset x$,
2. $\in$*-conexo* si $\bigwedge uv \in x(u \in v \vee v \in u \vee u = v)$,

[9] Al final del capítulo siguiente veremos que toda teoría de conjuntos consistente suficientemente potente tiene siempre modelos en los que se da este tercer caso.

3. *bien fundado* si $\bigwedge u(u \subset x \wedge u \neq \varnothing \to \bigvee v \in u\, v \cap u = \varnothing)$. Un conjunto v que cumpla esta definición recibe el nombre de *elemento $\in$-minimal* de u.

4. x es un *ordinal* si cumple las tres propiedades anteriores.

Llamaremos Ω a la clase de todos los ordinales.

Con más detalle: un conjunto x es transitivo si todos sus elementos son también subconjuntos suyos. Una forma equivalente de expresar esta propiedad es $\bigwedge uv(u \in v \wedge v \in x \to u \in x)$, y de ahí el nombre de "transitividad" (de la pertenencia).

Un conjunto es $\in$-conexo si dos cualesquiera de sus elementos están "conectados" por la pertenencia, en el sentido de que uno pertenece al otro (entendiendo que hablamos de dos elementos distintos). Notemos que si y es $\in$-conexo y $x \subset y$ entonces x también es $\in$-conexo, pues dos de sus elementos son también elementos de y, luego están conectados por la pertenencia.

Por último, un conjunto x está bien fundado si cada subconjunto no vacío u tiene un elemento $\in$-minimal, es decir un $v \in u$ tal que ningún elemento $w \in u$ pertenece a v. También se cumple que si y está bien fundado y $x \subset y$ entonces x está bien fundado, pues todo $u \subset x$ no vacío cumple también $u \subset y$, luego tiene un $\in$-minimal por la buena fundación de y. Necesitaremos una propiedad elemental de los conjuntos bien fundados:

Teorema 3.10 *Si x es un conjunto bien fundado entonces $x \notin x$.*

DEMOSTRACIÓN: Si $x \in x$ entonces $\{x\} \subset x \wedge \{x\} \neq \varnothing$. Sea u un elemento $\in$-minimal de $\{x\}$. Necesariamente, $u = x$, pero $x \in x \cap \{x\}$, contradicción. ∎

Con esto podemos probar:

Teorema 3.11 $0 \in \Omega \wedge \bigwedge x \in \Omega\ x' \in \Omega$.

DEMOSTRACIÓN: Notemos que $0 = \varnothing$ cumple trivialmente las tres condiciones de la definición de ordinal (es transitivo porque no existe ningún $u \in \varnothing$ que pueda incumplir la definición, es $\in$-conexo porque no existen $u, v \in \varnothing$ que puedan incumplir la definición, y está bien fundado porque no existe ningún $u \subset \varnothing$, $u \neq \varnothing$ que pueda incumplir la definición).

Supongamos ahora que x es un ordinal. Si $u \in x' = x \cup \{x\}$, entonces $u \in x \vee u = x$, pero en ambos casos $u \subset x$, en el primero porque x es transitivo. Esto prueba que x' es transitivo.

Si $u, v \in x'$, entonces $u \in x \vee u = x$ y $v \in x \vee v = x$. Esto nos da cuatro casos: $u \in x \wedge v \in x$ o bien $u \in x \wedge v = x$, o bien $u = x \wedge v \in x$, o bien $u = x = v$. En el primero tenemos que $u \in v \vee v \in u \vee u = v$ porque x es $\in$-conexo, y en los otros tres tenemos $u \in v$, $v \in u$, $u = v$ respectivamente. Esto prueba que x' es $\in$-conexo.

Tomemos $u \subset x' \wedge u \neq \varnothing$ y veamos que tiene $\in$-minimal. Tratemos aparte el caso en que $u = \{x\}$. Entonces $v = x$ es un $\in$-minimal de u, pues $x \cap \{x\} = \varnothing$. En efecto, si existiera $w \in x \cap \{x\}$, sería $x = w \in x$, en contradicción con el teorema anterior.

Como $u \subset x \cup \{x\}$, si no se da la igualdad $u = \{x\}$ es porque $u \cap x \neq \varnothing$, y tenemos así un subconjunto no vacío de x. Como x está bien fundado existe un $v \in u \cap x$ que es $\in$-minimal para esta intersección. Vamos a ver que es $\in$-minimal de u.

En efecto, si $w \in v \cap u$, entonces $w \in x'$, luego $w \in x \vee w = x$. En el primer caso $w \in u \cap x$ y $w \in v$, lo que contradice que v sea $\in$-minimal de $u \cap x$. En el segundo caso $x = w \in v \in x$, luego, por la transitividad de x, resulta que $x \in x$, en contradicción con el teorema anterior. ∎

En vista de este teorema resulta que 0 es un ordinal, luego $1 = 0'$ es un ordinal, luego $2 = 1'$ es un ordinal y, en definitiva, todos los números naturales son ordinales, pero no podemos demostrar tal cosa porque no tenemos una definición de número natural. En realidad vamos a definir los números naturales como los ordinales que cumplen una propiedad adicional, pero antes de ello necesitamos demostrar algunas propiedades sobre ordinales.

Notemos que, por ejemplo, $5 = \{0, 1, 2, 3, 4\}$, de modo que los elementos del ordinal 5 son otros ordinales. Esto es cierto en general:

Teorema 3.12 *Los elementos de los ordinales son ordinales.*

Demostración: Sea $y \in \Omega$ y sea $x \in y$. Por transitividad $x \subset y$ y por consiguiente x es conexo y bien fundado. Falta probar que es transitivo, es decir, que $\bigwedge uv(u \in v \wedge v \in x \to u \in x)$.

Si $u \in v \wedge v \in x$, tenemos $v \in x \wedge x \in y$, y como y es transitivo, $v \in y$, e igualmente $u \in y$. Así pues, $\{u, v, x\} \subset y$. Como y está bien fundado se cumplirá

$$u \cap \{u, v, x\} = \varnothing \quad \vee \quad v \cap \{u, v, x\} = \varnothing \quad \vee \quad x \cap \{u, v, x\} = \varnothing,$$

pero $u \in v \cap \{u, v, x\}$ y $v \in x \cap \{u, v, x\}$, luego ha de ser $u \cap \{u, v, x\} = \varnothing$. Como y es conexo ha de ser $u \in x \vee x \in u \vee u = x$, pero si $x \in u$ entonces $x \in u \cap \{u, v, x\} = \varnothing$, y si $x = u$ entonces $v \in u \cap \{u, v, x\} = \varnothing$. Así pues, se ha de cumplir $u \in x$, como queríamos. ∎

Notemos que el teorema anterior puede enunciarse así:

$$\bigwedge xy(x \in y \wedge y \in \Omega \to x \in \Omega),$$

pero esto es lo mismo que decir que la clase Ω es transitiva.

Definición 3.13 Si x, y son ordinales, escribiremos $x \leq y \equiv x \subset y$.

Alternativamente, podríamos definir una clase

$$\leq \equiv \{(x, y) \in \Omega \times \Omega \mid x \subset y\},$$

y trivialmente se trata de una relación de orden (porque la inclusión siempre es una relación de orden), pero restringida a cada ordinal resulta ser una relación de orden total. En efecto, si $x \in \Omega$ y $u, v \in x$, entonces $u \in v \vee v \in u \vee u = v$, y como sabemos que u, v son ordinales, en particular son transitivos, y resulta que $u \subset v \vee v \subset u$. Más aún:

Teorema 3.14 *Si x es un ordinal y $\varnothing \neq u \subset x$, entonces todo $\in$-minimal v de u es, de hecho, el mínimo de u para la relación de inclusión. Por lo tanto todo ordinal está bien ordenado por la inclusión.*

DEMOSTRACIÓN: En principio tenemos que $v \cap u = \varnothing$. Si $w \in u \subset x$, entonces $v, w \in x$, luego $w \in v \vee v \in w \vee v = w$. El caso $w \in v$ no puede darse, pues implicaría que $w \in v \cap u = \varnothing$. En los otros dos casos, como w es un ordinal (por ser elemento de x) es transitivo y se cumple que $v \subset w$, es decir, $v \leq w$. Así pues, v es menor o igual que cualquier elemento $w \in u$. Como un conjunto ordenado sólo puede tener un mínimo, u tiene un único $\in$-minimal. ∎

Cada número natural (en el sentido en que pensamos definirlos en B) está formado por los números menores que él. Así, el hecho de que (informalmente) $3 < 5$, se traduce en B en que $3 \in 5$. Para sacarle partido a esta observación demostramos lo siguiente:

Teorema 3.15 $\bigwedge xy \in \Omega(x \leq y \to x \in y \vee x = y)$.

DEMOSTRACIÓN: Si $x \neq y$ entonces $y \setminus x \neq \varnothing$. Como y es un ordinal, $y \setminus x$ tiene un elemento $\in$-minimal $u \in y \setminus x$, de modo que $u \cap (y \setminus x) = \varnothing$. Basta probar que $u = x$, pues entonces tenemos que $x \in y$.

Si $z \in u$, entonces $z \notin y \setminus x$ y $z \in y$ (por transitividad, pues $z \in u \in y$), luego $z \in x$. Por lo tanto $u \subset x$.

Si $z \in x$, entonces tenemos $z, u \in y$, luego $z \in u \vee u \in z \vee z = u$. Si $u \in z$, entonces $u \in z \in x$, luego $u \in x$, contradicción ($u \in y \setminus x$). Si $z = u$ entonces de nuevo $u \in x$, contradicción. Por lo tanto $z \in u$, y así $x \subset u$. En definitiva, tenemos la igualdad $u = x$. ∎

Notemos que no pueden darse a la vez los dos casos $x \in y \wedge x = y$, pues esto supondría que $y \in y$, cuando hemos probado que un conjunto bien fundado no se pertenece a sí mismo. Por lo tanto, si definimos

$$x < y \equiv x \leq y \wedge x \neq y,$$

tenemos que si x, y son ordinales entonces $x < y \leftrightarrow x \in y$.

Teorema 3.16 *La intersección de dos ordinales es un ordinal.*

DEMOSTRACIÓN: Sean x, y ordinales. Como $x \cap y \subset x$, trivialmente $x \cap y$ es conexo y bien fundado. Falta ver que es transitivo. En efecto:

Si $u \in x \cap y$, entonces $u \in x \wedge u \in y$, $u \subset x \wedge u \subset y$, luego $u \subset x \cap y$. ∎

Con esto podemos probar un resultado no trivial. Hasta ahora sabemos que si x y y son dos elementos de un ordinal (dos ordinales que pertenecen a otro ordinal), entonces $x \in y \vee y \in x \vee x = y$. Ahora podemos probar que esto es cierto sin suponer que x e y son elementos de otro ordinal:

Teorema 3.17 $\bigwedge xy \in \Omega\,(x \in y \vee y \in x \vee x = y)$.

DEMOSTRACIÓN: $x \cap y$ es un ordinal, $x \cap y \subset x$ y $x \cap y \subset y$. Por el teorema 3.15 tenemos $(x \cap y \in x \vee x \cap y = x) \wedge (x \cap y \in y \vee x \cap y = y)$. Esto nos da cuatro casos:

$$(x \cap y \in x \wedge x \cap y \in y) \vee (x \cap y \in x \wedge x \cap y = y)$$

$$\vee (x \cap y = x \wedge x \cap y \in y) \vee (x \cap y = x \wedge x \cap y = y),$$

o sea $x \cap y \in x \cap y \quad \vee \quad y \in x \quad \vee \quad x \in y \quad \vee \quad x = y$. El primer caso se descarta por el teorema 3.10. ∎

Notemos que el teorema anterior es lo que significa precisamente la sentencia "Ω es una clase $\in$-conexa". También hemos visto que es transitiva, luego estamos a un paso de probar lo siguiente:

Teorema 3.18 *Ω es un ordinal.*

DEMOSTRACIÓN: Sólo nos falta probar que Ω está bien fundada. Si tomamos la definición de "x está bien fundada" y sustituimos x por Ω vemos que lo que tenemos que probar es que

$$\bigwedge u(u \subset \Omega \wedge u \neq \varnothing \to \bigvee v \in u\, v \cap u = \varnothing),$$

donde a su vez hay que entender que $u \subset \Omega$ significa $\bigwedge w(w \in u \to w \in \Omega)$.

En definitiva, hemos de demostrar que si u es un conjunto no vacío cuyos elementos son ordinales, entonces tiene un $\in$-minimal. Podemos tomar $w \in u$ (que será entonces un ordinal) y distinguimos dos casos. Si w es un $\in$-minimal de u, no hay nada que probar. En caso contrario, existe un $z \in w \cap u$, luego $w \cap u \neq \varnothing$. Así $w \cap u$ es un subconjunto no vacío de w, que es un ordinal, luego existe $v \in w \cap u$ que es $\in$-minimal para la intersección. Basta probar que v es $\in$-minimal para u.

En caso contrario existiría un $z \in v \cap u$, pero $z \in v \in w$ y w es transitivo, luego $z \in w$, luego $z \in v \cap (w \cap u)$, en contradicción con la $\in$-minimalidad de v en $w \cap u$. ∎

De aquí podemos deducir un hecho interesante:

Teorema 3.19 *Ω es una clase propia.*

DEMOSTRACIÓN: El enunciado significa que no existe ningún conjunto cuyos elementos sean todos los ordinales. La demostración consiste en observar que si x fuera tal conjunto, toda la prueba del teorema anterior valdría para concluir que x es un ordinal, es decir, que $x \in \Omega$, y entonces debería cumplirse $x \in x$, en contradicción con el teorema 3.10. ∎

Este teorema es la versión atenuada de lo que en la teoría de conjuntos ingenua de Cantor era una paradoja conocida como la *Antinomia de Burali-Forti*. En efecto, Cantor definió los ordinales de forma muy distinta a como lo hemos hecho aquí, pero la situación era equivalente: el "conjunto" de todos los

ordinales debía ser el mayor ordinal posible, pero por otra parte se demostraba que era un ordinal, y como tal debía tener un siguiente. El error estaba en pretender que cualquier colección bien definida puede considerarse un conjunto, como si "conjunto" fuera lo mismo que "colección".

Así tenemos una prueba manifiesta de que quien identifique ingenuamente "conjunto" con "colección de objetos", no puede entender nada de lo que estamos diciendo y, lo que es peor, se verá abocado a contradicciones, pues se encuentra con una demostración de que la "colección de todos los ordinales" no es un conjunto y, al mismo tiempo, (supuestamente) es un conjunto (porque es una colección).

En nuestro contexto no hay contradicción alguna. Desde un punto de vista formal tenemos meramente la constatación irrefutable de que no podemos pretender que cualquier fórmula defina un conjunto. Desde un punto de vista semántico, si tenemos un modelo M de B, la clase Ω se interpreta en M como una colección de conjuntos bien definida (la colección de los conjuntos que satisfacen la definición de ordinal), pero resulta que no es (la extensión de) ninguno de los conjuntos de M, por lo que, aunque tenga las propiedades que definen a los ordinales, no se le pueden aplicar los teoremas de B (que son válidos para todos los objetos de M, entre los que no está la interpretación de Ω) y por ello no podemos afirmar que Ω tenga un siguiente, ni tampoco que $\Omega \in \Omega$ (pues Ω es una colección de objetos de M y Ω no es uno de ellos).

Antes de pasar a definir los números naturales demostramos un último teorema sobre ordinales. En lo sucesivo, cuando digamos que α es un ordinal se entenderá que α es un conjunto que es un ordinal.

Teorema 3.20 *Se cumple:*

1. 0 *es el mínimo ordinal.*
2. *Si α es un ordinal, entonces α' también lo es, y es el mínimo ordinal mayor que α (es decir, $\bigwedge\beta \in \Omega(\alpha < \beta \to \alpha' \leq \beta)$.*
3. *Todo conjunto de ordinales $x \subset \Omega$ tiene supremo $\sigma = \bigcup x$.*

DEMOSTRACIÓN: 1) ya hemos probado que 0 es un ordinal, y es el mínimo porque el conjunto vacío está contenido en cualquier conjunto.

2) Ya hemos probado que $\alpha' \in \Omega$. Si $\alpha < \beta$ entonces $\alpha \in \beta$, luego $\alpha \subset \beta$, luego $\alpha' = \alpha \cup \{\alpha\} \subset \beta$, luego $\alpha' \leq \beta$.

3) Como todo $\alpha \in x$ está contenido en Ω, es claro que $\sigma \subset \Omega$, luego es un conjunto conexo y bien fundado. Hemos de probar que es transitivo, pero si $\beta \in \sigma$, entonces existe un $\alpha \in x$ tal que $\beta \in \alpha$, luego por la transitividad de α es $\beta \subset \alpha \subset \sigma$. Por consiguiente $\sigma \in \Omega$. Teniendo en cuenta que el orden es la inclusión, es inmediato que σ es el supremo de x. ∎

Definición 3.21 Un conjunto n es un *número natural* si cumple:

$$n \in \Omega \wedge \bigwedge\alpha(\alpha \in n' \to \alpha = 0 \vee \bigvee\beta \in \alpha\, \alpha = \beta').$$

o sea, un número natural n es un ordinal tal que todos los ordinales $0 < \alpha \leq n$ tienen un inmediato anterior (es decir, son el siguiente de otro).

Tenemos así definida una fórmula precisa de $\mathcal{L}_{\mathrm{tc}}$, la cual nos permite a su vez considerar la clase de todos los números naturales, a la que llamaremos ω.

Teorema 3.22 *ω es un ordinal (aunque no necesariamente un conjunto).*

DEMOSTRACIÓN: Como $\omega \subset \Omega$, es trivialmente una clase $\in$-conexa y bien fundada, y basta ver que es transitiva. Si $u \in v \wedge v \in \omega$, entonces v es un número natural. Por definición tenemos que

$$\bigwedge\alpha(\alpha \in v' \to \alpha = 0 \vee \bigvee\beta \in \alpha\, \alpha = \beta'),$$

como $u < v$, se cumple que $u' \leq v < v'$ y en particular

$$\bigwedge\alpha(\alpha \in u' \to \alpha = 0 \vee \bigvee\beta \in \alpha\, \alpha = \beta'),$$

luego $u \in \omega$. ∎

Teorema 3.23 (Axiomas de Peano) *Se cumple:*

1) $0 \in \omega$,

2) $\bigwedge n \in \omega\, n' \in \omega$,

3) $\bigwedge n \in \omega\, n' \neq 0$,

4) $\bigwedge mn \in \omega(m' = n' \to m = n)$,

5) $\bigwedge y(y \subset \omega \wedge 0 \in y \wedge \bigwedge n \in y\ n' \in y \to y = \omega)$.

DEMOSTRACIÓN: 1) es trivial.

2) si $n \in \omega$ y $\alpha \in n''$, entonces, o bien $\alpha \in n'$ o bien $\alpha = n'$. En el primer caso $\alpha = 0 \vee \bigvee\beta \in \alpha\, \alpha = \beta'$, porque $n \in \omega$. Esto también se cumple en el segundo caso, tomando $\beta = n$. Por consiguiente $n' \in \omega$.

Las propiedades 3) y 4) son trivialmente válidas para ordinales cualesquiera, pues $0 \leq n < n'$, luego $0 \in n'$, luego $n' \neq 0$. Por otra parte, si $m' = n'$, tiene que ser $m = n$, ya que si fuera $m < n$ entonces $m' \leq n < n'$, luego $m' \neq n'$, e igualmente si $n < m$.

5) Si $y \subset \omega \wedge 0 \in y \wedge \bigwedge n \in y\, n' \in y$ pero $y \neq \omega$, tomemos $z \in \omega \setminus y$. Entonces $z \in z' \setminus y \subset z'$. Como z' es un ordinal, existe un $\in$-minimal $n \in z' \setminus y$. No puede ser $n = 0$, pues $0 \in y$, $n \notin y$. Como n es un número natural, por definición existe un $m \in n$ tal que $n = m'$. Como n es minimal, no puede ser que $m \in z' \setminus y$, pues entonces $m \in n \cap (z' \setminus y)$. Por lo tanto $m \in y$ (notemos que $m \in n \in z'$, luego $m \in z'$, por transitividad). Pero estamos suponiendo que $m \in y$ implica $n = m' \in y$, contradicción. ∎

En B no puede probarse que ω sea un conjunto. Las teorías de conjuntos que manejan habitualmente los matemáticos incluyen un "axioma de infinitud" que postula la existencia de conjuntos infinitos. Admite varias versiones equivalentes, y una de ellas es precisamente que ω es un conjunto. Sin embargo,

nosotros evitaremos ese axioma mientras nos sea posible por una razón: gran parte de la matemática básica puede demostrarse en teorías de conjuntos sin el axioma de infinitud, y a todas las teorías de conjuntos que consideraremos sin dicho axioma les podremos encontrar un modelo, es decir, tendremos la garantía de que son consistentes. En cambio, cuando añadimos el axioma de infinitud la situación es mucho más problemática.

Notemos que si ω es un conjunto, entonces $\omega \in \Omega$, y tenemos un ejemplo de ordinal que no es un número natural.

Terminamos aquí esta sección porque poco más puede decirse en B sobre ordinales y números naturales. Ni siquiera es posible definir la suma de números naturales. Para ello necesitamos un axioma más.

3.3 La teoría de Zermelo

Puestos a añadir axiomas a nuestra teoría de conjuntos B, resulta natural plantearse lo que podemos llamar el *axioma de comprensión,* que es en realidad el esquema axiomático que, para cada fórmula $\phi(u)$, (tal vez con más variables libres) postula que

$$\bigvee y \bigwedge u(u \in y \leftrightarrow \phi(u)).$$

Si combinamos esto con el axioma de extensionalidad resulta que y es único, por lo que podemos definir

$$\{u \mid \phi(u)\} \equiv y | \bigwedge u(u \in y \leftrightarrow \phi(u))$$

y tenemos en definitiva que cada fórmula $\phi(u)$ define un conjunto: el conjunto $\{u \mid \phi(u)\}$ formado por todos los conjuntos que cumplen $\phi(u)$. Desgraciadamente, este axioma es contradictorio, pues permite formalizar la paradoja de Russell: tomando $\phi(u) \equiv u \notin u$, el axioma de comprensión implica la existencia del conjunto

$$R = \{u \mid u \notin u\},$$

y de aquí se sigue inmediatamente la contradicción $R \in R \leftrightarrow R \notin R$.

Así pues, si queremos una teoría de conjuntos consistente, no podemos suponer que todas las fórmulas definen conjuntos. Zermelo encontró un sustituto del axioma de comprensión que no permite formalizar la paradoja de Russell. Consiste en suponer que todas las fórmulas especifican subconjuntos de cualquier conjunto dado, en el sentido que concretamos en la definición siguiente:

Definición 3.24 Llamaremos *teoría de conjuntos (restringida) de Zermelo* a la teoría axiomática Z^* sobre el lenguaje $\mathcal{L}_{\mathrm{tc}}$ cuyos axiomas son:

Extensionalidad	$\bigwedge xy(\bigwedge u(u \in x \leftrightarrow u \in y) \to x = y)$
Par	$\bigwedge xy \bigvee z \bigwedge u(u \in z \leftrightarrow u = x \vee u = y)$
Unión	$\bigwedge x \bigvee y \bigwedge u(u \in y \leftrightarrow \bigvee v(u \in v \wedge v \in x))$
Especificación	$\bigwedge x \bigvee y \bigwedge u(u \in y \leftrightarrow u \in x \wedge \phi(u))$

donde $\phi(u)$ es cualquier fórmula, tal vez con más variables libres (pero distintas de y).[10]

Así, el esquema de especificación afirma que toda fórmula especifica un subconjunto de cada conjunto x, a saber, el conjunto y formado por los elementos de x que cumplen $\phi(u)$. Nuevamente, el axioma de extensionalidad nos da que el conjunto y es único, por lo que podemos definir

$$\{u \in x \mid \phi(u)\} \equiv y|\bigwedge u(u \in y \leftrightarrow u \in x \wedge \phi(u)),$$

que siempre es un conjunto, sea cual sea la fórmula $\phi(u)$. Ahora bien, ya hemos visto que no podemos relajar el "$\{u \in x \mid \cdots$" por un mero "$\{u \mid \cdots$".

En particular, en Z^* podemos probar el axioma de la diferencia, que es el único axioma de la teoría básica B que falta en Z^*. En efecto, la diferencia de dos conjuntos x e y es el conjunto $\{u \in x \mid u \notin y\}$.

A partir de aquí tenemos que todos los teoremas de B son también teoremas de Z^*. En particular, en Z^* está definida la clase ω de los números naturales (pero seguimos sin poder probar que es un conjunto).

Notas El esquema de especificación puede enunciarse diciendo que toda subclase de un conjunto es un conjunto. En efecto, una clase es de la forma $A = \{u \mid \phi(u)\}$, para cierta fórmula ϕ, y la fórmula $A \subset x$ significa que

$$\bigwedge u \in A(u \in x) \qquad \text{o, equivalentemente,} \qquad \bigwedge u(\phi(u) \to u \in x).$$

Por lo tanto, A tiene los mismos elementos que el conjunto $\{u \in x \mid \phi(u)\}$, y esto es lo que significa precisamente que A sea un conjunto.

Otra consecuencia es que la clase V de todos los conjuntos no puede ser un conjunto. En efecto, si lo fuera, también lo sería

$$R = \{x \mid x \notin x\} = \{x \in V \mid x \notin x\},$$

y tendríamos la paradoja de Russell: $R \in R \leftrightarrow R \notin R$. ∎

La relación entre fórmulas y conjuntos que establece el esquema de especificación nos permite demostrar en Z^* el principio de inducción de AP:

Teorema 3.25 *Para toda fórmula $\phi(n)$ (tal vez con más variables libres), la fórmula siguiente es un teorema de Z^*:*

$$\phi(0) \wedge \bigwedge n \in \omega\,(\phi(n) \to \phi(n')) \to \bigwedge n \in \omega\,\phi(n).$$

Demostración: Supongamos que $\phi(0) \wedge \bigwedge n \in \omega\,(\phi(n) \to \phi(n'))$, pero que existe un $m \in \omega$ tal que $\neg\phi(m)$. Sea $x = \{i \in m' \mid \neg\phi(i)\}$. Se trata de un conjunto no vacío, luego por definición de ordinal existe un $i \in x$ tal que $i \cap x = \varnothing$. Entonces i es un número natural tal que $\neg\phi(i)$. Por lo tanto $i \neq 0$, luego existe un $n \in i$ tal que $i = n'$. Por hipótesis $\neg\phi(n)$, luego $n \in i \cap x$, contradicción. ∎

[10] En este tipo de esquemas axiomáticos o teoremáticos, supondremos siempre, aunque no lo digamos explícitamente, que los posibles parámetros de las fórmulas son variables distintas de las que aparecen explícitamente en el esquema.

Definiremos la suma y el producto de números naturales como casos particulares del teorema siguiente:

Teorema 3.26 *Sea X una clase, sea $G : \omega \times X \longrightarrow X$ y sea $a \in X$. Entonces existe una función $F : \omega \longrightarrow X$ tal que*

$$F(0) = a \wedge \bigwedge n \in \omega\ F(n') = G(n, F(n)).$$

DEMOSTRACIÓN: Definimos

$$F = \{(n, x) \in \omega \times X \mid \bigvee s(s : n' \longrightarrow X \wedge s(n) = x \wedge s(0) = a \wedge$$

$$\bigwedge i < n\ s(i') = G(i, s(i)))\}.$$

Veamos que F cumple lo pedido. Para ello probamos por inducción que

$$\bigwedge n \in \omega \bigvee x \in X\ (n, x) \in F.$$

En efecto, para $n = 0$ basta tomar $s = \{(0, a)\}$ y es claro que cumple lo necesario para justificar que $(0, a) \in F$. Si $(n, x) \in F$, existe s según la definición de F, y podemos considerar $s^* = s \cup \{(n', G(n, x))\}$, y es fácil ver que cumple lo requerido para justificar que $(n', G(n, x)) \in F$.

Si $s, s^* : n' \longrightarrow X$ cumplen la definición de F, entonces $\bigwedge i \in n'\ s(i) = s^*(i)$. Esto se prueba trivialmente por inducción sobre i. En particular, ahora es claro que $\bigwedge n \in \omega \overset{1}{\bigvee} x \in X\ (n, x) \in F$, es decir, que $F : \omega \longrightarrow X$. Ya hemos probado que $F(0) = a$. Dado $n \in \omega$, si $F(n') = x$, tomamos $s : n'' \longrightarrow X$ según la definición de F. Es claro que $s|_{n'} : n' \longrightarrow X$ también cumple la definición de F, luego $F(n) = s(n)$ y $F(n') = s(n')$. Por consiguiente concluimos que $F(n') = s(n') = G(n, s(n)) = G(n, F(n))$. ∎

Nota Si $X = \{u \mid \phi(u)\}$, $G = \{(n, u, v) \mid \psi(n, u, v)\}$, un enunciado sin clases propias del teorema anterior sería el siguiente: Dadas fórmulas $\phi(u)$ y $\psi(n, u, v)$ (tal vez con más variables libres), existe una fórmula $\chi(n, x, a)$ (con las mismas variables libres adicionales que las dos fórmulas dadas) tal que si

$$\phi(a) \wedge \bigwedge n \in \omega \bigwedge u(\phi(u) \to \overset{1}{\bigvee} v(\phi(v) \wedge \psi(n, u, v))),$$

entonces $\bigwedge n \in \omega \overset{1}{\bigvee} x(\phi(x) \wedge \chi(n, x, a)) \wedge \chi(0, a, a) \wedge$

$$\bigwedge n \in \omega \bigvee v_1 v_2(\phi(v_1) \wedge \phi(v_2) \wedge \chi(n, v_1, a) \wedge \chi(n', v_2, a) \wedge \psi(n, v_1, v_2)).$$

Concretamente:

$$\chi(n, x, a) \equiv n \in \omega \wedge \phi(x) \wedge \bigvee s(s \text{ es una función} \wedge \mathcal{D}s = n' \wedge \bigwedge i \in n'\ \phi(s(i))$$

$$\wedge\ s(n) = x \wedge s(0) = a \wedge \bigwedge i \in n\ \psi(i, s(i), s(i'))).$$

∎

Si aplicamos este teorema a la función sucesor $S : \omega \longrightarrow \omega$ (es decir, a la fórmula $\psi(n, x, y) \equiv y = x'$), para cada $m \in \omega$ obtenemos una $F_m : \omega \longrightarrow \omega$ que cumple

$$F_m(0) = m \wedge \bigwedge n \in \omega\ F_m(n') = F_m(n)'.$$

Definimos $m + n = F_m(n)$. En estos términos,[11]

$$\bigwedge m \in \omega\ m + 0 = m \wedge \bigwedge mn \in \omega\ m + n' = (m + n)'.$$

Similarmente, si partimos de la función $G_m : \omega \longrightarrow \omega$ definida mediante $G_m(n) = n + m$, el teorema anterior nos da una función $F_m : \omega \longrightarrow \omega$ tal que $F_m(0) = 0 \wedge \bigwedge n \in \omega\, F_m(n') = F_m(n)+m$. Si llamamos $m \cdot n = F_m(n)$, tenemos que

$$\bigwedge m \in \omega\ m \cdot 0 = 0 \wedge \bigwedge mn \in \omega\ m \cdot n' = (m \cdot n) + m.$$

En definitiva, hemos demostrado lo siguiente:

Teorema 3.27 *En* Z^* *se demuestran los axiomas de Peano:*

1. $0 \in \omega$,
2. $\bigwedge n \in \omega\ n' \in \omega$,
3. $\bigwedge n \in \omega\ n' \neq 0$,
4. $\bigwedge mn \in \omega(m' = n' \to m = n)$,
5. $\bigwedge m \in \omega\ m + 0 = m$,
6. $\bigwedge mn \in \omega\ m + n' = (m + n)'$,
7. $\bigwedge m \in \omega\ m \cdot 0 = 0$,
8. $\bigwedge mn \in \omega\ m \cdot n' = m \cdot n + m$,
9. $\phi(0) \wedge \bigwedge n \in \omega(\phi(n) \to \phi(n')) \to \bigwedge n \in \omega\, \phi(n)$,

para toda fórmula $\phi(n)$*, tal vez con más variables libres.*

3.4 Teorías aritméticas

La aritmética de Peano está pensada para hablar sólo de números naturales, de modo que en ella $\bigwedge n$ significa (o pretende significar) "para todo número natural n". Sin embargo, los matemáticos trabajan en teorías que permiten hablar de los números naturales "entre otras cosas". Es el caso de la teoría Z^*, en la que $\bigwedge x$ no pretende significar "para todo número natural x", sino "para todo conjunto x", de modo que los números naturales se identifican con ciertos conjuntos (los que satisfacen la fórmula $n \in \omega$). Para dar cuenta de esta situación introducimos el concepto siguiente:

[11] Más precisamente, el teorema anterior define una fórmula $\chi(n, x, m)$ tal que

$$F_m = \{(n, x) \mid \chi(n, x, m)\},$$

y entonces $m + n \equiv x|\chi(n, x, m)$.

Definición 3.28 Sea T una teoría axiomática sobre un lenguaje $\mathcal{L}$, sea $x \in \mathbb{N}$ una fórmula de $\mathcal{L}$ con x como única variable libre, sea 0 un designador de T, sean $x', x+y, x\cdot y$ tres términos de $\mathcal{L}$ cuyas variables libres sean las indicadas. Diremos que T, con estas expresiones, *interpreta a* $\mathcal{L}_a$ si las fórmulas siguientes son teoremas[12] de T:

1. $0 \in \mathbb{N}$
2. $\bigwedge n \in \mathbb{N}\ n' \in \mathbb{N}$
3. $\bigwedge mn \in \mathbb{N}\, m+n \in \mathbb{N}$
4. $\bigwedge mn \in \mathbb{N}\, m\cdot n \in \mathbb{N}$

En estas condiciones, llamaremos *expresiones aritméticas* de $\mathcal{L}$ a las determinadas por las condiciones siguientes:

1. 0 y las variables son términos aritméticos.
2. Si t_1 y t_2 son términos aritméticos, también lo son t_1', t_1+t_2 y $t_1\cdot t_2$.
3. Si t_1 y t_2 son términos aritméticos, $t_1 = t_2$ es una fórmula aritmética.
4. Si α y β son fórmulas aritméticas, también lo son $\neg\alpha$, $\alpha\to\beta$ y $\bigwedge x \in \mathbb{N}\,\alpha$.
5. Si α es una fórmula aritmética, $x \in \mathbb{N}|\alpha$ es un término aritmético,

donde hemos usado el primero y el tercero de los convenios siguientes:

$$\bigwedge x \in \mathbb{N}\,\alpha \equiv \bigwedge x(x\in\mathbb{N}\to\alpha), \quad \bigvee x \in \mathbb{N}\,\alpha \equiv \bigvee x(x\in\mathbb{N}\wedge\alpha),$$

$$(x \in \mathbb{N}|\alpha) \equiv x|(x\in\mathbb{N}\wedge\alpha).$$

Del apartado 4) se desprende que también son aritméticas $\alpha\vee\beta$, $\alpha\wedge\beta$ y $\alpha\leftrightarrow\beta$. En la práctica consideraremos como fórmulas aritméticas (en sentido amplio) todas las fórmulas equivalentes (en la teoría) a una fórmula aritmética (en sentido estricto), y ello incluye a las de la forma $\bigvee x \in \mathbb{N}\,\alpha$, donde α es aritmética.

Más aún, podemos definir la *traducción* $\bar\theta$ de $\mathcal{L}$ de una expresión θ de $\mathcal{L}_a$ mediante las reglas siguientes:

1. $\bar{x}_i \equiv x_i$.
2. $\bar{0} \equiv 0$.
3. $\overline{t'} \equiv (\bar{t})'$.
4. $\overline{t_1+t_2} \equiv \bar{t}_1+\bar{t}_2$.
5. $\overline{t_2\cdot t_2} \equiv \bar{t}_1\cdot\bar{t}_2$.

[12]Para tratar más cómodamente con las descripciones impropias convendremos en que en una teoría aritmética se puede demostrar también que $x|(x=x) = 0$.

6. $\overline{t_1 = t_2} \equiv \bar{t}_2 = \bar{t}_2$.
7. $\overline{\neg\alpha} \equiv \neg\bar{\alpha}$.
8. $\overline{\alpha \to \beta} \equiv \bar{\alpha} \to \bar{\beta}$.
9. $\overline{\bigwedge x_i\, \alpha} \equiv \bigwedge x_i \in \mathbb{N}\, \bar{\alpha}$.
10. $\overline{x_i|\alpha} \equiv x_i \in \mathbb{N}|\bar{\alpha}$.

Aquí hay que entender que los signos que hay bajo la barra a la izquierda son los de $\mathcal{L}_a$, mientras que los de la derecha son los términos de $\mathcal{L}$ dados por la definición de teoría aritmética.

Es inmediato que $\overline{\alpha \vee \beta} \equiv \bar{\alpha} \vee \bar{\beta}$, $\overline{\alpha \wedge \beta} \equiv \bar{\alpha} \wedge \bar{\beta}$, $\overline{\alpha \leftrightarrow \beta} \equiv \bar{\alpha} \leftrightarrow \bar{\beta}$, mientras que $\overline{\bigvee x\alpha}$ es lógicamente equivalente a $\bigvee \bar{x} \in \mathbb{N}\, \bar{\alpha}$. También es fácil ver que $\overline{\bigvee\limits^{1} x\alpha}$ equivale a $\bigvee\limits^{1} \bar{x} \in \mathbb{N}\, \bar{\alpha}$.

Es claro entonces que las fórmulas aritméticas de $\mathcal{L}$ son precisamente las (equivalentes a) traducciones de fórmulas de $\mathcal{L}_a$.

Diremos que una teoría axiomática T es una *teoría aritmética* si interpreta a $\mathcal{L}_a$ y en ella pueden demostrarse las traducciones de los axiomas de Peano (tomando clausuras universales en el principio de inducción):

$(\overline{\text{AP1}})$ $\bigwedge x \in \mathbb{N}\; x' \neq 0$
$(\overline{\text{AP2}})$ $\bigwedge xy \in \mathbb{N}(x' = y' \to x = y)$
$(\overline{\text{AP3}})$ $\bigwedge x \in \mathbb{N}\; x + 0 = x$
$(\overline{\text{AP4}})$ $\bigwedge xy \in \mathbb{N}(x + y' = (x + y)')$
$(\overline{\text{AP5}})$ $\bigwedge x \in \mathbb{N}\; x \cdot 0 = 0$
$(\overline{\text{AP6}})$ $\bigwedge xy \in \mathbb{N}(xy' = xy + x)$
$(\overline{\text{AP7}})$ $\bigwedge x_1 \cdots x_n \in \mathbb{N}(\phi(0) \wedge \bigwedge x \in \mathbb{N}(\phi(x) \to \phi(x')) \to \bigwedge x \in \mathbb{N}\, \phi(x))$,

donde ϕ es cualquier fórmula aritmética con variables libres $x, x_1, \ldots, x_n$.[13]

Nota Las traducciones de los axiomas de Peano ya implican que las sentencias 2, 3 y 4 de la definición 3.28 son teoremas de T, pues se pueden demostrar por inducción respecto de las fórmulas aritméticas

$$\bigvee z \in \mathbb{N}\, z = x, \qquad \bigvee z \in \mathbb{Z}\, z = x + y, \qquad \bigvee z \in \mathbb{N}\, z = x \cdot y.$$

■

Por ejemplo, AP es una teoría aritmética con $x \in \mathbb{N} \equiv x = x$ y las expresiones obvias (de modo que todo es un número natural y la traducción de una expresión de $\mathcal{L}_a$ es ella misma), mientras que 3.27 prueba que Z^* es una teoría aritmética con $x \in \mathbb{N} \equiv x \in \omega$ y las definiciones de $x', x + y, x \cdot y$ que hemos dado.

[13] En principio, $(\overline{\text{AP7}})$ se tiene que probar para fórmulas aritméticas en sentido estricto, pero es claro que si una fórmula ϕ es equivalente a una fórmula aritmética en sentido estricto, entonces el principio de inducción para dicha fórmula equivalente es lógicamente equivalente al principio de inducción para ϕ, por lo que también es un teorema.

Vamos a probar que en una teoría aritmética se pueden demostrar las traducciones de todos los teoremas de AP. Para ello necesitamos un resultado técnico. Notemos que θ y $\bar{\theta}$ tienen "las mismas" variables libres, donde las comillas indican que esto supone identificar cada variable x_i de $\mathcal{L}_a$ con la variable x_i de $\mathcal{L}$, aunque en principio puedan ser signos distintos.

Teorema 3.29 *Sea T una teoría axiomática sobre un lenguaje formal $\mathcal{L}$ que interprete a $\mathcal{L}_a$.*

1. *Si $t(x_1,\dots,x_n)$ es un término de $\mathcal{L}_a$ cuyas variables libres están entre las indicadas, entonces $\underset{T}{\vdash}\bigwedge x_1\cdots x_n\in\mathbb{N}\,\bar{t}\in\mathbb{N}$.*

2. *Si θ, x, t son, respectivamente una expresión, una variable y un término de $\mathcal{L}_a$ cuyas variables libres estén entre $x_1,\dots,x_n$, entonces, si θ es un término,*

$$\underset{T}{\vdash}\bigwedge x_1\cdots x_n\in\mathbb{N}\ \overline{\mathtt{S}^t_x\theta}=\mathtt{S}^{\bar{t}}_x\bar{\theta}$$

y si θ es una fórmula

$$\underset{T}{\vdash}\bigwedge x_1\cdots x_n\in\mathbb{N}\ (\overline{\mathtt{S}^t_x\theta}\leftrightarrow\mathtt{S}^{\bar{t}}_x\bar{\theta}).$$

DEMOSTRACIÓN: 1) Por inducción sobre la longitud de t. Si $t\equiv x_i$, entonces trivialmente

$$\underset{T}{\vdash}\bigwedge x_1\cdots x_n\in\mathbb{N}\,x_i\in\mathbb{N}.$$

Si $t\equiv 0$, por definición de teoría aritmética $\underset{T}{\vdash}0\in\mathbb{N}$.

Si $t\equiv t'_0$, por hipótesis de inducción $\underset{T}{\vdash}\bigwedge x_1\cdots x_n\in\mathbb{N}\,t_0\in\mathbb{N}$ y, por definición de teoría aritmética, de aquí se sigue que $\underset{T}{\vdash}\bigwedge x_1\cdots x_n\in\mathbb{N}\,t'_0\in\mathbb{N}$.

Igualmente se razona si $t\equiv t_1+t_2$ o $t\equiv t_1\cdot t_2$.

Si $t\equiv x|\alpha$, hay que probar que $\underset{T}{\vdash}\bigwedge x_1\cdots x_n\ (x\in\mathbb{N}|\bar{\alpha})\in\mathbb{N}$, para lo cual razonamos así: fijados $x_1,\dots,x_n\in\mathbb{N}$, si $\overset{1}{\bigvee}x(x\in\mathbb{N}\wedge\bar{\alpha})$, entonces por DP se cumple que $(x\in\mathbb{N}|\bar{\alpha})\in\mathbb{N}$. En caso contrario, por DI

$$(x\in\mathbb{N}|\bar{\alpha})=x|(x=x)=0\in\mathbb{N},$$

luego en ambos casos tenemos la conclusión.

2) Por inducción sobre la longitud de θ.

Si $\theta\equiv x$, entonces $\mathtt{S}^t_x\theta\equiv t$, y trivialmente $\bar{t}=\mathtt{S}^{\bar{t}}_x x$.

Si $\theta\equiv x_i\not\equiv x$, entonces $\mathtt{S}^t_x\theta\equiv x_i$, y lo que hay que probar es que $x_i=x_i$.

Si $\theta\equiv t'_0$, entonces $\mathtt{S}^t_x\theta\equiv(\mathtt{S}^t_x t_0)'$, luego $\overline{(\mathtt{S}^t_x t_0)'}\equiv(\overline{\mathtt{S}^t_x\bar{t}_0})'$, mientras que $\mathtt{S}^{\bar{t}}_x\bar{t}'_0\equiv\mathtt{S}^{\bar{t}}_x(\bar{t}_0)'\equiv(\mathtt{S}^{\bar{t}}_x t_0)'$ y hay que probar que

$$\bigwedge x_1\cdots x_n\in\mathbb{N}\ (\overline{\mathtt{S}^t_x t_0})'=(\mathtt{S}^{\bar{t}}_x t_0)',$$

pero, por hipótesis de inducción, tenemos que $\bigwedge x_1\cdots x_n\in\mathbb{N}\ \overline{\mathtt{S}^t_x t_0}=\mathtt{S}^{\bar{t}}_x\bar{t}_0$, y la conclusión es inmediata.

Si $\theta \equiv t_1 + t_2$ o $\theta \equiv t_1 \cdot t_2$ se razona análogamente.

Si $\theta \equiv t_1 = t_2$, fijados $x_1, \ldots, x_n \in \mathbb{N}$, por hipótesis de inducción tenemos que $\overline{\mathrm{S}^t_x t_i} = \mathrm{S}^{\bar{t}}_x \bar{t}_i$ y, por otra parte,

$$\overline{\mathrm{S}^t_x(t_1 = t_2)} \equiv \overline{\mathrm{S}^t_x t_1 = S^t_x t_2)} \equiv \overline{\mathrm{S}^t_x t_1} = \overline{\mathrm{S}^t_x t_2},$$

$$S^{\bar{t}}_x\ \overline{t_1 = t_2} \equiv \mathrm{S}^{\bar{t}}_x \bar{t}_1 = \mathrm{S}^{\bar{t}}_x \bar{t}_2,$$

de donde se sigue inmediatamente que $\overline{\mathrm{S}^t_x(t_1 = t_2)} \leftrightarrow S^{\bar{t}}_x\ \overline{t_1 = t_2}$.

Si $\theta \equiv \neg\alpha$, entonces, fijados $x_1, \ldots, x_n \in \mathbb{N}$ y usando la hipótesis de inducción,

$$\overline{\mathrm{S}^t_x \neg\alpha} \equiv \neg\overline{\mathrm{S}^t_x \alpha} \leftrightarrow \neg\mathrm{S}^{\bar{t}}_x \bar{\alpha} \equiv \mathrm{S}^{\bar{t}}_x \neg\bar{\alpha}.$$

Si $\theta \equiv \alpha \to \beta$ el razonamiento es análogo.

Si $\theta \equiv \bigwedge u\alpha$ y x no está libre en θ, luego tampoco en $\bar{\theta}$, tenemos

$$\overline{\mathrm{S}^t_x \bigwedge u\alpha} \equiv \overline{\bigwedge u\alpha} \equiv \mathrm{S}^{\bar{t}}_x \overline{\bigwedge u\alpha}.$$

Si x está libre en θ y u no está libre en t (luego tampoco en $\bar{t}$), supuesto que $x_1, \ldots, x_n \in \mathbb{N}$

$$\overline{\mathrm{S}^t_x \bigwedge u\alpha} \equiv \overline{\bigwedge u\, \mathrm{S}^t_x \alpha} \equiv \bigwedge u \in \mathbb{N}\, \overline{\mathrm{S}^t_x \alpha}.$$

Por hipótesis de inducción, $x_1, \ldots, x_n, u \in \mathbb{N}$ implica que $\overline{\mathrm{S}^t_x \alpha} \leftrightarrow \mathrm{S}^{\bar{t}}_x \bar{\alpha}$, luego, usando que x no está libre en $u \in \mathbb{N}$,

$$\overline{\mathrm{S}^t_x \bigwedge u\alpha} \leftrightarrow \bigwedge u(u \in \mathbb{N} \to \mathrm{S}^{\bar{t}}_x \bar{\alpha}) \equiv \bigwedge u \mathrm{S}^{\bar{t}}_x (u \in \mathbb{N} \to \bar{\alpha}) \leftrightarrow \mathrm{S}^{\bar{t}}_x (\bigwedge u \in \mathbb{N}\ \bar{\alpha}) \equiv \mathrm{S}^t_x \overline{\bigwedge u\alpha}.$$

Si x está libre en θ, pero u está libre en t, tomamos la menor variable v no libre en θ ni en t, de modo que, supuesto que $x_1, \ldots, x_n \in \mathbb{N}$,

$$\overline{\mathrm{S}^t_x \bigwedge u\alpha} \equiv \overline{\bigwedge v \mathrm{S}^t_x \mathrm{S}^v_u \alpha} \equiv \bigwedge v \in \mathbb{N}\, \overline{\mathrm{S}^t_x \mathrm{S}^v_u \alpha}.$$

Por la hipótesis de inducción aplicada primero a $\mathrm{S}^v_u \alpha$, tenemos que

$$\overline{\mathrm{S}^t_x \bigwedge u\alpha} \leftrightarrow \bigwedge v \in \mathbb{N}\, \mathrm{S}^{\bar{t}}_x \overline{\mathrm{S}^v_u \alpha},$$

pero, por la hipótesis de inducción aplicada a α, tenemos que

$$\bigwedge x_1 \cdots x_n, x, v \in \mathbb{N}\, (\overline{\mathrm{S}^v_u \alpha} \leftrightarrow \mathrm{S}^v_u \bar{\alpha})$$

y, eliminando el $\bigwedge x$ (usando que $\bar{t} \in \mathbb{N}$), queda que

$$\mathrm{S}^{\bar{t}}_x \overline{\mathrm{S}^v_u \alpha} \leftrightarrow \mathrm{S}^{\bar{t}}_x \mathrm{S}^v_u \bar{\alpha},$$

luego

$$\overline{\mathrm{S}^t_x \bigwedge u\alpha} \leftrightarrow \bigwedge v(v \in \mathbb{N} \to \mathrm{S}^{\bar{t}}_x \mathrm{S}^v_u \bar{\alpha}) \leftrightarrow \mathrm{S}^{\bar{t}}_x \bigwedge u(u \in \mathbb{N} \to \bar{\alpha}) \equiv \mathrm{S}^{\bar{t}}_x \overline{\bigwedge u\alpha}.$$

Si $\theta \equiv u|\alpha$ y x no está libre en θ (luego tampoco en $\bar\theta$), entonces

$$\overline{\mathtt{S}^t_x(u|\alpha)} \equiv \overline{u|\alpha} \equiv \mathtt{S}^{\bar t}_x\ \overline{u|\alpha}.$$

Si x está libre en θ y u no está libre en t (luego tampoco en $\bar t$), supuesto que $x_1, \ldots, x_n \in \mathbb{N}$,

$$\overline{\mathtt{S}^t_x(u|\alpha)} \equiv \overline{u|\mathtt{S}^t_x\alpha} \equiv u \in \mathbb{N}|\overline{\mathtt{S}^t_x\alpha}.$$

Por hipótesis de inducción, si $x_1, \ldots, x_n, u \in \mathbb{N}$, tenemos que $\overline{\mathtt{S}^t_x\alpha} \leftrightarrow \mathtt{S}^{\bar t}_x\bar\alpha$, luego, usando que x no está en $u \in \mathbb{N}$,

$$\overline{\mathtt{S}^t_x(u|\alpha)} = u \in \mathbb{N}|\mathtt{S}^{\bar t}_x\bar\alpha \equiv \mathtt{S}^{\bar t}_x(u \in \mathbb{N}|\bar\alpha) \equiv \mathtt{S}^{\bar t}_x\overline{u|\alpha}.$$

Si x está libre en θ, pero u está libre en t, tomamos la menor variable v no libre en θ ni en t, de modo que, supuesto que $x_1, \ldots, x_n \in \mathbb{N}$,

$$\overline{\mathtt{S}^t_x u|\alpha} \equiv \overline{v|\mathtt{S}^t_x\mathtt{S}^v_u\alpha} \equiv v \in \mathbb{N}|\overline{\mathtt{S}^t_x\mathtt{S}^v_u\alpha}.$$

Aplicando la hipótesis de inducción como en el caso anterior,

$$\overline{\mathtt{S}^t_x u|\alpha} = v \in \mathbb{N}|\mathtt{S}^{\bar t}_x\overline{S^v_u\alpha} \equiv \mathtt{S}^{\bar t}_x(v \in \mathbb{N}|\mathtt{S}^v_u\bar\alpha)) =$$
$$\mathtt{S}^{\bar t}_x(u \in \mathbb{N}|\bar\alpha) = \mathtt{S}^{\bar t}_x\overline{u|\alpha}.$$

■

Con esto ya estamos en condiciones de probar:

Teorema 3.30 *Sea T una teoría axiomática sobre un lenguaje $\mathcal{L}$ que interpreta a $\mathcal{L}_a$. Si unas sentencias de $\mathcal{L}_a$ cumplen $\alpha_1, \ldots, \alpha_n \vdash \alpha$, entonces*

$$\bar\alpha_1, \ldots, \bar\alpha_n \underset{T}{\vdash} \bar\alpha.$$

DEMOSTRACIÓN: Si α es una fórmula de $\mathcal{L}_a$ cuyas variables libres sean $x_1, \ldots, x_n$, llamaremos $\bar\alpha^c \equiv \bigwedge x_1 \cdots x_n \in \mathbb{N}\,\bar\alpha$.

Veamos en primer lugar que si γ es un axioma lógico de $\mathcal{L}_a$, entonces $\underset{T}{\vdash} \bar\gamma^c$.

Por simplificar la notación vamos a suponer que las variables libres en γ son x, y, aunque todos los argumentos valen sin cambio alguno cualquiera que sea el número de variables libres (y se simplifican si no hay ninguna).

Si γ es uno de los axiomas K1, K2, K3, es inmediato que $\bar\gamma$ es un axioma del mismo tipo, luego $\vdash \bar\gamma$, luego $\vdash (x \in \mathbb{N} \wedge y \in \mathbb{N} \to \bar\gamma)$, luego, introduciendo generalizadores, $\vdash \bar\gamma^c$.

Si $\gamma \equiv \bigwedge u\alpha \to \mathtt{S}^t_u\alpha$, entonces, por el teorema anterior, $\bar\gamma \equiv \bigwedge u \in \mathbb{N}\,\bar\alpha \to \mathtt{S}^{\bar t}_u\bar\alpha$, y podemos razonar como sigue (la línea 2 es válida por el teorema anterior):

1)	$x \in \mathbb{N} \wedge y \in \mathbb{N}$	Hipótesis
2)	$\bar t \in \mathbb{N}$	Consecuencia de 1)
3)	$\bigwedge u \in \mathbb{N}\,\bar\alpha$	Hipótesis
4)	$\bar t \in \mathbb{N} \to \mathtt{S}^{\bar t}_u\bar\alpha$	EG 3
5)	$\mathtt{S}^{\bar t}_u\bar\alpha$	MP 2, 4
6)	$\bigwedge u \in \mathbb{N}\,\bar\alpha \to \mathtt{S}^{\bar t}_u\bar\alpha$	
7)	$x \in \mathbb{N} \wedge y \in \mathbb{N} \to (\bigwedge u \in \mathbb{N}\,\bar\alpha \to \mathtt{S}^{\bar t}_u\bar\alpha)$	
8)	$\bigwedge xy \in \mathbb{N}(\bigwedge u \in \mathbb{N}\,\bar\alpha \to \mathtt{S}^{\bar t}_u\bar\alpha)$	IG 7

Si $\gamma \equiv \bigwedge u(\alpha \to \beta) \to (\alpha \to \bigwedge u\beta)$, entonces

$$\bar\gamma \equiv \bigwedge u \in \mathbb{N}\,(\bar\alpha \to \bar\beta) \to (\bar\alpha \to \bigwedge u \in \mathbb{N}\,\bar\beta).$$

Razonamos como sigue:

1)	$\bigwedge u \in \mathbb{N}(\bar\alpha \to \bar\beta)$	Hipótesis
2)	$\bar\alpha$	Hipótesis
3)	$u \in \mathbb{N}$	Hipótesis
4)	$u \in \mathbb{N} \to (\bar\alpha \to \bar\beta)$	EG 1
5)	$\bar\alpha \to \bar\beta$	MP 3, 4
6)	$\bar\beta$	MP 2, 5
7)	$u \in \mathbb{N} \to \bar\beta$	
8)	$\bigwedge u \in \mathbb{N}\ \bar\beta$	IG 7
9)	$\bar\alpha \to \bigwedge u \in \mathbb{N}\ \bar\beta$	
10)	$\bigwedge u \in \mathbb{N}(\bar\alpha \to \bar\beta) \to (\bar\alpha \to \bigwedge u \in \mathbb{N}\ \bar\beta)$	

De aquí podemos pasar a $\bar\gamma^c$ de forma obvia.

Si $\gamma \equiv \bigwedge u(u = t \to \alpha) \leftrightarrow \mathtt{S}^t_u\alpha$, entonces, usando el teorema anterior, $\bar\gamma$ equivale a

$$\bigwedge u \in \mathbb{N}(u = \bar t \to \bar\alpha) \leftrightarrow \mathtt{S}^{\bar t}_u\bar\alpha.$$

Razonamos como sigue:

1)	$x \in \mathbb{N} \wedge y \in \mathbb{N}$	Hipótesis
2)	$\bar t \in \mathbb{N}$	Consecuencia de 1)
3)	$\bigwedge u \in \mathbb{N}(u = \bar t \to \bar\alpha)$	Hipótesis
4)	$\bar t \in \mathbb{N} \to (\bar t = \bar t \to \mathtt{S}^{\bar t}_u\bar\alpha)$	EG 3
5)	$\mathtt{S}^{\bar t}_u\bar\alpha$	MP 2, 4, I
6)	$\bigwedge u \in \mathbb{N}(u = \bar t \to \bar\alpha) \to \mathtt{S}^{\bar t}_u\bar\alpha$	
7)	$\mathtt{S}^{\bar t}_u\bar\alpha$	Hipótesis
8)	$\bigwedge u(u = \bar t \to \bar\alpha)$	II 7
9)	$u = \bar t \to \bar\alpha$	EG 8
10)	$u \in \mathbb{N} \to (u = \bar t \to \bar\alpha)$	Consecuencia de 9)
11)	$\bigwedge u \in \mathbb{N}(u = \bar t \to \bar\alpha)$	IG 10
12)	$\mathtt{S}^{\bar t}_u\bar\alpha \to \bigwedge u \in \mathbb{N}(u = \bar t \to \bar\alpha)$	
13)	$\bigwedge u \in \mathbb{N}(u = \bar t \to \bar\alpha) \leftrightarrow \mathtt{S}^{\bar t}_u\bar\alpha$	IB 6, 12
14)	$x \in \mathbb{N} \wedge y \in \mathbb{N} \to \bigwedge u \in \mathbb{N}\,(u = \bar t \to \bar\alpha) \leftrightarrow \mathtt{S}^{\bar t}_u\bar\alpha$	
15)	$\bigwedge xy \in \mathbb{N}(\bigwedge u \in \mathbb{N}(u = \bar t \to \bar\alpha) \leftrightarrow \mathtt{S}^{\bar t}_u\bar\alpha)$	IG 14

Si $\gamma \equiv \overset{1}{\bigvee} u\alpha \to \mathtt{S}^{u|\alpha}_u\alpha$, entonces es fácil ver que (bajo las hipótesis $x, y \in \mathbb{N}$) la fórmula $\bar\gamma$ es equivalente a

$$\overset{1}{\bigvee} u(u \in \mathbb{N} \wedge \bar\alpha) \to \mathtt{S}^{u\in\mathbb{N}|\bar\alpha}_u\bar\alpha,$$

y esto es un axioma lógico.

Si $\gamma \equiv \neg\overset{1}{\bigvee} u\alpha \to (u|\alpha) = x|(x = x)$, entonces $\bar\gamma$ es equivalente a

$$\neg\overset{1}{\bigvee} u(u \in \mathbb{N} \wedge \bar\alpha) \to (u \in \mathbb{N}|\bar\alpha) = x \in \mathbb{N}|(x = x).$$

Suponemos $\neg\overset{1}{\bigvee} u(u \in \mathbb{N} \wedge \bar\alpha)$ y distinguimos dos casos: si $\neg\overset{1}{\bigvee} x\ x \in \mathbb{N}$, las dos descripciones son impropias, por lo que ambas son iguales a $x|(x = x)$.

Si, por el contrario, $\overset{1}{\bigvee} x\ x \in \mathbb{N}$, entonces,[14] como $\underset{T}{\vdash} 0 \in \mathbb{N}$, resulta que 0 es el único elemento de $\mathbb{N}$, y por el teorema anterior tenemos también que $\underset{T}{\vdash}(u \in \mathbb{N}|\bar\alpha) \in \mathbb{N}$, luego concluimos que

$$(u \in \mathbb{N}|\bar\alpha) = 0 = x \in \mathbb{N}|(x = x).$$

Supongamos ahora que tenemos una deducción lógica $\gamma_1, \ldots, \gamma_r$ de $\alpha \equiv \gamma_r$ con premisas $\alpha_1, \ldots, \alpha_n$ y vamos a probar que, para todo i,

$$\bar\alpha_1^c, \ldots, \bar\alpha_n^c \underset{T}{\vdash} \bar\gamma_i^c.$$

En el caso particular en que todas las α_i son sentencias, tenemos lo que afirma el teorema.

Razonamos por inducción sobre i, es decir, suponemos que la conclusión es cierta para $j < i$. Ya hemos probado que también es cierta si γ_i es un axioma lógico, y lo es trivialmente si es una premisa.

Si $\gamma_i \equiv \beta$ se deduce por MP de $\gamma_j \equiv \alpha$ y $\gamma_k \equiv \alpha \to \beta$, supongamos, por concretar, que $\alpha \equiv \alpha(x, y, z)$ y $\beta \equiv \beta(x, w)$ (es decir, que las variables libres son las indicadas) y razonamos así:

1)	$\bigwedge xyzw \in \mathbb{N}\ (\bar\alpha \to \bar\beta)$	Premisa
2)	$\bigwedge xyz \in \mathbb{N}\, \bar\alpha$	Premisa
3)	$x \in \mathbb{N} \wedge 0 \in \mathbb{N} \wedge 0 \in \mathbb{N} \wedge w \in \mathbb{N} \to (\bar\alpha(x, 0, 0) \to \bar\beta)$	EG 1
4)	$x \in \mathbb{N} \wedge 0 \in \mathbb{N} \wedge 0 \in \mathbb{N} \to \bar\alpha(x, 0, 0)$	EG 2
5)	$x \in \mathbb{N} \wedge w \in \mathbb{N}$	Hipótesis
6)	$0 \in \mathbb{N}$	Teorema
7)	$x \in \mathbb{N} \wedge 0 \in \mathbb{N} \wedge 0 \in \mathbb{N} \wedge w \in \mathbb{N}$	IC 5, 6
8)	$\bar\alpha(x, 0, 0) \to \bar\beta(x, w)$	MP 3, 7
9)	$\bar\alpha(x, 0, 0)$	MP 4, 7
10)	$\bar\beta$	MP 8, 9
11)	$x \in \mathbb{N} \wedge w \in \mathbb{N} \to \bar\beta$	
12)	$\bigwedge xw \in \mathbb{N}\, \bar\beta(x, w)$	IG 11

Si α_i se deduce por generalización, tenemos una fórmula α previa tal que $\alpha_i \equiv \bigwedge x\alpha$ y, por hipótesis de inducción, en T se demuestra $\bar\alpha^c$. Si las variables libres de α_i son $x, x_1, \ldots, x_n$, es claro que $\bar\alpha^c$ es equivalente a

$$\bigwedge u_1 \cdots u_n x \in \mathbb{N}\, \bar\alpha,$$

que a su vez equivale a $(\bigwedge x \in \mathbb{N}\, \bar\alpha)^c$. ∎

[14] Este caso no puede darse si en T se demuestran los axiomas de Peano, pero no estamos suponiendo tal cosa.

En particular, en toda teoría aritmética se demuestran las traducciones de todos los teoremas de AP que sean sentencias. Esto no es ninguna restricción, porque una fórmula es un teorema si y sólo si lo es su clausura universal. Necesitamos cuantificar las variables libres para que en la traducción queden restringidas a números naturales.

3.5 Descriptores

Esta sección correspondería en principio al capítulo anterior, pero la hemos pospuesto porque así ahora contamos ya con numerosos ejemplos prácticos del uso del descriptor. En realidad el descriptor tiene esencialmente un único uso práctico, que es el de introducir nuevos términos en una teoría sin necesidad de incorporar nuevos signos a su lenguaje. Un ejemplo típico (en realidad todos lo son) nos lo encontramos cuando en la teoría básica de conjuntos B hemos demostrado

$$\bigwedge xy\overset{1}{\bigvee} z\bigwedge u(u \in z \leftrightarrow u \in x \wedge u \in y)$$

y hemos querido dar nombre a ese único conjunto que contiene los elementos comunes de x e y. La forma en que lo hemos hecho ha sido definir

$$x \cap y \equiv z|\bigwedge u(u \in z \leftrightarrow u \in x \wedge u \in y),$$

es decir, "la intersección es el conjunto z tal que sus elementos son los elementos comunes de x e y".

Aplicamos la regla de las descripciones propias cuando afirmamos que la intersección tiene la propiedad que la define, y escribimos

$$\bigwedge u(u \in x \cap y \leftrightarrow u \in x \wedge u \in y).$$

Este paso habría sido incorrecto de no contar con la existencia y unicidad, que es la premisa de la regla DP. Así pues: *Sólo podemos afirmar que una descripción tiene la propiedad que la define si tenemos la garantía de que hay un único objeto que la posee.* Éste es el contenido de la regla DP.

Tenemos también la regla de las descripciones impropias, DI, que hemos usado, por ejemplo, en la prueba del teorema 3.29. En general, esta regla tiene poca aplicación práctica, porque lo habitual es no usar nunca una descripción sin tener garantizado que es propia, pero esto no es más que un "código de buenas prácticas", no una obligación lógica.

Así, en Z* definimos $f(x) \equiv y|(x,y) \in f$. Si, por ejemplo,

$$f = \{(0,2),(1,7)\},$$

tenemos que $f(1) = 7$, pero ¿qué es $f(2)$? En la práctica un matemático nunca se preguntará esto. Simplemente, no escribirá nunca $f(2)$ en este contexto, pues sabe que no existe ningún y tal que $(2,y) \in f$. Pero lo cierto es que $f(2)$ es un término de $\mathcal{L}_{\mathrm{tc}}$ y, aunque no queramos, podemos razonar sobre él. La regla DI

atribuye un sentido a $f(2)$ al establecer que $f(2) = x|x = x$. Por ejemplo, como lo mismo vale para $f(3)$ por el mismo motivo, podemos afirmar que, en este contexto $f(2) = f(3)$. En un modelo de Z^*, los términos $f(2)$, $f(3)$ o $x|x = x$ denotarán al objeto que hayamos designado como descripción impropia.

Como decimos, estos convenios nunca tendrán ninguna relevancia práctica, pero en los resultados teóricos como el teorema 3.29, en el que queremos demostrar algo para todos los términos de un lenguaje formal (y ello incluye tanto a los que estamos dispuestos a usar como a los que nunca usaremos "en la práctica"), necesitamos una regla de inferencia que nos permita tratar las descripciones impropias, y ésa es DI. El teorema siguiente es un ejemplo especialmente simple de esta situación:

Teorema 3.31 $\vdash \bigwedge x(\alpha \leftrightarrow \beta) \rightarrow x|\alpha = x|\beta$.

DEMOSTRACIÓN: Es evidente (o, si se prefiere, por el teorema 2.11), que la hipótesis implica $\bigvee^1 x\, \alpha \leftrightarrow \bigvee^1 x\, \beta$. Por lo tanto, tenemos dos casos: o bien ambas descripciones son propias, o ambas son impropias.

Si suponemos que son propias, tenemos que $x|\alpha$ cumple α y $x|\beta$ cumple β. Técnicamente esto es $\mathtt{S}_x^{x|\alpha}\alpha$ y $\mathtt{S}_x^{x|\beta}\beta$, pero por EG en la hipótesis tenemos también $\mathtt{S}_x^{x|\alpha}\beta$, y por la unicidad $x|\alpha = x|\beta$.

Ahora bien, aunque éste es el único caso de interés práctico, la demostración no estará acabada si no consideramos también el caso en que ambas descripciones son impropias, y en ese caso la regla de las descripciones impropias nos dice que

$$x|\alpha = (x|x = x) = x|\beta. \qquad \blacksquare$$

Cuando los términos definidos, como $x \cap y$ o $x - y$ se consideran funtores "nuevos" añadidos al lenguaje, o como "trozos de fórmula eliminables", es necesario justificar de algún modo que todas las fórmulas que contienen estos signos son equivalentes a otras fórmulas "auténticas" que no los contienen. Cuando se consideran descripciones esa necesidad teórica desaparece, pues los conceptos definidos son términos del lenguaje propiamente dichos, pero sigue existiendo la necesidad práctica de justificar que los descriptores pueden eliminarse, pues en algunos contextos es necesario suponer que se trabaja sin ellos. El capítulo siguiente será una buena muestra de contexto en el que es necesario suprimir temporalmente los descriptores de los razonamientos, y la justificación de que eso puede hacerse la vamos a dar ahora.

Antes de probar un resultado general, veamos algunos ejemplos concretos. Consideremos la fórmula

$$\{x, y\} \in z.$$

En ella aparece la descripción

$$\{x, y\} \equiv u|\bigwedge v(v \in u \leftrightarrow u = x \vee u = y),$$

pero podemos considerar una fórmula equivalente sin descriptores:

$$\bigvee u(u \in z \wedge \bigwedge v(v \in u \leftrightarrow v = x \vee v = y)).$$

En efecto, esta fórmula dice: existe un $u \in z$ cuyos elementos son x, y, que, ciertamente, es otra forma de decir que $\{x, y\} \in z$. Esta forma de eliminar descriptores se puede acumular. Por ejemplo, si recordamos que $(x, y) = \{\{x\}, \{x, y\}\}$, una fórmula sin descriptores equivalente a $(x, y) \in z$ es

$$\bigvee u(u \in z \wedge \bigwedge v(v \in u \leftrightarrow v = \{x\} \vee v = \{x, y\})),$$

que aún tiene descriptores (aunque menos), pero se pueden eliminar en una segunda fase:

$$\bigvee u(u \in z \wedge \bigwedge v(v \in u \leftrightarrow \bigwedge w(w \in v \leftrightarrow w = x) \vee \bigwedge w(w \in v \leftrightarrow w = x \vee w = y))).$$

Así hemos encontrado una fórmula $\phi(x, y, z)$ sin descriptores que es equivalente (por ejemplo, en B) a $(x, y) \in z$.

Ahora bien, estos ejemplos eran sencillos porque en B se demuestra que las descripciones que hemos eliminado son siempre propias, cosa que no tiene por qué cumplirse en general. Consideremos por ejemplo la fórmula $f(x) \in z$, donde

$$f(x) \equiv u|(x, u) \in f.$$

Esta descripción puede ser impropia, bien porque f no contenga ningún par de la forma (x, u), bien porque contenga varios, en cuyo caso $f(x) = u|(u = u)$. Teniendo esto en cuenta, una fórmula equivalente a $f(x) \in z$ es

$$\bigvee y(y \in z \wedge \bigwedge u((x, u) \in f \leftrightarrow u = y)) \vee (\neg \overset{1}{\bigvee} y(x, y) \in f \wedge x|(x = x) \in z).$$

Aquí dice que, o bien hay un $y \in z$ que es el único que cumple $(x, y) \in f$, o bien no hay un único y que cumpla $(x, y) \in f$ y la descripción impropia está en z. Esto equivale a $f(x) \in z$ y tiene en cuenta la posibilidad de que $f(x)$ sea una descripción impropia.

Si usamos la fórmula $\phi(x, y, z)$ sin descriptores que hemos obtenido antes, nos queda:

$$\bigvee y(y \in z \wedge \bigwedge u(\phi(x, u, f) \leftrightarrow u = y)) \vee$$
$$(\neg \overset{1}{\bigvee} y\phi(x, y, f) \wedge \bigvee w(w = x|(x = x) \wedge w \in z)).$$

Así tenemos una fórmula equivalente a $f(x) \in z$ que tiene un único descriptor, el que aparece al nombrar la descripción impropia $x|x = x$. Éste no se puede eliminar salvo que tengamos una caracterización sin descriptores de la descripción impropia. Por ejemplo, si convenimos que $x|(x = x) = \varnothing$, entonces $w = x|(x = x)$ es equivalente a $\bigwedge u\, u \notin w$, y así llegamos a una fórmula equivalente sin descriptores:

$$\bigvee y(y \in z \wedge \bigwedge u(\phi(x, u, f) \leftrightarrow u = y)) \vee$$
$$(\neg \overset{1}{\bigvee} y\phi(x, y, f) \wedge \bigvee w(\bigwedge u\, u \notin w \wedge w \in z)).$$

El teorema siguiente prueba que esto puede hacerse en general, que cualquier fórmula es equivalente a una fórmula sin descriptores siempre que la descripción impropia admita una caracterización sin descriptores y, más aún, todo teorema sin descriptores puede demostrarse sin descriptores:

Teorema 3.32 (Eliminación de descriptores) *Sea T una teoría axiomática sobre un lenguaje formal $\mathcal{L}$ sin descriptor, y llamemos $\mathcal{L}'$ al lenguaje formal que resulta de añadirle a $\mathcal{L}$ un descriptor. Supongamos que existe una fórmula $\phi(x)$ de $\mathcal{L}$, con x como única variable libre, tal que*

$$\vdash_T \bigvee\nolimits^1 u\, \phi(u).$$

Llamamos T' a la teoría axiomática sobre $\mathcal{L}'$ cuyos axiomas propios son los de T más el axioma $\phi(x|x = x)$. Entonces:

1. *Toda fórmula de $\mathcal{L}'$ es equivalente en T' a una fórmula de $\mathcal{L}$ con las mismas variables libres.*

2. *Si una fórmula de $\mathcal{L}$ es demostrable en T', también lo es en T.*

Si el lenguaje $\mathcal{L}$ tiene una constante c, siempre podemos tomar $\phi(x) \equiv x = c$. Así, el axioma $x|(x = x) = c$ establece que todas las descripciones impropias, todo lo que esté mal definido, será c, por definición.

Por ejemplo, si llamamos T a la teoría axiomática sobre el lenguaje $\mathcal{L}_a^*$ que resulta de eliminar el descriptor en $\mathcal{L}_a$ y cuyos axiomas son los axiomas de Peano salvo que el principio de inducción se restringe a fórmulas sin descriptores, la teoría T' sobre $\mathcal{L}_a$ cuyos axiomas son los de T mas la sentencia $x|(x = x) = 0$ es equivalente a AP, en el sentido de que sus teoremas son los mismos que los de AP.

En efecto, por el teorema anterior, toda fórmula de $\mathcal{L}_a$ es equivalente en T' a una fórmula sin descriptores, y es claro que el principio de inducción para una fórmula cualquiera es equivalente en T' al principio de inducción para una fórmula equivalente sin descriptores, que es un axioma de T', luego todos los casos del principio de inducción (con o sin descriptores) son teoremas de T', luego todos los teoremas de AP son teoremas de T'.

En otras palabras, T es una teoría axiomática sin descriptor que equivale a AP en cuanto le añadimos el descriptor y el axioma $x|(x = x) = 0$. Por ello, podemos referirnos a T como la aritmética de Peano sin descriptor. El axioma $x|(x = x) = 0$ lo consideraremos como un axioma "semilógico", es decir, que no lo incluiremos entre los axiomas propios de AP, sino que lo consideraremos como parte de los axiomas lógicos, aunque no lo es en sentido estricto (pues no es una fórmula lógicamente válida). Tenemos que todo teorema de AP sin descriptores puede probarse en T.

Otro ejemplo típico de aplicación del teorema anterior es cualquier teoría de conjuntos dotada al menos del axioma de extensionalidad y que permita probar la existencia del conjunto vacío. En tal caso podemos aplicar el teorema anterior considerando el axioma $\bigwedge u u \notin x|(x = x)$, que equivale a $x|(x = x) = \varnothing$.

La situación es la misma que acabamos de analizar para la Aritmética de Peano. Por ejemplo, para Z^* podemos considerar la teoría T sobre el lenguaje $\mathcal{L}_{\mathrm{tc}}$ sin descriptor cuyos axiomas son los de Z^* salvo que el esquema de especificación lo restringimos a fórmulas sin descriptores, y nos encontramos con que

la teoría T' sobre $\mathcal{L}_{\mathrm{tc}}$ (con descriptor) que resulta de añadirle a T el axioma $\bigwedge uu \notin x|(x = x)$ tiene los mismos teoremas que Z^*, pues los casos del axioma de especificación correspondientes a fórmulas con descriptores son equivalentes en T' a los casos correspondientes a fórmulas equivalentes sin descriptores. Por ello, podemos referirnos a T como la teoría Z^* sin descriptor, que se convierte en Z^* en cuanto añadimos el descriptor y el axioma $\bigwedge uu \notin x|(x = x)$. Además, todo teorema de Z^* sin descriptores puede demostrarse en T.

Para demostrar 3.32, empezamos observando que, bajo sus hipótesis, por la unicidad tenemos que

$$\underset{T'}{\vdash} y = x|(x = x) \leftrightarrow \phi(y).$$

El teorema siguiente nos proporciona el medio de expresar una descripción sin descriptores:

Teorema 3.33 *Si la variable y no está en $u|\alpha$, se cumple*

$$\vdash y = u|\alpha \leftrightarrow \bigwedge u(\alpha \leftrightarrow u = y) \vee (\neg \overset{1}{\bigvee} u\, \alpha \wedge y = x|(x = x)).$$

Demostración: Esbozamos la prueba. Bajo la hipótesis $y = u|\alpha$, distinguimos dos casos, o bien $\overset{1}{\bigvee} u\ \alpha$ o bien $\neg \overset{1}{\bigvee} u\ \alpha$.

En el primer caso, por la definición de unicidad, tenemos $\bigvee v \bigwedge u(\alpha \leftrightarrow u = v)$. Eliminamos el particularizador, con lo que $\bigwedge u(\alpha \leftrightarrow u = z)$. Eliminando el generalizador llegamos a $S_u^{u|\alpha}\alpha \leftrightarrow (u|\alpha) = z$, pero la parte izquierda la tenemos por la regla de las descripciones propias, con lo que $z = u|\alpha$. Por hipótesis, $y = z$ y por la equivalencia de términos idénticos $\bigwedge u(\alpha \leftrightarrow u = y)$.

En el segundo caso, la regla de las descripciones impropias nos permite afirmar que $u|\alpha = x|x = x$ y por hipótesis $y = u|\alpha$, lo que nos lleva a que $\neg \overset{1}{\bigvee} u\alpha \wedge y = x|(x = x)$.

Supongamos ahora la parte derecha del teorema. Por la regla del dilema basta probar que ambas disyuntivas nos llevan a $y = u|\alpha$.

Si suponemos $(*)$: $\bigwedge u(\alpha \leftrightarrow u = y)$, introduciendo un particularizador obtenemos $\overset{1}{\bigvee} u\, \alpha$, luego la regla de las descripciones propias nos da $\mathtt{S}_u^{u|\alpha}\alpha$. Por otro lado, eliminando el generalizador en $(*)$ obtenemos $\mathtt{S}_u^{u|\alpha}\alpha \leftrightarrow u|\alpha = y$, luego concluimos que $y = u|\alpha$.

Si suponemos $\neg \overset{1}{\bigvee} u\, \alpha \wedge y = x|(x = x)$, la regla de las descripciones impropias nos da que $u|\alpha = x|(x = x)$, luego concluimos igualmente que $y = u|\alpha$. ∎

Combinando esto con la observación previa al teorema, vemos que

$$\underset{T'}{\vdash} y = u|\alpha \leftrightarrow \bigwedge u(\alpha \leftrightarrow u = y) \vee (\neg \overset{1}{\bigvee} u\, \alpha \wedge \phi(y)).$$

A continuación definimos, para cada término t de $\mathcal{L}'$, una fórmula $\psi_t(y)$ de $\mathcal{L}$ que tenga las mismas variables libres que t más una variable libre adicional y y, para cada fórmula α de $\mathcal{L}'$, una fórmula α^* de $\mathcal{L}$ que tenga sus mismas variables libres (de modo que, si partimos de una expresión, obtenemos una fórmula).

La definición es la siguiente:

1. Si $t \equiv x_i$, entonces $\psi_t(y) \equiv y = x_i$.
2. Si $t \equiv c_i$, entonces $\psi_t(y) \equiv y = c_i$.
3. Si $t \equiv f_i^n t_1 \cdots t_n$, entonces
$$\psi_t(y) \equiv \bigvee u_1 \cdots u_n(\psi_{t_1}(u_1) \wedge \cdots \wedge \psi_{t_n}(u_n) \wedge y = f_i^n u_1 \cdots u_n).$$
4. Si $\alpha \equiv R_i^n t_1 \cdots t_n$, entonces
$$\alpha^* \equiv \bigvee u_1 \cdots u_n(\psi_{t_1}(u_1) \wedge \cdots \wedge \psi_{t_n}(u_n) \wedge R_i^n u_1 \ldots u_n).$$
5. Si $\alpha \equiv \neg\beta$, entonces $\alpha^* \equiv \neg\beta^*$.
6. Si $\alpha \equiv \beta \to \gamma$, entonces $\alpha^* \equiv \beta^* \to \gamma^*$.
7. Si $\alpha \equiv \bigwedge u\, \beta$, entonces $\alpha^* \equiv \bigwedge u\, \beta^*$.
8. Si $t \equiv u|\alpha$, entonces
$$\psi_t(y) \equiv \bigwedge u(\alpha^* \leftrightarrow u = y) \vee (\neg \overset{1}{\bigvee} u\, \alpha^* \wedge \phi(y)).$$

Necesitamos probar varios hechos sobre estas fórmulas:

1. Si y es una variable que no está en el término t de $\mathcal{L}'$ y α es una fórmula de $\mathcal{L}'$,
$$\underset{T'}{\vdash} \psi_t(y) \leftrightarrow y = t, \qquad \underset{T'}{\vdash} \alpha^* \leftrightarrow \alpha.$$
Y si t o α no tienen descriptores, la equivalencia correspondiente puede probarse en T.

   En efecto, razonando por inducción sobre la longitud de una expresión, todos los casos son inmediatos salvo el correspondiente a las descripciones $t \equiv u|\alpha$, donde usamos
$$\underset{T'}{\vdash} y = u|\alpha \leftrightarrow \bigwedge u(\alpha \leftrightarrow u = y) \vee (\neg \overset{1}{\bigvee} u\, \alpha \wedge \phi(y)), \quad \underset{T'}{\vdash} \alpha^*(y) \leftrightarrow \alpha(y).$$

2. $\underset{T}{\vdash} \overset{1}{\bigvee} y\, \psi_t(y)$, donde la variable y no está en t.

   Nuevamente, todos los casos son inmediatos salvo el correspondiente a $t \equiv u|\alpha$. Distinguimos dos casos: o bien $\overset{1}{\bigvee} u\, \alpha^*(u)$, o bien $\neg \overset{1}{\bigvee} u\, \alpha^*(u)$. En el primer caso $\psi_t(y)$ sólo lo cumple el único y que cumple $\alpha^*(y)$ y en el segundo caso sólo lo cumple el único y que cumple $\phi(y)$.

3. Si t, t' son términos de $\mathcal{L}'$, α es una fórmula e $y \notequiv x$ es una variable que no esté en $\mathtt{S}^t_x t'$ ni en $\mathtt{S}^t_x \alpha$, entonces

$$\underset{T}{\vdash} \psi_{\mathtt{S}^t_x t'}(y) \leftrightarrow \bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x \psi_{t'}(y)), \quad \underset{T}{\vdash} (S^t_x \alpha)^* \leftrightarrow \bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x \alpha^*).$$

Probamos por inducción sobre la longitud de una expresión θ que cumple lo requerido para t' o α según si es un término o una fórmula.

Si $\theta \equiv z$ es una variable, distinguimos dos casos, según si $x \equiv z$ o $x \notequiv z$. En el primer caso hay que probar (en T) que

$$\psi_t(y) \leftrightarrow \bigvee u(\psi_t(u) \wedge y = u),$$

lo cual es obvio.

En el segundo caso hay que probar

$$y = z \leftrightarrow \bigvee u(\psi_t(u) \wedge y = z),$$

y esto es también un teorema porque, según el apartado precedente, en T podemos probar que $\bigvee u \psi_t(u)$.

El caso en que $\theta \equiv c$ es una constante es idéntico a la segunda parte del caso anterior, cambiando z por c.

Si $\theta \equiv f^n_i t_1 \cdots t_n$, llamando $t'_i \equiv \mathtt{S}^t_x t_i$, tenemos que

$$\psi_{\mathtt{S}^t_x \theta}(y) \equiv \bigvee u_1 \cdots u_n (\psi_{t'_1}(u_1) \wedge \cdots \wedge \psi_{t'_n}(u_n) \wedge y = f^n_i u_1 \cdots u_n).$$

Por hipótesis de inducción,

$$\psi_{t'_i}(u_i) \leftrightarrow \bigvee v_i (\psi_t(v_i) \wedge \mathtt{S}^{v_i}_x \psi_{t_i}(u_i)),$$

pero en T se prueba $\overset{1}{\bigvee} v\, \psi_t(v)$, luego todos los v_i tienen que ser iguales, y $\psi_{\mathtt{S}^t_x \theta}(y)$ equivale a

$$\bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x \bigvee u_1 \cdots u_n (\psi_{t_1}(u_1) \wedge \cdots \wedge \psi_{t_n}(u_n) \wedge y = f^n_i u_1 \ldots u_n)),$$

que es lo mismo que $\bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x \psi_\theta(y))$.

Si $\theta \equiv R^n_i t_1 \cdots t_n$, llamando $t' \equiv \mathtt{S}^t_x t_i$, como antes, tenemos que

$$\theta^* \equiv \bigvee u_1 \cdots u_n (\psi_{t'_1}(u_1) \wedge \cdots \wedge \psi_{t'_n}(u_n) \wedge R^n_i u_1 \ldots u_n)$$

y, aplicando como antes la hipótesis de inducción, esto equivale a

$$\bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x \bigvee u_1 \cdots u_n (\psi_{t_1}(u_1) \wedge \cdots \wedge \psi_{t_n}(u_n) \wedge R^n_i u_1 \ldots u_n)),$$

que es lo mismo que $\bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x \theta^*)$.

Si $\theta \equiv \neg\alpha$, por hipótesis de inducción, en T se prueba

$$(S^t_x\alpha)^* \leftrightarrow \bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x\alpha^*).$$

Teniendo en cuenta que también tenemos $\overset{1}{\bigvee} u\psi_t(u)$, es fácil concluir que

$$\neg(S^t_x\alpha)^* \leftrightarrow \bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x\neg\alpha^*).$$

El caso $\theta \equiv \alpha \to \beta$ es análogo.

Si $\theta \equiv \bigwedge v\,\alpha$, en el caso en que x no está libre en θ hay que probar que

$$\theta^* \leftrightarrow \bigvee u(\psi_t(u) \wedge \theta^*),$$

lo cual se cumple porque en T se prueba $\bigvee u\,\psi_t(u)$.

Si x está libre en θ y v no está libre en t, aplicando la hipótesis de inducción (y cambiando la variable u por otra, si hace falta, para que $u \not\equiv v$):

$$(\mathtt{S}^t_x\theta)^* \equiv \bigwedge v(\mathtt{S}^t_x\alpha^*) \leftrightarrow \bigwedge v\bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x\alpha^*),$$

y esto equivale a $\bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x\bigwedge v\alpha^*)$, que es $\bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x\theta^*)$.

Si x está libre en θ y v está libre en t, llamamos w a la variable de menor índice que no está en θ ni en t. Entonces $\mathtt{S}^t_x\theta \equiv \bigwedge w\mathtt{S}^t_x\mathtt{S}^w_v\alpha$, luego

$$(\mathtt{S}^t_x\theta)^* \equiv \bigwedge w(\mathtt{S}^t_x\mathtt{S}^w_v\alpha)^*.$$

Por la hipótesis de inducción para $\mathtt{S}^w_v\alpha$ (que tiene la misma longitud que α), tenemos que

$$(\mathtt{S}^t_x\mathtt{S}^w_v\alpha)^* \leftrightarrow \bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x(\mathtt{S}^w_v\alpha)^*).$$

A su vez, por la hipótesis de inducción para α,

$$(\mathtt{S}^w_v\alpha)^* \leftrightarrow \bigvee s(\psi_w(s) \wedge \mathtt{S}^s_v\alpha^*) \leftrightarrow \mathtt{S}^w_v\alpha^*,$$

donde la variable s la podemos tomar que no esté en θ. En total queda que

$$(\mathtt{S}^t_x\theta)^* \leftrightarrow \bigwedge w\bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x\mathtt{S}^w_v\alpha^*).$$

Teniendo en cuenta que $\overset{1}{\bigvee} u\psi_t(u)$, esto equivale a

$$(\mathtt{S}^t_x\theta)^* \leftrightarrow \bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x\bigwedge w\,\mathtt{S}^w_v\alpha^*) \leftrightarrow \bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x\bigwedge v\alpha^*)$$

$$\leftrightarrow \bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x\theta^*).$$

Si $\theta \equiv v|\alpha$, el caso en que x no está libre en θ es idéntico a la parte correspondiente del caso anterior.

Si x está libre en θ y v no está libre en t,

$$\psi_{\mathtt{S}^t_x\theta}(y) \equiv \bigwedge v((\mathtt{S}^t_x\alpha)^* \leftrightarrow v = y) \vee (\neg \overset{1}{\bigvee} v(\mathtt{S}^t_x\alpha)^* \wedge \phi(y)).$$

Aplicando la hipótesis de inducción, esto equivale a

$$\bigwedge v(\bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x\alpha^*) \leftrightarrow v = y) \vee (\neg \overset{1}{\bigvee} v \bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x\alpha^*) \wedge \phi(y)).$$

Usando una vez más que en T se prueba que $\overset{1}{\bigvee} u\, \psi_t(u)$, esto equivale a

$$\bigvee u(\psi_t(u) \wedge (\bigwedge v(\mathtt{S}^u_x\alpha^* \leftrightarrow v = y) \vee (\neg \overset{1}{\bigvee} v\, \mathtt{S}^u_x\alpha^* \wedge \phi(y)))),$$

que a su vez equivale a $\bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x\psi_\theta(y))$.

Por último, si x está libre en θ y v está libre en t, tomamos w como en el caso anterior, de modo que $\mathtt{S}^t_x\theta \equiv w|S^t_x\mathtt{S}^w_v\alpha$, luego

$$\psi_{\mathtt{S}^t_x\theta}(y) \equiv \bigwedge w((\mathtt{S}^t_x\mathtt{S}^w_v\alpha)^* \leftrightarrow w = y) \vee (\neg \overset{1}{\bigvee} w(\mathtt{S}^t_x\mathtt{S}^w_v\alpha)^* \wedge \phi(y)).$$

Aplicando la hipótesis de inducción a $\mathtt{S}^w_v\alpha$ y a α como en el caso anterior llegamos a que

$$\psi_{\mathtt{S}^t_x\theta}(y) \leftrightarrow \bigwedge w(\bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x\mathtt{S}^w_v\alpha^*) \leftrightarrow w = y) \vee$$

$$(\neg \overset{1}{\bigvee} w \bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x\mathtt{S}^w_v\alpha^*) \wedge \phi(y)).$$

Por la unicidad de $\psi_t(u)$, esto equivale a

$$\psi_{\mathtt{S}^t_x\theta}(y) \leftrightarrow \bigvee u(\psi_t(u) \wedge (\bigwedge w(\mathtt{S}^u_x\mathtt{S}^w_v\alpha^* \leftrightarrow w = y) \vee (\neg \overset{1}{\bigvee} w\, \mathtt{S}^u_x\mathtt{S}^w_v\alpha^* \wedge \phi(y))))$$

$$\leftrightarrow \bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x(\bigwedge w(\mathtt{S}^w_v\alpha^* \leftrightarrow w = y) \vee (\neg \overset{1}{\bigvee} w\, \mathtt{S}^w_v\alpha^* \wedge \phi(y)))) \leftrightarrow$$

$$\bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x(\bigwedge v(\alpha^* \leftrightarrow v = y) \vee (\neg \overset{1}{\bigvee} v\, \alpha^* \wedge \phi(y)))) \leftrightarrow$$

$$\bigvee u(\psi_t(u) \wedge \mathtt{S}^u_x\psi_\theta(y)).$$

4. Si θ es un axioma de $K_{\mathcal{L}'}$, entonces $\underset{T}{\vdash} \theta^*$.

   Esto es inmediato para los axiomas K1, K2, K3, pues al eliminar sus descriptores obtenemos otro axioma del mismo tipo. Por ejemplo,

   $$(\alpha \to (\beta \to \alpha))^* \equiv \alpha^* \to (\beta^* \to \alpha^*).$$

   Analizamos únicamente los casos en los que la comprobación no es inmediata:

   Si $\theta \equiv \bigwedge u\alpha \to \mathtt{S}^t_u\alpha$, entonces

   $$\theta^* \equiv \bigwedge u\alpha^* \to (S^t_u\alpha)^*,$$

   y por el punto precedente esto equivale en T a

   $$\bigwedge u\alpha^* \to \bigvee v(\psi_t(v) \wedge \mathtt{S}^v_u\alpha^*),$$

   y es fácil probar esto en T teniendo en cuenta que podemos probar $\bigvee v\, \psi_t(v)$.

Si $\theta \equiv \bigwedge u(u = t \to \alpha) \leftrightarrow \mathtt{S}^t_u\alpha$, observemos en primer lugar que

$$(x = t)^* \equiv \bigvee uv(u = x \wedge \psi_t(v) \wedge u = v),$$

que equivale en T a $\psi_t(x)$. Por lo tanto, θ^* es equivalente a

$$\bigwedge u(\psi_t(u) \to \alpha^*) \leftrightarrow \bigvee v(\psi_t(v) \wedge \mathtt{S}^v_u\alpha^*),$$

y esto es ciertamente un teorema de T (siempre teniendo en cuenta la unicidad de $\psi_t(u)$).

Por ejemplo, si suponemos el miembro derecho, tomamos un x que cumpla $\psi_t(x) \wedge \alpha^*(x)$. Si se cumple $\psi_t(y)$, por la unicidad de ψ_t, tiene que ser $y = x$, luego también se cumple $\alpha^*(y)$, lo que nos da la implicación $\bigwedge u(\psi_t(u) \to \alpha^*)$. El recíproco es más sencillo.

Si $\theta \equiv \overset{1}{\bigvee} u\,\alpha \to \mathtt{S}^{u|\alpha}_u\alpha$, entonces

$$\theta^* \leftrightarrow \overset{1}{\bigvee} u\,\alpha^* \to \bigvee x(\psi_{u|\alpha}(x) \wedge \mathtt{S}^x_u\alpha^*),$$

donde

$$\psi_{u|\alpha}(x) \equiv \bigwedge u(\alpha^* \leftrightarrow u = x) \vee (\neg\overset{1}{\bigvee} u\,\alpha^* \wedge \phi(x)).$$

Por lo tanto, si suponemos que existe un único u que cumple α^*, como también sabemos que existe un único x que cumple $\psi_{u|\alpha}(x)$, dicho x cumple la primera fórmula de la disyunción, de donde se sigue que cumple $\alpha^*(x)$, luego tenemos que $\bigvee x(\psi_{u|\alpha}(x) \wedge \mathtt{S}^x_u\alpha^*)$.

Por último, si $\theta \equiv \neg\overset{1}{\bigvee} u\,\alpha \to u|\alpha = v|(v = v)$, entonces θ^* equivale a

$$\neg\overset{1}{\bigvee} u\,\alpha^* \to \bigvee u_1u_2(\psi_{u|\alpha}(u_1) \wedge \psi_{v|v=v}(u_2) \wedge u_1 = u_2),$$

que a su vez equivale a

$$\neg\overset{1}{\bigvee} u\,\alpha^* \to \bigvee x(\psi_{u|\alpha}(x) \wedge \psi_{v|v=v}(x)).$$

Ahora bien, bajo la hipótesis $\neg\overset{1}{\bigvee} u\,\alpha^*$, en T podemos razonar que el único x que cumple $\psi_{u|\alpha}(x)$ es también el único x que cumple $\phi(x)$. Por otro lado tenemos que distinguir dos casos, según si $\overset{1}{\bigvee} v\,v = v$ o no. En el primer caso todo es igual a todo, y el único x que cumple $\psi_{u|\alpha}(x)$ es también el único x que cumple $\psi_{v|v=v}(x)$, mientras que en el segundo caso se razona que el único x que cumple $\psi_{v|v=v}(x)$ es también el único x que cumple $\phi(x)$. En ambos casos tenemos que $\psi_{u|\alpha}(x) \wedge \psi_{v|v=v}(x)$, de donde obtenemos la conclusión.

5. Si θ es un axioma de T', entonces $\underset{T}{\vdash} \theta^*$.

   En efecto, ya lo hemos probado para los axiomas lógicos, también es cierto para los axiomas de T, pues son fórmulas de $\mathcal{L}$ y entonces θ^* equivale

a θ en T, luego θ^* es un teorema. Sólo falta probarlo para el axioma $\theta \equiv \phi(v|v=v) \equiv \mathtt{S}_x^{v|v=v}\phi(x)$. Sabemos entonces que θ^* equivale a

$$\bigvee u(\psi_{v|v=v}(u) \wedge \phi(u)),$$

y en el apartado anterior ya hemos visto que esto es un teorema de T, pues, tanto si $\overset{1}{\bigvee} v\, v = v$ como si no, se cumple que el único x que cumple $\psi_{v|v=v}(x)$ coincide con el único x que cumple $\phi(x)$.

Ahora ya es inmediata la prueba del teorema 3.32: La primera parte la hemos demostrado ya, pues toda fórmula α de $\mathcal{L}'$ es equivalente en T' a la fórmula α^*. Para probar la segunda parte, suponemos que $\alpha_1, \ldots, \alpha_m$ es una demostración en T' de una fórmula $\alpha \equiv \alpha_m$ de $\mathcal{L}$. Basta probar inductivamente que cada α_i^* es un teorema de T, pues en particular tendremos que α^* es un teorema de T, y α^* es equivalente a α en T.

Si α_i es un axioma de T', ya hemos visto que α_i^* es un teorema de T. Si α_i se deduce de las líneas anteriores α_j y $\alpha_j \to \alpha_i$ por MP y α_j^* y $\alpha_j^* \to \alpha_i^*$ son teoremas de T, es obvio que α_i^* también lo es.

Por último, si $\alpha_i \equiv \bigwedge x\alpha_j$ se deduce por IG, entonces $\alpha_i^* \equiv \bigwedge x\alpha_j^*$ también se deduce por IG del teorema α_j^*. ∎

En particular, si una teoría axiomática T sobre un lenguaje sin descriptor es consistente, sigue siéndolo si le añadimos un descriptor (y un axioma de tipo $\phi(v|v=v)$ que determine la descripción impropia).

Capítulo IV

La completitud semántica

Al estudiar las teorías axiomáticas de conjuntos que hemos presentado en el capítulo anterior hemos razonado como razonan habitualmente los matemáticos, es decir, sin cuidarnos de comprobar que, en efecto, todo cuanto decimos puede justificarse paso a paso aplicando oportunamente las reglas de inferencia de $K_{\mathcal{L}}$. Esto puede interpretarse de dos modos distintos: La interpretación más formal es que las demostraciones que hemos dado son en realidad esbozos de demostración, lo suficientemente detallados como para que cualquiera que se lo proponga pueda desarrollarlos hasta convertirlos en auténticas demostraciones, paso a paso, en $K_{\mathcal{L}}$. Pero la realidad es que cuando un matemático razona y valora si sus razonamientos son correctos o no, no tiene en cuenta para nada si tal o cual cosa se podrá justificar con EDI o con MB, sino que da por bueno un argumento si se convence de que si sus hipótesis son ciertas sus conclusiones también tienen que serlo. Pero ¿ciertas en qué sentido? El matemático no se plantea eso, pero no es difícil responder: se da por satisfecho cuando se convence de que si unos objetos (los que sean) cumplen sus hipótesis, necesariamente tienen que cumplir también sus conclusiones. Se puede decir que el matemático en la práctica razona semánticamente (es decir, preocupado de no pasar nunca de afirmaciones verdaderas a falsas) y no formalmente (preocupado de aplicar sólo unas reglas de razonamiento prefijadas).

El problema de razonar semánticamente (si uno quiere a la vez ser riguroso) es que en principio nos pone en la obligación de explicar de qué estamos hablando (es decir, de explicar qué objetos se supone que cumplen nuestras afirmaciones para que podamos decir con sentido "si nuestras hipótesis son verdaderas...") Lo que suele hacer el matemático si le piden este tipo de explicaciones es dar la respuesta fácil (y válida) de que, en realidad, se podría comprobar que todo cuanto razona puede ser formalizado en $K_{\mathcal{L}}$, y eso basta para que sus razonamientos sean rigurosos. En efecto, esto resuelve el problema de fundamentar su trabajo "sin agujeros", porque razonar formalmente en $K_{\mathcal{L}}$ es un proceso objetivo y bien definido, pero, ¿de verdad hace falta recurrir al "se podría comprobar"? ¿Es casualidad que todo lo que un matemático razona sin preocuparse de $K_{\mathcal{L}}$ al final resulta que, en efecto, puede formalizarse en $K_{\mathcal{L}}$? ¿No puede justificarse que razonar semánticamente tiene sentido sin refugiarse en $K_{\mathcal{L}}$?

En este capítulo responderemos afirmativamente a las dos últimas preguntas. De hecho, veremos que la respuesta a la segunda lleva fácilmente a la respuesta a la primera. Notemos en primer lugar que para que tenga sentido razonar formalmente no es necesario dar un modelo explícito de los axiomas aceptados, porque el argumento no es: "tales objetos concretos cumplen lo que digo", sino "si unos objetos (los que sean) cumplen mis axiomas, también tienen que cumplir mis teoremas", por lo que sólo necesitamos justificar que existen unos objetos (los que sean) que cumplan mis axiomas.

Eso es precisamente lo que afirma el teorema siguiente, debido a Gödel (aunque la prueba que daremos se debe a Henkin):

Teorema 4.1 *Si una teoría axiomática es consistente, entonces tiene un modelo numerable.*[1]

Combinando esto con el teorema 3.4, podemos afirmar que una teoría axiomática es consistente si y sólo si tiene un modelo. Por lo tanto, el único requisito para que el razonamiento semántico tenga sentido es garantizar que sus axiomas son consistentes (requisito obviamente necesario, por otra parte). Ahora bien, una cosa es que tenga sentido y otra distinta que razonar semánticamente sea lo mismo que razonar formalmente. El teorema que garantiza esto es el que se conoce propiamente como teorema de completitud semántica de Gödel, pero como es una consecuencia sencilla de 4.1 y en la práctica se apela más veces a éste, no es raro llamar teorema de completitud a 4.1, aunque la denominación no sea exacta.

4.1 Conjuntos maximalmente consistentes

Si tuviéramos probado el teorema 4.1 podríamos concluir fácilmente que toda teoría consistente puede extenderse (añadiéndole axiomas) a una teoría consistente y completa. Basta fijar un modelo y tomar como axiomas todas las sentencias verdaderas en él. Obviamente la teoría con tales axiomas es consistente y completa. De hecho, no hay nada que deducir de ella: toda sentencia que sea consecuencia lógica de los axioma es también un axioma. Planteado así es una obviedad, pero vamos a ver que es posible extender una teoría consistente hasta una teoría consistente y completa sin apoyarnos en ningún modelo, y a partir de ella construir el modelo que buscamos.

Conviene trabajar con un concepto de extensión de una teoría axiomática un poco más general que el consistente en añadir más axiomas:

Definición 4.2 Diremos que una teoría axiomática S es una *extensión* de una teoría T si el lenguaje formal de S contiene a todos los signos del lenguaje de T y todos los axiomas de T son teoremas de S.

[1] Un modelo numerable es un modelo tal que es posible ordenar los objetos de su universo, ya sea en una sucesión finita $a_0, \ldots, a_n$, ya infinita $a_0, a_1, a_2, \ldots$ La numerabilidad no es relevante para lo que estamos discutiendo ahora, pero tiene consecuencias muy interesantes que comentaremos luego sobre las limitaciones de la teoría de conjuntos.

Es claro que si S es una extensión de T, entonces todos los teoremas de T son teoremas de S. La forma habitual de extender una teoría T consiste en formar una nueva teoría S que tenga más axiomas, pero la definición que hemos dado permite que se eliminen algunos axiomas de T y en su lugar se añadan otros axiomas más fuertes.[2]

Para probar el teorema de completitud necesitaremos una versión refinada del teorema siguiente, pero incluimos también esta versión porque el resultado tiene interés en sí mismo y muestra más claramente la idea básica. El lector debería prestar especial atención a su demostración, porque sucede que la demostración del teorema de completitud no es constructiva, no es finitista, y el núcleo no finitista del argumento es justo la demostración siguiente. Si el lector se ha preguntado alguna vez qué presupuestos son necesarios para dar sentido a todos los planteamientos metamatemáticos que estamos presentando en este libro, la respuesta es: ni más ni menos que los necesarios para aceptar que la prueba del teorema siguiente tiene sentido y es concluyente.

Teorema 4.3 *Toda teoría axiomática consistente tiene una extensión consistente y completa.*

DEMOSTRACIÓN: Sea T una teoría axiomática consistente sobre un lenguaje formal $\mathcal{L}$. Fijemos una enumeración $\alpha_0, \alpha_1, \alpha_2, \ldots$ de las sentencias de $\mathcal{L}$.

Sea Γ el conjunto de los axiomas de T y sea Γ_0 el conjunto de las clausuras universales de las fórmulas de Γ. Es claro que para toda fórmula α se cumple $\Gamma \vdash \alpha$ si y sólo si $\Gamma_0 \vdash \alpha$. Por el teorema 3.5 tenemos que Γ_0 es un conjunto consistente de sentencias de $\mathcal{L}$. Para cada número natural n, definimos

$$\Gamma_{n+1} \equiv \begin{cases} \Gamma_n & \text{si } \Gamma_n \cup \{\alpha_n\} \text{ es contradictorio,} \\ \Gamma_n \cup \{\alpha_n\} & \text{si } \Gamma_n \cup \{\alpha_n\} \text{ es consistente.} \end{cases}$$

Por construcción todos los conjuntos Γ_n son consistentes y si $m \leq n$ entonces Γ_m está contenido en Γ_n. Sea Γ_∞ la unión de todos los conjuntos Γ_n. Es claro que Γ_∞ es consistente, pues si de sus sentencias se dedujera una contradicción, ésta se deduciría de hecho de un número finito de ellas, y todas estarían contenidas en un cierto Γ_n, que sería, pues, contradictorio.

Sea S la teoría axiomática cuyo conjunto de axiomas es Γ_∞. Ciertamente es consistente. Como Γ está contenido en Γ_∞, es claro que S es una extensión de T (los teoremas de T son las consecuencias de Γ, luego también son consecuencias de Γ_0 y de Γ_∞). Veamos por último que S es completa.

Sea α una sentencia de $\mathcal{L}$. Entonces $\alpha \equiv \alpha_i$ para un cierto i. Supongamos que no $\underset{S}{\vdash} \neg\alpha_i$, o sea, que no $\Gamma_\infty \vdash \neg\alpha_i$. Entonces tampoco $\Gamma_i \vdash \neg\alpha_i$. Por el teorema 3.6, el conjunto $\Gamma_i \cup \{\alpha_i\}$ es consistente, y así $\Gamma_{i+1} = \Gamma_i \cup \{\alpha_i\}$, luego α_i está en Γ_∞ y por consiguiente $\underset{S}{\vdash} \alpha_i$. ∎

[2]Por ejemplo, hemos visto que la teoría Z* extiende a la teoría básica B a pesar de que un axioma de B no lo es de Z*.

Observaciones La enumeración $\alpha_0, \alpha_1, \ldots$ de las sentencias de un lenguaje formal puede hacerse explícitamente. Más adelante veremos detalladamente una forma de hacerlo, por lo que por ahora no insistiremos más en ello. Lo importante es que una tal ordenación es un concepto finitista.

Un punto mucho más delicado es la definición de los conjuntos Γ_n, pues, según veremos más adelante, en los casos de interés matemático no es posible calcular explícitamente cada uno de sus términos. Aunque pudiéramos calcular los primeros, digamos hasta Γ_5, nada nos garantiza que seamos capaces de determinar quién es Γ_6 o, más concretamente, si la sentencia α_5 forma parte o no de Γ_6. El problema es que para ello tendríamos que decidir si $\Gamma_5 \cup \{\alpha_5\}$ es o no consistente, y no tenemos ningún algoritmo que nos permita decidir si un conjunto de fórmulas, aunque sea finito, es consistente o no.

Pese a ello, lo cierto es que $\Gamma_5 \cup \{\alpha_5\}$ será consistente o contradictorio y, según el caso Γ_6 coincidirá con este conjunto o se reducirá a Γ_5. Puesto que uno de los dos casos, y sólo uno, ha de ser cierto, podemos afirmar que Γ_6 está bien definido con independencia de si sabemos o no determinar sus elementos.

Así pues, el conjunto Γ_∞ de los axiomas de la extensión S está completamente determinado por T y por la ordenación de las sentencias de $\mathcal{L}$ que hemos escogido, a pesar de que no sabemos determinar qué sentencias contiene. Estamos ante el tipo de colecciones de objetos más general que vamos a considerar desde un punto de vista metamatemático.

La teoría S es bastante "patológica", pues, aunque conozcamos perfectamente la teoría de partida T, lo cierto es que no sabemos qué sentencias son axiomas de S y, *a fortiori,* qué sentencias son teoremas de S. Más adelante veremos que esta patología es inevitable si partimos de una teoría aritmética consistente.

A diferencia de lo que sucede con el teorema de completitud, este teorema afirma simplemente la existencia de un objeto bien definido que escapa a nuestro control. En sí mismo no tiene repercusiones finitistas. Por ello no es un resultado indicado para valorar si tenemos realmente motivos para aceptar razonamientos no finitistas. Ciertamente, si las técnicas no finitistas nos llevaran únicamente a conclusiones de este tipo, resultarían ser totalmente superfluas. ∎

El papel que representa la completitud en la prueba del teorema 4.1 es, a grandes rasgos, el siguiente: un modelo de una teoría axiomática determina si una sentencia dada es verdadera o falsa. Por consiguiente, para construir un modelo debemos contar con toda la información necesaria para aceptar o rechazar cualquier sentencia y, para ello, uno de los primeros pasos que daremos será completar la teoría de partida. Si nos fijamos en la teoría S construida en la prueba del teorema anterior veremos que una sentencia es un teorema de S si y sólo si es un axioma. Para no trabajar con teorías "hinchadas" de axiomas, conviene tratar directamente con el conjunto de las sentencias demostrables en una teoría axiomática, ahorrándonos así el darles artificialmente rango de axiomas. Esto nos lleva al concepto siguiente:

Definición 4.4 Un conjunto Γ de sentencias de un lenguaje formal $\mathcal{L}$ es *maximalmente consistente* si Γ es consistente y para toda sentencia α de $\mathcal{L}$ que no esté en Γ se cumple que $\Gamma \cup \{\alpha\}$ es contradictorio.

La relación con la completitud es la siguiente:

Teorema 4.5 *Una teoría axiomática T es consistente y completa si y sólo si el conjunto Γ de todas las sentencias demostrables en T es maximalmente consistente.*

DEMOSTRACIÓN: Supongamos que T es consistente y completa. Entonces Γ es consistente, pues las consecuencias lógicas de los teoremas de T son teoremas de T, luego si a partir de Γ pudiera demostrarse α y $\neg\alpha$, estas fórmulas serían teoremas de T.

Sea α una sentencia que no esté en Γ, es decir, tal que no $\underset{T}{\vdash}\alpha$. Como T es completa, $\underset{T}{\vdash}\neg\alpha$, luego $\neg\alpha$ está en Γ. Es claro entonces que $\Gamma\cup\{\alpha\}$ es contradictorio. Por lo tanto Γ es maximalmente consistente.

Recíprocamente, si Γ es maximalmente consistente, entonces T es consistente, pues hay sentencias que no son teoremas de T (las negaciones de las sentencias de Γ). Además, si α es una sentencia, o bien $\underset{T}{\vdash}\alpha$ o, en caso contrario, α no está en Γ, luego $\Gamma\cup\{\alpha\}$ es contradictorio luego, por el teorema 3.6, tenemos que $\Gamma\vdash\neg\alpha$, y a su vez esto implica que $\underset{T}{\vdash}\neg\alpha$. ∎

El teorema siguiente recoge las propiedades básicas de los conjuntos maximalmente consistentes. Notemos que en él aparecen conexiones estrictamente sintácticas (es decir, no basadas en ningún modelo) entre los signos lógicos y su significado.

Teorema 4.6 *Sea Γ un conjunto maximalmente consistente de sentencias de un lenguaje formal $\mathcal{L}$ y α, β dos sentencias de $\mathcal{L}$. Entonces*

1. *$\Gamma\vdash\alpha$ syss α está en Γ,*
2. *Si $\vdash\alpha$, entonces α está en Γ,*
3. *$\neg\alpha$ está en Γ syss α no está en Γ,*
4. *$\alpha\to\beta$ está en Γ syss α no está en Γ o β está en Γ,*
5. *$\alpha\vee\beta$ está en Γ syss α está en Γ o β está en Γ,*
6. *$\alpha\wedge\beta$ está en Γ syss α está en Γ y β está en Γ,*
7. *$\alpha\leftrightarrow\beta$ está en Γ syss α y β están ambas en Γ o ninguna lo está.*

DEMOSTRACIÓN: 1) Si $\Gamma\vdash\alpha$ entonces no $\Gamma\vdash\neg\alpha$, porque Γ es consistente, luego $\Gamma\cup\{\alpha\}$ es consistente, por el teorema 3.6, luego α está en Γ. El recíproco es obvio.

2) Es consecuencia de 1).

3) Si $\neg\alpha$ está en Γ, entonces α no puede estar en Γ, porque Γ es consistente. Si α no está en Γ entonces $\Gamma\cup\{\alpha\}$ es contradictorio, luego por el teorema 3.6 se cumple que $\Gamma\vdash\neg\alpha$ y por 1) concluimos que $\neg\alpha$ está en Γ.

4) Si $\alpha \to \beta$ está en Γ y α está en Γ, entonces $\Gamma \vdash \beta$, luego por 1) concluimos que β está en Γ.

Si α no está en Γ o β está en Γ, por 3) $\neg\alpha$ está en Γ o β está en Γ. Por consiguiente $\Gamma \vdash \neg\alpha$ o $\Gamma \vdash \beta$. En cualquier caso $\Gamma \vdash \alpha \to \beta$ y por 1) $\alpha \to \beta$ está en Γ.

5), 6) y 7) se deducen de 3) y 4) por las definiciones de los conectores. ∎

Necesitamos propiedades análogas a las de este teorema pero en relación a los cuantificadores. Para ello introducimos una nueva noción:

Definición 4.7 Un conjunto Γ de sentencias de un lenguaje formal $\mathcal{L}$ es *ejemplificado* si cuando $\bigvee x\alpha$ está en Γ existe un designador t de $\mathcal{L}$ tal que $\mathtt{S}^t_x\alpha$ está en Γ.

Es decir, Γ está ejemplificado si cuando afirma la existencia de un objeto que cumple algo, nos da también un ejemplo concreto t de objeto que cumple lo indicado. En realidad se cumple mucho más:

Teorema 4.8 *Sea Γ un conjunto de sentencias de un lenguaje formal $\mathcal{L}$ maximalmente consistente y ejemplificado. Sea α una fórmula de $\mathcal{L}$ en la que a lo sumo esté libre la variable x. Entonces*

1. *$\bigvee x\alpha$ está en Γ syss existe un designador t de $\mathcal{L}$ tal que $\mathtt{S}^t_x\alpha$ está en Γ.*
2. *$\bigwedge x\alpha$ está en Γ syss para todo designador t de $\mathcal{L}$ se cumple que $\mathtt{S}^t_x\alpha$ está en Γ.*

Demostración: 1) Si $\bigvee x\alpha$ está en Γ, hay un designador t de $\mathcal{L}$ tal que $\mathtt{S}^t_x\alpha$ está en Γ por ser Γ ejemplificado.

Si $\mathtt{S}^t_x\alpha$ está en Γ, por (IP) obtenemos que $\Gamma \vdash \bigvee x\alpha$, con lo que $\bigvee x\alpha$ está en Γ por el teorema anterior.

2) Si $\bigwedge x\alpha$ está en Γ y t es un designador de $\mathcal{L}$, por (EG) se cumple $\Gamma \vdash \mathtt{S}^t_x\alpha$ y por consiguiente $\mathtt{S}^t_x\alpha$ está en Γ (por el teorema anterior).

Si $\mathtt{S}^t_x\alpha$ está en Γ para todo designador t de $\mathcal{L}$, entonces el teorema anterior nos da que $\neg\mathtt{S}^t_x\alpha \equiv \mathtt{S}^t_x\neg\alpha$ no está en Γ para todo t. Por 1) $\bigvee x\neg\alpha$ no está en Γ y por el teorema 4.6 otra vez concluimos que $\neg\bigvee x\neg\alpha$ sí lo esta. Aplicando (NP) resulta que $\Gamma \vdash \bigwedge x\alpha$, luego, por el teorema anterior una vez más, concluimos que $\bigwedge x\alpha$ está en Γ. ∎

Con esto tenemos todos los conceptos necesarios para demostrar el teorema 4.1. Nos dedicamos a ello en la sección siguiente.

4.2 La prueba del teorema de completitud

El teorema 4.3 puede reformularse como que todo conjunto consistente de sentencias está contenido en un conjunto maximalmente consistente. Sin embargo, necesitaremos un conjunto que además sea ejemplificado y ello plantea un problema técnico que hemos de resolver previamente. La clave será el teorema siguiente:

Teorema 4.9 *Sea $\mathcal{L}$ un lenguaje formal y sea $\mathcal{L}'$ un lenguaje formal que conste de los mismos signos que $\mathcal{L}$ más una constante c (aunque admitimos el caso de que c esté en $\mathcal{L}$ y, por consiguiente, que $\mathcal{L}$ coincida con $\mathcal{L}'$). Si c no está en una fórmula α, y $\underset{K_{\mathcal{L}'}}{\vdash} \mathtt{S}^c_x\alpha$ con una demostración en la que no aparezca la variable x, entonces $\underset{K_{\mathcal{L}}}{\vdash} \alpha$.*

DEMOSTRACIÓN: Veamos que $\underset{K_{\mathcal{L}}}{\vdash} \alpha$ por inducción sobre el número de líneas de una demostración en la que no aparezca la variable x. Si $\mathtt{S}^c_x\alpha$ se demuestra en una línea, entonces es un axioma de $K_{\mathcal{L}'}$. Veamos que α ha de ser un axioma de $K_{\mathcal{L}}$. Para probarlo nos basaremos en dos hechos obvios:

A Si θ es una expresión de $\mathcal{L}'$ que no contiene la variable x, entonces $\theta \equiv \mathtt{S}^c_x\theta_0$, para una expresión θ_0 de $\mathcal{L}$ que no contiene a c. (θ_0 es la expresión que resulta de cambiar por x cada aparición de c en θ.)

B Si β_0 y β_1 son fórmulas de $\mathcal{L}$ que no contienen a c y $\mathtt{S}^c_x\beta_0 \equiv \mathtt{S}^c_x\beta_1$ entonces $\beta_0 \equiv \beta_1$.

Así, si por ejemplo $\mathtt{S}^c_x\alpha \equiv \beta \to (\gamma \to \beta)$, entonces por A

$$\mathtt{S}^c_x\alpha \equiv \beta \to (\gamma \to \beta) \equiv \mathtt{S}^c_x\beta_0 \to (\mathtt{S}^c_x\gamma_0 \to \mathtt{S}^c_x\beta_0) \equiv \mathtt{S}^c_x\big(\beta_0 \to (\gamma_0 \to \beta_0)\big),$$

luego por B tenemos que $\alpha \equiv \beta_0 \to (\gamma_0 \to \beta_0)$.

Similarmente, si $\mathtt{S}^c_x\alpha \equiv \bigwedge y\, \beta \to \mathtt{S}^t_y\beta$, entonces

$$\mathtt{S}^c_x\alpha \equiv \bigwedge y\, \mathtt{S}^c_x\beta_0 \to \mathtt{S}^{\mathtt{S}^c_x t_0}_y \mathtt{S}^c_x\beta_0 \equiv \bigwedge y\, \mathtt{S}^c_x\beta_0 \to \mathtt{S}^c_x\mathtt{S}^{t_0}_y\beta_0 \equiv \mathtt{S}^c_x(\bigwedge y\, \beta_0 \to \mathtt{S}^{t_0}_y\beta_0),$$

luego $\alpha \equiv \bigwedge y\, \beta_0 \to \mathtt{S}^{t_0}_y\beta_0$.

La comprobación para los restantes esquemas axiomáticos es análoga.

Si el teorema es cierto para las fórmulas demostrables en menos de n pasos, supongamos que $\mathtt{S}^c_x\alpha$ se demuestra en n pasos.

a) Si $\mathtt{S}^c_x\alpha$ se deduce por (MP) de β y $\beta \to \mathtt{S}^c_x\alpha$, líneas anteriores. Por A) podemos expresar $\beta \equiv \mathtt{S}^c_x\gamma$, donde γ no contiene a c.

Observemos que $\beta \to \mathtt{S}^c_x\alpha \equiv \mathtt{S}^c_x\gamma \to \mathtt{S}^c_x\alpha \equiv \mathtt{S}^c_x(\gamma \to \alpha)$, la constante c no está en γ ni en $\gamma \to \alpha$ y x no está ligada en γ ni en $\gamma \to \alpha$. Por hipótesis de inducción $\underset{K_{\mathcal{L}}}{\vdash} \gamma$ y $\underset{K_{\mathcal{L}}}{\vdash} \gamma \to \alpha$, luego $\underset{K_{\mathcal{L}}}{\vdash} \alpha$.

b) Si $\mathtt{S}^c_x\alpha \equiv \bigwedge y\beta$ se deduce de β por (IG), entonces $y \not\equiv x$, pues x no aparece en la demostración. Sea γ la fórmula resultante de sustituir c por x en β. Como antes $\beta \equiv \mathtt{S}^c_x\gamma$. Además $\mathtt{S}^c_x\alpha \equiv \bigwedge y\mathtt{S}^c_x\gamma \equiv \mathtt{S}^c_x\bigwedge y\gamma$. De aquí se sigue que $\alpha \equiv \bigwedge y\gamma$. Por hipótesis de inducción $\underset{K_{\mathcal{L}}}{\vdash} \gamma$, luego $\underset{K_{\mathcal{L}}}{\vdash} \alpha$. ∎

En definitiva, la prueba del teorema muestra que basta reemplazar todas las apariciones de c por apariciones de x en una demostración de $\mathtt{S}^c_x\alpha$ para tener una demostración de α.

El teorema siguiente es el que nos permitirá ejemplificar un conjunto consistente de sentencias para volverlo a la vez ejemplificado y maximalmente consistente.

Teorema 4.10 *Si $\Gamma \cup \{\bigvee x\, \alpha\}$ es un conjunto consistente de sentencias de un lenguaje formal $\mathcal{L}$, el lenguaje $\mathcal{L}'$ es como en el teorema anterior y la constante c no está en ninguna sentencia de $\Gamma \cup \{\bigvee x\, \alpha\}$, entonces $\Gamma \cup \{\bigvee x\, \alpha\} \cup \{\mathtt{S}^c_x \alpha\}$ es consistente.*

Demostración: Si $\Gamma \cup \{\bigvee x\, \alpha\} \cup \{\mathtt{S}^c_x \alpha\}$ es contradictorio, por el teorema 3.6 tenemos que $\Gamma \cup \{\bigvee x\, \alpha\} \underset{K_{\mathcal{L}'}}{\vdash} \neg \mathtt{S}^c_x \alpha$.

Existen $\gamma_1, \ldots, \gamma_n$ en Γ tales que $\gamma_1 \wedge \cdots \wedge \gamma_n \wedge \bigvee x\, \alpha \underset{K_{\mathcal{L}'}}{\vdash} \neg \mathtt{S}^c_x \alpha$. Entonces

$$\gamma_1 \wedge \cdots \wedge \gamma_n \wedge \bigvee x\, \alpha \underset{K_{\mathcal{L}'}}{\vdash} \mathtt{S}^c_x \neg \alpha,$$

luego por el teorema de deducción

$$\underset{K_{\mathcal{L}'}}{\vdash} \gamma_1 \wedge \cdots \wedge \gamma_n \wedge \bigvee x\, \alpha \to \mathtt{S}^c_x \neg \alpha.$$

Sea y una variable que no aparezca en la demostración. Tenemos que

$$\underset{K_{\mathcal{L}'}}{\vdash} \gamma_1 \wedge \cdots \wedge \gamma_n \wedge \bigvee x\, \alpha \to \mathtt{S}^c_y \mathtt{S}^y_x \neg \alpha,$$

y esto equivale a

$$\underset{K_{\mathcal{L}'}}{\vdash} \mathtt{S}^c_y (\gamma_1 \wedge \cdots \wedge \gamma_n \wedge \bigvee x\, \alpha \to \mathtt{S}^y_x \neg \alpha),$$

pues y no está libre en $\gamma_1 \wedge \cdots \wedge \gamma_n \wedge \bigvee x\, \alpha$. Por el teorema anterior

$$\underset{K_{\mathcal{L}}}{\vdash} \gamma_1 \wedge \cdots \wedge \gamma_n \wedge \bigvee x\, \alpha \to \mathtt{S}^y_x \neg \alpha,$$

y de aquí que $\Gamma \cup \{\bigvee x\, \alpha\} \underset{K_{\mathcal{L}}}{\vdash} \neg \mathtt{S}^y_x \alpha$. Aplicando (IG) y (NP) llegamos a que $\Gamma \cup \{\bigvee x\, \alpha\} \underset{K_{\mathcal{L}}}{\vdash} \neg \bigvee y \mathtt{S}^y_x \alpha$, de donde se concluye que $\Gamma \cup \{\bigvee x\, \alpha\} \underset{K_{\mathcal{L}}}{\vdash} \neg \bigvee x\, \alpha$, con lo que $\Gamma \cup \{\bigvee x\, \alpha\}$ resulta ser contradictoria. ∎

Aunque el teorema siguiente no lo necesitaremos hasta un poco más adelante, lo incluimos aquí porque su prueba es completamente análoga a la del teorema anterior.

Teorema 4.11 *Si Γ es un conjunto consistente de sentencias de un lenguaje formal con descriptor $\mathcal{L}$, el lenguaje $\mathcal{L}'$ es como en el teorema anterior y la constante c no está en $\mathcal{L}$, entonces $\Gamma \cup \{c = x | (x = x)\}$ es consistente.*

DEMOSTRACIÓN: Si $\Gamma \cup \{c = x|(x = x)\}$ es contradictorio, por el teorema 3.6 tenemos que $\Gamma \underset{K_{\mathcal{L}'}}{\vdash} \neg c = x|(x = x)$. Existen sentencias $\gamma_1, \ldots, \gamma_n$ en Γ tales que $\gamma_1 \wedge \cdots \wedge \gamma_n \underset{K_{\mathcal{L}'}}{\vdash} \neg c = x|(x = x)$ y por el teorema de deducción

$$\underset{K_{\mathcal{L}'}}{\vdash} \gamma_1 \wedge \cdots \wedge \gamma_n \to \neg c = x|(x = x).$$

Sea y una variable que no aparezca en la demostración. Tenemos que

$$\underset{K_{\mathcal{L}'}}{\vdash} \mathtt{S}_y^c(\gamma_1 \wedge \cdots \wedge \gamma_n \to \neg y = x|(x = x)),$$

luego por el teorema 4.9 también $\underset{K_{\mathcal{L}}}{\vdash} \gamma_1 \wedge \cdots \wedge \gamma_n \to \neg y = x|(x = x)$. Así pues, $\Gamma \underset{K_{\mathcal{L}}}{\vdash} \neg y = x|(x = x)$. Aplicando (IG) resulta que $\Gamma \underset{K_{\mathcal{L}}}{\vdash} \bigwedge y \neg y = x|(x = x)$, y aplicando (EG) llegamos a $\Gamma \underset{K_{\mathcal{L}}}{\vdash} \neg(x|(x = x)) = (x|(x = x))$, de donde se sigue que Γ es contradictorio. ∎

Ahora ya podemos probar la base de nuestro argumento hacia el teorema de completitud:

Teorema 4.12 *Sea $\mathcal{L}$ un lenguaje formal, sea $\mathcal{L}'$ un lenguaje formal que conste de los mismos signos que $\mathcal{L}$ más una sucesión de constantes $d_0, d_1, \ldots$ que no estén en $\mathcal{L}$, sea Γ un conjunto consistente de sentencias de $\mathcal{L}$. Existe un conjunto Γ_∞ maximalmente consistente y ejemplificado de sentencias de $\mathcal{L}'$ que contiene a Γ.*

DEMOSTRACIÓN: Sea $\alpha_0, \alpha_1, \ldots$ una enumeración de las sentencias de $\mathcal{L}'$. Definimos $\Gamma_0 \equiv \Gamma$ y, supuesto definido Γ_n, sea

$$\Gamma_{n+1} \equiv \begin{cases} \Gamma_n & \text{si } \Gamma_n \cup \{\alpha_n\} \text{ es contradictorio,} \\ \Gamma_n \cup \{\alpha_n\} & \text{si } \Gamma_n \cup \{\alpha_n\} \text{ es consistente y } \alpha_n \text{ no es} \\ & \text{de la forma } \bigvee x\, \beta, \\ \Gamma_n \cup \{\alpha_n\} \cup \{\mathtt{S}_x^{d_k}\beta\} & \text{si } \Gamma_n \cup \{\alpha_n\} \text{ es consistente, } \alpha_n \equiv \bigvee x\, \beta \text{ y} \\ & k \text{ es el menor natural tal que } d_k \text{ no está} \\ & \text{en } \Gamma_n \cup \{\alpha_n\}. \end{cases}$$

Por el teorema 4.10, cada Γ_n es consistente. Sea Γ_∞ la unión de todos los conjuntos Γ_n. Como en el teorema 4.3, es claro que Γ_∞ es consistente. De hecho es maximalmente consistente, pues si una sentencia $\alpha \equiv \alpha_i$ de $\mathcal{L}'$ no está en Γ_∞, entonces $\Gamma_i \cup \{\alpha_i\}$ es contradictorio o, de lo contrario, α_i estaría en Γ_{i+1}, luego en Γ_∞. Por consiguiente $\Gamma_\infty \cup \{\alpha_i\}$ también es contradictorio.

Por último veamos que Γ_∞ es ejemplificado. Si $\bigvee x\, \alpha$ está en Γ_∞, entonces $\bigvee x\, \alpha \equiv \alpha_j$ para algún natural j. Como $\Gamma_j \cup \{\bigvee x\, \alpha\}$ está contenido en Γ_∞, ciertamente es consistente. Por construcción $\Gamma_{j+1} = \Gamma_j \cup \{\bigvee x\, \alpha\} \cup \{\mathtt{S}_x^{d_k}\alpha\}$, luego $\mathtt{S}_x^{d_k}\alpha$ está en Γ_∞ y d_k es un designador de $\mathcal{L}'$. ∎

El conjunto de sentencias Γ_∞ construido en la prueba del teorema anterior tiene las mismas características que el construido en 4.3, es decir, está unívocamente determinado a partir de Γ y de la enumeración fijada de las sentencias de $\mathcal{L}$, pero no tenemos ningún algoritmo que nos permita decidir si una sentencia dada está o no en Γ_∞.

A partir de un conjunto de sentencias Γ ejemplificado y maximalmente consistente es fácil construir un modelo. Para ello hemos de buscar un conjunto de objetos que, con las relaciones y funciones adecuadas, verifiquen cuanto se afirma en Γ. Ahora bien, puesto que Γ proporciona designadores concretos que nombran a cada uno de los objetos de los que pretende hablar, podemos tomar como universo del modelo los propios designadores del lenguaje formal de Γ, y arreglar las definiciones de forma que cada designador se denote a sí mismo. Aquí nos aparece un problema técnico, y es que si Γ contiene una sentencia de tipo $t_1 = t_2$, donde t_1 y t_2 son designadores distintos, entonces t_1 y t_2 deben denotar al mismo objeto, lo cual no sucederá si, como pretendemos, cada uno se denota a sí mismo. Para corregir este inconveniente no tomaremos como universo a los designadores exactamente, sino a las clases de equivalencia respecto a la relación que satisfacen dos designadores t_1 y t_2 precisamente cuando $t_1 = t_2$ está en Γ. Veamos los detalles:

Teorema 4.13 *Sea $\mathcal{L}$ un lenguaje formal (con o sin descriptor) y sea Γ un conjunto consistente de sentencias sin descriptores de $\mathcal{L}$. Entonces Γ tiene un modelo (de universo) numerable.*

DEMOSTRACIÓN: Recordemos que $\underline{\mathcal{L}}$ es el lenguaje que resulta de eliminar el descriptor de $\mathcal{L}$. Sea Γ_∞ un conjunto de sentencias de $\underline{\mathcal{L}}'$ maximalmente consistente y ejemplificado que contenga a Γ según el teorema anterior. Basta probar que Γ_∞ tiene un modelo numerable M_∞, pues entonces ciertamente $M_\infty \models \Gamma$ y si llamamos M al modelo de $\underline{\mathcal{L}}$ que se diferencia de M_∞ en que no interpreta las constantes nuevas, es claro que $M \models \Gamma$. Finalmente, seleccionando si es necesario una descripción impropia, podemos considerar a M como modelo de $\mathcal{L}$ (pues la descripción impropia no influye en la interpretación de fórmulas sin descriptores).

Equivalentemente, podemos suponer que $\mathcal{L}$ no tiene descriptor y que Γ es un conjunto maximalmente consistente y ejemplificado de sentencias de $\mathcal{L}$.

Sea T el conjunto de todos los designadores de $\mathcal{L}$. Consideramos en T la relación diádica dada por $t_1 \sim t_2$ syss la sentencia $t_1 = t_2$ está en Γ. Veamos que se trata de una relación de equivalencia.

Dado un designador t, ciertamente $\vdash t = t$, luego $t = t$ está en Γ por el teorema 4.6. Por consiguiente la relación es reflexiva.

Dados dos designadores t_1 y t_2 tales que $t_1 \sim t_2$, esto significa que $t_1 = t_2$ está en Γ, luego $\Gamma \vdash t_2 = t_1$ y por consiguiente $t_2 = t_1$ está también en Γ. Esto prueba la simetría y similarmente se prueba la transitividad.

Representaremos por $[t]$ a la clase de equivalencia de t respecto a $\sim$ y llamaremos U al conjunto de todas las clases de equivalencia. Es claro que U es un conjunto numerable, ya que lo es el conjunto de los designadores de $\mathcal{L}$.

Definimos como sigue un modelo M de $\mathcal{L}$:

- El universo de M es U.
- Si c es una constante de $\mathcal{L}$, entonces $M(c) = [c]$.

- Si R_i^n es un relator n-ádico de $\mathcal{L}$, entonces $M(R_i^n)$ es la relación n-ádica dada por

$$M(R_i^n)([t_1],\ldots,[t_n]) \quad \text{syss} \quad R_i^n t_1\cdots t_n \text{ está en } \Gamma.$$

- Si f_i^n es un funtor n-ádico de $\mathcal{L}$, entonces $M(f_i^n)$ es la función n-ádica en U dada por

$$M(f_i^n)([t_1],\ldots,[t_n]) = [f_i^n t_1\cdots t_n].$$

Hemos de comprobar que las interpretaciones de los relatores y los funtores están bien definidas. En el caso de los relatores esto significa que si se cumple $[t_1] = [t'_1],\ldots,[t_n] = [t'_n]$ entonces $R_i^n t_1\cdots t_n$ está en Γ syss $R_i^n t'_1\cdots t'_n$ está en Γ.

En efecto, tenemos que las sentencias $t_1 = t'_1,\ \ldots,\ t_n = t'_n$ están en Γ, de donde se sigue fácilmente que $\Gamma \vdash R_i^n t_1\cdots t_n \leftrightarrow R_i^n t'_1\cdots t'_n$. Por la consistencia maximal, la sentencia $R_i^n t_1\cdots t_n \leftrightarrow R_i^n t'_1\cdots t'_n$ está en Γ, y por el teorema 4.6 tenemos que $R_i^n t_1\cdots t_n$ está en Γ syss $R_i^n t'_1\cdots t'_n$ está en Γ.

Análogamente se prueba que las funciones $M(f_i^n)$ están bien definidas. También hemos de comprobar que $M(=)$ es la relación de igualdad, pero esto es inmediato a partir de las definiciones.

Ahora probamos que si t es un designador de $\mathcal{L}$ y v es cualquier valoración, entonces $M(t)[v] = [t]$. Si t es una constante esto es cierto por definición de M. Como $\mathcal{L}$ no tiene descriptores, la única posibilidad adicional es que $t \equiv f_i^n t_1\cdots t_n$, en cuyo caso, razonando por inducción sobre la longitud de t, podemos suponer que $M(t_i)[v] = [t_i]$, y por definición de $M(f_i^n)$ llegamos a que $M(t)[v] = [t]$.

Finalmente probamos que si α es una sentencia de $\mathcal{L}$, entonces $M \vDash \alpha$ si y sólo si α está en Γ. En particular $M \vDash \Gamma$, que es lo que queríamos probar. Razonamos por inducción sobre el número de signos lógicos ($\neg$, $\rightarrow$, $\bigwedge$) que contiene α. Notar que los términos de $\mathcal{L}$ no tienen descriptores, por lo que no pueden contener signos lógicos.

Si α no tiene signos lógicos, entonces $\alpha \equiv R_i^n t_1\cdots t_n$ y se cumple que

$$M \vDash \alpha \text{ syss } M(R_i^n)([t_1],\ldots,[t_n]) \text{ syss } \alpha \equiv R_i^n t_1\cdots t_n \text{ está en } \Gamma.$$

Supuesto cierto para sentencias con menos signos lógicos que α, distinguimos tres casos:

a) Si $\alpha \equiv \neg\beta$, entonces $M \vDash \alpha$ syss no $M \vDash \beta$ syss (hip. de ind.) β no está en Γ syss (teorema 4.6) $\neg\beta$ está en Γ syss α está en Γ.

b) Si $\alpha \equiv \beta \rightarrow \gamma$, entonces $M \vDash \alpha$ syss no $M \vDash \beta$ o $M \vDash \gamma$ syss (hip. ind.) β no está en Γ o γ está en Γ syss (teorema 4.6) $\beta \rightarrow \gamma$ está en Γ syss α está en Γ.

c) Si $\alpha \equiv \bigwedge x\beta$, entonces $M \vDash \alpha$ syss para todo $[t]$ de U (y cualquier valoración v en M) se cumple $M \vDash \beta[v_x^{[t]}]$ syss para todo designador t de $\mathcal{L}$ (y cualquier valoración) $M \vDash \beta[v_x^{M(t)[v]}]$ syss (por 1.12) para todo designador t de $\mathcal{L}$ se cumple $M \vDash \mathtt{S}_x^t\beta$.

Notemos que las sentencias $\mathtt{S}^t_x\beta$ —aunque no tienen necesariamente menor longitud que α— tienen menos signos lógicos, luego podemos aplicarles la hipótesis de inducción: $M \vDash \alpha$ syss $S^t_x\beta$ está en Γ para todo designador t de $\mathcal{L}$ syss (por el teorema 4.8) $\alpha \equiv \bigwedge x\beta$ está en Γ. ∎

Ahora sólo nos queda el problema técnico de ir eliminando hipótesis en el teorema anterior. En primer lugar nos ocupamos de la restricción sobre los descriptores:

Teorema 4.14 *Sea $\mathcal{L}$ un lenguaje formal (con o sin descriptor) y Γ un conjunto consistente de sentencias de $\mathcal{L}$. Entonces Γ tiene un modelo numerable.*

Demostración: Sea $\mathcal{L}'$ un lenguaje formal que conste de los mismos signos que $\mathcal{L}$ más una nueva constante c. Sea x una variable. Por el teorema 4.11 tenemos que $\Gamma \cup \{c = x|(x = x)\}$ es consistente.

Por el teorema 3.32, para cada sentencia γ de Γ existe una sentencia γ' de $\mathcal{L}'$ sin descriptores y tal que

$$c = x|(x = x) \vdash \gamma \leftrightarrow \gamma'.$$

Sea Δ el conjunto de todas estas sentencias γ'. Toda sentencia de Δ es consecuencia de $\Gamma \cup \{c = x|(x = x)\}$. Por el teorema 3.5 tenemos que Δ es consistente. Por el teorema anterior Δ tiene un modelo M' de universo numerable. Como las sentencias de Δ no tienen descriptores, si cambiamos la descripción impropia en M' éste no deja de ser un modelo de Δ. Tomamos concretamente $M'(c)$ como descripción impropia, de modo que ahora se cumple $M' \vDash c = x|(x = x)$.

Así pues, $M' \vDash \Delta \cup \{c = x|(x = x)\}$ y, como toda fórmula de Γ es consecuencia de $\Delta \cup \{c = x|(x = x)\}$, por el teorema de corrección tenemos que $M' \vDash \Gamma$. Finalmente, sea M el modelo de $\mathcal{L}$ que se diferencia de M' en que no interpreta la constante c. Claramente M es numerable y $M \vDash \Gamma$. ∎

Finalmente eliminamos la restricción de que las fórmulas sean sentencias, con lo que tenemos 4.1 y, combinándolo con 3.4, tenemos el teorema siguiente:

Teorema 4.15 *Un conjunto Γ de fórmulas de un lenguaje formal $\mathcal{L}$ es consistente si y sólo si tiene un modelo (que podemos tomar numerable).*

Demostración: Supongamos que Γ es consistente. Sea Γ^c el conjunto de las clausuras universales de las fórmulas de Γ. Como todas las sentencias de Γ^c se deducen de las de Γ, el teorema 3.5 nos da que Γ^c es consistente. Por el teorema anterior Γ^c tiene un modelo numerable M que, claramente, también es un modelo de Γ. El recíproco es 3.4. ∎

4.3 Consecuencias del teorema de completitud

Del teorema 4.15 se sigue que el cálculo deductivo que hemos introducido en principio de forma arbitraria es exactamente lo que tiene que ser:

Teorema 4.16 (Teorema de adecuación) *Toda fórmula lógicamente válida es un teorema lógico.*

DEMOSTRACIÓN: Sea α una fórmula y α^c su clausura universal. Si no $\vdash \alpha$, entonces no $\vdash \alpha^c$, luego no $\vdash \neg\neg\alpha^c$. Por el teorema 3.6 obtenemos que $\neg\alpha^c$ es consistente, luego tiene un modelo M, es decir, $M \models \neg\alpha^c$, luego no $M \models \alpha^c$, luego no $M \models \alpha$, es decir, α no es lógicamente válida. ∎

Nota Es crucial comprender que el enunciado de este teorema carecería de un significado metamatemático preciso si no fuera por la propia demostración del teorema. Recordemos que no sabemos dar un sentido a una afirmación del tipo "α es lógicamente válida" salvo en el caso en que dispongamos de un argumento que nos convenza de que α ha de ser verdadera en cualquier modelo. El teorema de corrección nos da que esto sucede siempre que α es un teorema lógico. Lo que ahora hemos probado es que si α no es un teorema lógico (y está claro qué significa esto) entonces sabemos construir un modelo en el que α es falsa, y esto podemos expresarlo diciendo que α no es lógicamente válida. En resumen, ahora sabemos que toda fórmula α ha de encontrarse en uno de los dos casos siguientes: o bien es un teorema lógico y entonces es verdadera en cualquier modelo, o bien no es un teorema lógico y entonces sabemos describir un modelo concreto en el cual es falsa. Esto nos permite considerar como equivalentes las afirmaciones $\vdash \alpha$ y $\models \alpha$, con lo que, dado que la primera tiene un significado preciso, lo mismo podemos decir, a partir de ahora, de la segunda. ∎

La adecuación del cálculo deductivo queda plasmada más claramente en lo que propiamente se conoce como el *teorema de completitud semántica para la lógica de primer orden:*

Teorema 4.17 (Teorema de completitud semántica (de Gödel)) *Sea T una teoría axiomática consistente y sea α una fórmula de su lenguaje formal. Si α es verdadera en todo modelo (numerable) de T, entonces $\underset{T}{\vdash} \alpha$.*

DEMOSTRACIÓN: Sea Γ el conjunto de los axiomas de T. Hemos de probar que $\Gamma \vdash \alpha$. En caso contrario no $\Gamma \vdash \alpha^c$, luego no $\Gamma \vdash \neg\neg\alpha^c$. Por el teorema 3.6 tenemos que $\Gamma \cup \{\neg\alpha^c\}$ es consistente, luego tiene un modelo numerable M. Como $M \models \Gamma$ se cumple que M es un modelo de T, pero no $M \models \alpha$, en contra de lo supuesto. ∎

Así pues, si el teorema de corrección garantizaba que el cálculo deductivo jamás nos lleva de premisas verdaderas a conclusiones falsas, el teorema de completitud semántica nos garantiza que el cálculo deductivo es completo, no en el sentido sintáctico de que nos responda afirmativa o negativamente a cualquier pregunta, sino en el sentido semántico de que cualquier otro cálculo deductivo "más generoso" que permitiera deducir más consecuencias que el nuestro de unas premisas dadas, necesariamente nos permitiría deducir consecuencias falsas en un modelo a partir de premisas verdaderas en él, por lo que no sería semánticamente aceptable. En resumen, ahora sabemos que nuestro cálculo deductivo

se corresponde exactamente con la noción metamatemática de razonamiento lógico. Cuando un matemático se convence de que si unos objetos cumplen sus hipótesis Γ tienen que cumplir también su conclusión α, está demostrando que $\Gamma \vDash \alpha$, y el teorema de completitud nos da que $\Gamma \vdash \alpha$, es decir, que su razonamiento "podría desarrollarse"[3] hasta convertirse en una demostración en $K_{\mathcal{L}}$.

Aritmética no estándar Si el teorema de completitud nos ha mostrado que el cálculo deductivo es exactamente lo que tiene que ser, a la vez nos muestra ciertas limitaciones que, por esta misma razón, resultan ser esenciales a toda posible formalización y axiomatización de una teoría matemática.

Observemos que si un conjunto de fórmulas Γ tiene la propiedad de que todos sus subconjuntos finitos son consistentes, entonces es consistente. En efecto, si a partir de Γ se dedujera una contradicción, en la deducción sólo podría aparecer una cantidad finita de premisas, las cuales formarían un subconjunto finito y contradictorio de Γ, en contra de lo supuesto. El teorema de completitud traduce este hecho obvio en un hecho nada trivial:

Teorema 4.18 (Teorema de compacidad de Gödel) *Un conjunto de fórmulas tiene un modelo si y sólo si lo tiene cada uno de sus subconjuntos finitos.*

Lo importante en este teorema es que ninguno de los modelos de ninguno de los subconjuntos finitos tiene por qué ser un modelo de la totalidad de las fórmulas y, pese a ello, podemos garantizar que existe un modelo que cumple simultáneamente todas ellas.

De aquí se deduce que la lógica de primer orden no es *categórica,* es decir, que —en la mayoría de casos de interés— es imposible caracterizar unívocamente unos objetos que pretendamos estudiar a través de una colección de axiomas. Concretamente vamos a probarlo con las nociones de "finitud" y de "número natural".

Nosotros hemos presentado los números naturales como los objetos 0, 1, 2, 3, 4, 5, etc., es decir, los objetos generados por un proceso de cómputo perfectamente determinado que nos permite continuar indefinidamente y sin vacilación la sucesión anterior. Así aprenden todos los niños lo que son los números naturales y esta definición les basta para manejarlos en todos los contextos distintos del de la matemática formal. Muchos matemáticos piensan que esta noción "intuitiva", en el más despectivo sentido de la palabra, puede ser suficiente para usos no sofisticados, como contar monedas, o sellos, o piedras, pero no para las matemáticas serias, donde es necesaria una definición más precisa y rigurosa de número natural. Ahora vamos a probar que esto, aunque tiene algo de cierto, también tiene mucho de falso. Es verdad que la matemática, desde el momento en que pretende estudiar objetos abstractos que involucran la noción general de conjunto, requiere ser axiomatizada en su totalidad, lo cual incluye el tratar

[3]Supuesto que todo el razonamiento conste de afirmaciones formalizables en el lenguaje formal que esté considerando. Obviamente, si pasa a hablar de cosas no expresables en $\mathcal{L}$, su razonamiento no será formalizable en $K_{\mathcal{L}}$, pero no por limitaciones de la lógica de $K_{\mathcal{L}}$, sino del lenguaje formal elegido.

axiomáticamente los números naturales. Sin embargo, no es cierto que una presentación axiomática de los números naturales sea más precisa y rigurosa que una presentación no axiomática como la que hemos dado aquí. Al contrario, vamos a demostrar que una presentación axiomática de los números naturales será rigurosa, pero nunca precisa, en el sentido de que será necesariamente ambigua.

En efecto, supongamos que el lector cree que puede definir con total precisión los números naturales en el seno de una teoría axiomática formal. El intento más simple es la aritmética de Peano que ya hemos presentado, pero si el lector considera que es demasiado débil, podemos admitir cualquier otra teoría. Por ejemplo podríamos pensar en una teoría axiomática de conjuntos. No importa cuáles sean sus axiomas en concreto. El argumento que vamos a emplear se aplica a cualquier teoría axiomática T que cumpla los siguientes requisitos mínimos:

- T es consistente (es obvio que una teoría contradictoria no sería una buena forma de presentar los números naturales).
- El lenguaje de T contiene un designador 0, un término x' con x como única variable libre y una fórmula $x \in \mathbb{N}$ con x como única variable libre de modo que en T puedan demostrarse los teoremas siguientes:
  1. $0 \in \mathbb{N}$,
  2. $\bigwedge x \in \mathbb{N}\ x' \in \mathbb{N}$,
  3. $\bigwedge x \in \mathbb{N}\ x' \neq 0$,
  4. $\bigwedge xy \in \mathbb{N}(x' = y' \to x = y)$.

En otras palabras, admitimos que en la teoría T se definan los números naturales, el cero y la operación "siguiente" como se considere oportuno, con tal de que se puedan demostrar las cuatro propiedades elementales que hemos exigido.

Si es posible determinar axiomáticamente los números naturales, la forma de hacerlo será, sin duda, una teoría T que cumpla los requisitos anteriores. Ahora probaremos que existen unos objetos que satisfacen la definición de número natural que ha propuesto el lector —cualquiera que ésta sea— y que, pese a ello, nadie en su sano juicio los aceptaría como números naturales. Más precisamente, vamos a construir un modelo de T en el que existen objetos que satisfacen la definición de número natural del lector y que son distintos de lo que el lector ha decidido llamar 0, y de lo que el lector ha decidido llamar 1, etc.

Sea $\mathcal{L}$ el lenguaje de T y sea $\mathcal{L}'$ el lenguaje que resulta de añadir a $\mathcal{L}$ una constante c. Consideramos la teoría T' sobre $\mathcal{L}'$ que resulta de añadir a los axiomas de T la siguiente colección de sentencias:

$$c \in \mathbb{N}, \quad c \neq 0, \quad c \neq 0', \quad c \neq 0'', \quad c \neq 0''', \quad c \neq 0'''', \quad \ldots$$

La teoría T' es consistente. En virtud del teorema de compacidad basta encontrar un modelo de cada colección finita de axiomas de T. De hecho, es

claro que basta encontrar un modelo de cada teoría T'_n formada por los axiomas de T más los axiomas $c \in \mathbb{N}$, $c \neq 0$, ..., $c \neq 0^{(n)}$, donde $0^{(n)}$ representa a un 0 seguido de n veces el operador "siguiente".

Por hipótesis T es consistente, luego tiene un modelo M. Llamaremos M_n al modelo de $\mathcal{L}'$ que es igual que M salvo por que interpreta la constante c como el objeto denotado por $0^{(n+1)}$. Así, M_n es un modelo de T en el que además es verdadera la sentencia $c = 0^{(n+1)}$. De esta sentencia más las sentencias 1), 2), 3), 4), que estamos suponiendo que son teoremas de T, se deducen las sentencias $c \in \mathbb{N}$, $c \neq 0$, ..., $c \neq 0^{(n)}$, es decir, M_n es un modelo de T'_n.

Por el teorema de compacidad T' tiene un modelo M'. En particular M' es un modelo de T, es decir, sus objetos cumplen todos los axiomas que el lector ha considerado razonables. Más aún, los objetos a de M' que cumplen $M' \vDash x \in \mathbb{N}\,[v^a_x]$, para una valoración cualquiera v, satisfacen todos los requisitos que el lector ha tenido bien exigir a unos objetos para que merezcan el calificativo de números naturales.

Llamemos $\xi = M'(c)$. Puesto que $M' \vDash c \in \mathbb{N}$, tenemos que ξ es uno de esos objetos que el lector está dispuesto a aceptar como números naturales. Ahora bien, como $M' \vDash c \neq 0$, tenemos que ξ es distinto del objeto denotado por el designador 0, es decir, es distinto del objeto que satisface todo lo que el lector ha tenido a bien exigir para que merezca el calificativo de "número natural cero". Similarmente, como $M' \vDash c \neq 0'$, tenemos que ξ es distinto de lo que el lector ha tenido a bien llamar 1, etc. En resumen, la definición de número natural propuesta por el lector es satisfecha por unos objetos entre los cuales hay uno ξ que no es lo que el lector ha llamado 0, ni 1, ni 2, ni 3, ni, en general, ningún número que pueda obtenerse a partir del 0 por un número finito de aplicaciones de la operación "siguiente". Cualquier niño de 10 años al que se le explique esto adecuadamente comprenderá que el lector se equivoca si cree haber definido correctamente los números naturales.

En general, diremos que un modelo M de una teoría que satisface los requisitos que hemos exigido a T es un modelo *no estándar* de la aritmética si en su universo hay un objeto ξ tal que, para una valoración v cualquiera, $M \vDash x \in \mathbb{N}\,[v^\xi_x]$ y para todo número natural n se cumple $M \vDash x \neq 0^{(n)}[v^\xi_x]$. A tales objetos ξ los llamaremos *números no estándar* del modelo M.

Hemos probado que cualquier formalización mínimamente razonable de la aritmética tiene modelos no estándar, modelos en los que hay "números naturales" que no pueden obtenerse a partir del cero en un número finito de pasos. Vemos así que el razonamiento metamatemático que estamos empleando desde el primer capítulo, aunque inútil para tratar con la matemática abstracta, es mucho más preciso que el razonamiento axiomático formal a la hora de tratar con objetos intuitivamente precisos. Así, aunque la noción de finitud es totalmente precisa y rigurosa, tan simple que hasta un niño de 10 años comprende sin dificultad que hay un número finito de dedos en la mano pero hay infinitos números naturales, resulta que el más sofisticado aparato matemático es incapaz de caracterizarla con precisión.

En efecto, nosotros nunca hemos dado una definición de finitud, pues si el lector no supiera perfectamente lo que es ser finito debería entretenerse leyendo libros más elementales que éste. Ahora bien, el lector no sólo debe ser consciente de que ya sabe lo que es ser finito, sino que además debe comprender que no estamos "siendo poco rigurosos" al eludir una definición formal de finitud, ya que no se puede pecar de poco riguroso por no hacer algo imposible. Supongamos que el lector se siente capaz de corregirnos y enunciar una definición razonable de "conjunto finito". Sin duda, para ello deberá hacer uso de algunas propiedades elementales de los conjuntos. Todo cuanto utilice podrá enunciarse explícitamente como los axiomas de una teoría de conjuntos T. No importa cuál sea la teoría T. Pongamos que el lector construye el lenguaje formal que considere oportuno y en él enuncia unos axiomas que digan "los conjuntos cumplen esto y lo otro". Sólo exigimos las siguientes condiciones mínimas:

- T es consistente.
- En T pueden definirse los números naturales en las mismas condiciones que antes, y además ha de poder demostrarse un principio de inducción similar al esquema axiomático de la aritmética de Peano. También ha de ser posible definir la relación de orden en los números naturales y demostrar sus propiedades básicas.
- En T puede definirse una fórmula "x es un conjunto finito" con la cual pueda probarse que todo conjunto con un elemento es finito y que si a un conjunto finito le añadimos un elemento obtenemos un conjunto finito. Por lo demás, el lector es libre de exigir cuanto estime oportuno a esta definición para que sea todo lo exacta que considere posible.
- En T tiene que poder demostrarse que para cada número natural x existe el conjunto de los números naturales menores o iguales que x.

Si se cumplen estos requisitos, en la teoría T puede probarse que el conjunto de los números naturales menores o iguales que un número n es finito, es decir, que satisface la definición de finitud que ha decidido adoptar el lector. Ahora bien, la teoría T tiene un modelo M no estándar, en el cual podemos considerar el conjunto Ξ de todos los números naturales menores o iguales que un número no estándar fijo ξ. Este conjunto Ξ satisface, pues, la definición de finitud del lector, pero contiene al objeto que en M satisface la definición de 0 (ya que ξ no es 0 y en T ha de poder probarse que todo número distinto de 0 es mayor que 0), y también contiene a lo que el lector ha llamado 1 (ya que ξ es distinto de 1 y en T ha de poder probarse que todo número mayor que 0 y distinto de 1 es mayor que 1) y ha de contener a lo que el lector ha llamado 2, y 3, y 4, etc. En definitiva, tenemos un conjunto infinito que satisface la definición de finitud que haya propuesto el lector, cualquiera que ésta sea.

Las nociones de finitud y de número natural están íntimamente relacionadas: si tuviéramos una definición formal precisa de finitud podríamos definir los números naturales definiendo el 0 y la operación siguiente y estipulando que

ésta ha de aplicarse un número finito de veces para obtener cada número natural; recíprocamente, si tuviéramos una definición formal precisa de los números naturales podríamos definir a partir de ella la noción de finitud; pero sucede que no existe ni lo uno ni lo otro, lo cual a su vez no es obstáculo para que cualquier niño de 10 años —al igual que el lector— tenga una noción precisa (intuitiva, no axiomática) de lo que es la finitud y de lo que son los números naturales.

Por otra parte, el lector debe tener presente que todos los teoremas de la aritmética de Peano, o de otra teoría similar, son afirmaciones verdaderas sobre los números naturales. Lo que hemos probado es que también son afirmaciones verdaderas sobre otros objetos que no son los números naturales, pero esto no contradice a lo primero, que es lo que realmente importa. Más en general, una teoría axiomática con axiomas razonables nos permite probar cosas razonables, independientemente que que pueda aplicarse también a objetos no razonables.

Aquí el lector se encuentra nuevamente ante un dilema: o concede que el tratamiento metamatemático que estamos dando a los números naturales es legítimo, o concluye que todo lo dicho en este apartado es, no ya falso, sino un completo sinsentido. Por supuesto, los números naturales son simplemente el ejemplo más sencillo. Lo mismo se puede decir de cualquier concepto de naturaleza "numerable", como puedan ser los números enteros y racionales, las sucesiones finitas de números racionales, los polinomios con coeficientes racionales, los números algebraicos, los grupos finitos, etc. En teoría es posible trabajar metamatemáticamente con todos estos conceptos, aunque en muchos casos puede ser delicado y requerir una extrema atención para no caer en palabras sin significado. Nadie dice que convenga hacerlo, pues la alternativa de trabajar en una teoría axiomática es mucho más ventajosa, pero lo cierto es que es posible. Nosotros sólo trataremos con los estrictamente imprescindibles para estudiar la lógica matemática, donde el uso de una teoría axiomática nos llevaría a un círculo vicioso.

La paradoja de Skolem Veamos ahora una última consecuencia del teorema de completitud del que a su vez se siguen implicaciones muy profundas sobre la naturaleza del razonamiento matemático. En realidad no es nada que no sepamos ya: se trata de enfatizar la numerabilidad de los modelos que sabemos construir. Según el teorema 4.15, una colección de fórmulas tiene un modelo si y sólo si tiene un modelo numerable. Equivalentemente:

Teorema 4.19 (Teorema de Löwenheim-Skolem) *Una teoría axiomática tiene un modelo si y sólo si tiene un modelo numerable.*

En definitiva, este teorema garantiza que no perdemos generalidad si trabajamos sólo con modelos numerables, los únicos que en realidad sabemos entender metamatemáticamente. Lo sorprendente de este resultado estriba en que los matemáticos están convencidos de que en sus teorías tratan con conjuntos no numerables. Vamos a explicar aquí cómo llegan a esa conclusión y cómo encaja eso con el teorema anterior.

Cantor fue el primer matemático en afirmar que hay distintos tamaños posibles de conjuntos infinitos. Su primer resultado a este respecto fue su demostración de que el conjunto $\mathbb{R}$ de los números reales tiene un tamaño mayor que el conjunto $\mathbb{N}$ de los números naturales. Esto significa, más precisamente que, aunque puede definirse una aplicación inyectiva $f : \mathbb{N} \longrightarrow \mathbb{R}$ (por ejemplo, la inclusión), no hay una aplicación biyectiva en esas condiciones, es decir, no es posible "contar" los números reales ni siquiera usando para ello todos los números naturales.

Posteriormente Cantor encontró un argumento mucho más simple y conceptual para demostrar la no numerabilidad de otro conjunto, el conjunto $\mathcal{P}\mathbb{N}$ de todos los subconjuntos de $\mathbb{N}$. Lo hemos discutido superficialmente en la introducción, pero ahora podemos demostrarlo formalmente en Z^* más un axioma adicional, el *axioma de partes,* que afirma que, para todo conjunto X, existe un (obviamente único) conjunto, que representamos por $\mathcal{P}X$, tal que

$$\bigwedge u(u \in \mathcal{P}X \leftrightarrow u \subset X).$$

Teorema de Cantor *Si X es un conjunto, existe una aplicación inyectiva $f : X \longrightarrow \mathcal{P}X$, pero no existen aplicaciones suprayectivas (ni en particular biyectivas) entre ambos conjuntos.*

Demostración: Como aplicación inyectiva basta tomar la definida mediante $f(u) = \{u\}$. Supongamos ahora que existiera una aplicación f suprayectiva. Entonces podríamos definir el conjunto $A = \{u \in X \mid u \notin f(u)\} \in \mathcal{P}X$.

Como f es suprayectiva existe un $u \in X$ tal que $f(u) = A$, pero esto nos lleva a una contradicción, porque, o bien $u \in A$ o bien $u \notin A$, pero si $u \in A = f(u)$, debería ser $u \notin A$, por definición de A, mientras que si $u \notin A = f(u)$, debería ser $u \in A$. Concluimos que f, sea la aplicación que sea, no puede ser suprayectiva. ∎

El hecho de que no podamos emparejar cada elemento de X con un elemento de $\mathcal{P}X$ se interpreta como que el conjunto $\mathcal{P}X$ tiene más elementos que el conjunto X. En el caso en que X es un conjunto finito, digamos con n elementos, es fácil ver[4] que $\mathcal{P}X$ tiene 2^n elementos y, en efecto, $n < 2^n$, pero el teorema de Cantor no depende para nada de que el conjunto X sea finito. En la modesta teoría Z^* + AP no puede demostrarse que haya conjuntos infinitos, pero podemos añadir como axioma que la clase ω es un conjunto, es decir, que existe un conjunto ω (que se representa más habitualmente con la letra $\mathbb{N}$) cuyos elementos son los números naturales. Las teorías de conjuntos con las que trabajan usualmente los matemáticos son mucho más potentes que Z^* y, sean cuales sean sus axiomas, lo cierto es que permiten demostrar la existencia de conjuntos de partes, el teorema de Cantor, y la existencia del conjunto $\mathbb{N}$ de los números naturales. En cualquiera de estas teorías tenemos, pues, el conjunto $\mathcal{P}\mathbb{N}$ de todos los subconjuntos de $\mathbb{N}$. Se trata obviamente de un conjunto infinito, pero el teorema de Cantor implica que no puede numerarse, sus elementos no pueden disponerse en forma de sucesión $A_0, A_1, A_2, A_3, \ldots$

[4]Todo esto puede demostrarse en Z^* + AP, aunque no hemos desarrollado la teoría hasta el punto de que sea obvio cómo hacerlo. Lo veremos más adelante.

Como decíamos, los matemáticos interpretan esto como que $\mathcal{P}\mathbb{N}$ (al igual que muchos otros conjuntos infinitos, como el conjunto de los números reales, etc.) es un conjunto *no numerable,* un conjunto con infinitos elementos, pero *más elementos* que el conjunto $\mathbb{N}$, también infinito.

Sea T cualquier teoría de conjuntos que nos lleve hasta esta conclusión y supongamos que es consistente (en caso contrario deberíamos buscar otra teoría). Entonces T tiene un modelo numerable M. Digamos que los elementos de su universo son

$$c_0, \quad c_1, \quad c_2, \quad c_3, \quad c_4, \quad c_5, \quad \ldots$$

Estos objetos, con las relaciones y funciones adecuadas, satisfacen todos los axiomas y teoremas de la teoría de conjuntos, por lo que podemos llamarlos "conjuntos". A lo largo de este apartado, la palabra "conjunto" se referirá a los objetos c_n y a nada más. Retocando la enumeración si es preciso, podemos suponer que $c_0 = M(\mathbb{N})$, es decir, c_0 es el único objeto que satisface la definición de "conjunto de los números naturales". Así mismo podemos suponer que su extensión la forman los conjuntos c_{2^n}, para $n \geq 1$. Concretamente, c_2 es el conjunto que satisface la definición de número natural 0, c_4 el que satisface la definición de número natural 1, c_8 la de 2, etc.

También estamos afirmando que $M(\in)(c_n, c_0)$ si y sólo si $n = 2^{k+1}$ para algún k. Con esto estamos suponiendo tácitamente que M es un modelo estándar, es decir, que no tiene números naturales infinitos. No tendría por qué ser así, pero vamos a suponerlo por simplicidad.

No perdemos generalidad si suponemos que $M(\mathcal{P}\mathbb{N}) = c_1$, es decir, que c_1 es el único conjunto que tiene por elementos exactamente a todos los subconjuntos de $\mathbb{N}$, (de c_0). Así mismo podemos suponer que los elementos de c_1 son los conjuntos de la forma c_{3^n}, para $n \geq 1$. Así, c_3 podría ser el conjunto de los números pares, c_9 el conjunto de los números primos, c_{27} el conjunto vacío, c_{81} el conjunto de los números menores que 1000, etc.[5]

Más concretamente, estamos suponiendo que si un conjunto c_n es un subconjunto de c_0, es decir, si todo c_i que cumpla $M(\in)(c_i, c_n)$ cumple también $M(\in)(c_i, c_0)$, entonces $n = 3^{k+1}$, así como que $M(\in)(c_n, c_1)$ si y sólo si se cumple $n = 3^{k+1}$.

La llamada paradoja de Skolem consiste en que este modelo que estamos describiendo existe realmente, y ello no contradice el hecho de que $\mathcal{P}\mathbb{N}$, es decir, el conjunto cuyos elementos son c_3, c_9, c_{27}, etc. es un conjunto no numerable: no es posible biyectar sus elementos con los números naturales.

Quien crea ver una contradicción en todo esto necesita aclararse algunas ideas confusas. Por ejemplo, una presunta contradicción que probara que en este modelo $\mathcal{P}\mathbb{N}$ sí que es numerable sería considerar la "biyección" $n \mapsto c_{3^{n+1}}$. Pero esto no es correcto. La sentencia de T que afirma que $\mathcal{P}\mathbb{N}$ no es numerable

[5]Así suponemos que ningún número natural está contenido en $\mathbb{N}$. De acuerdo con la construcción más habitual del conjunto de los números naturales como ordinales sucede justo lo contrario: todo número natural es un subconjunto de $\mathbb{N}$, pero es posible modificar la definición de los números naturales para que esto no suceda (por ejemplo, cambiando n por $(n, 0)$) y hemos preferido evitar las confusiones que podría producir este tecnicismo.

se interpreta en M como que no existe ninguna biyección entre los conjuntos que en M satisfacen la definición de número natural y los conjuntos que en M satisfacen la definición de subconjunto de $\mathbb{N}$. En nuestro caso, lo que tendríamos que encontrar es una biyección entre los elementos de c_0 y los elementos de c_1, es decir, entre los conjuntos c_2, c_4, c_8, etc. y los conjuntos c_3, c_9, c_{27}, etc.

Quizá el lector ingenuo aún crea ver una biyección en estas condiciones, a saber, la dada por $c_{2^k} \mapsto c_{3^k}$. Pero esto tampoco es una biyección. Una biyección entre dos conjuntos es un conjunto que satisface la definición de biyección: un conjunto de pares ordenados cuyas primeras componentes estén en el primer conjunto, sus segundas componentes en el segundo conjunto y de modo que cada elemento del primer conjunto está emparejado con un único elemento del segundo y viceversa. Tratemos de conseguir eso. Ante todo, en Z y en cualquier otra teoría de conjuntos razonable podemos demostrar que, dados dos conjuntos x e y, existe un único conjunto z tal que $z = (x, y)$, es decir, z es el par ordenado formado por x e y en este orden. Este teorema tiene que cumplirse en nuestro modelo M. Si lo aplicamos a los conjuntos $c_{2^{k+1}}$ y $c_{3^{k+1}}$, concluimos que tiene que haber otro conjunto, y reordenando los índices podemos suponer que es $c_{5^{k+1}}$, tal que $M \vDash z = (x, y)[v]$, donde v es cualquier valoración que cumpla $v(x) = c_{2^{k+1}}$, $v(y) = c_{3^{k+1}}$ y $v(z) = c_{5^{k+1}}$.

Así, el "conjunto" formado por los conjuntos c_5, c_{25}, c_{125}, ... *sería* una biyección entre c_0 y c_1, es decir, entre $\mathbb{N}$ y $\mathcal{P}\mathbb{N}$. Lo sería... si fuera un conjunto.

Estamos al borde de la contradicción, pero no vamos a llegar a ella. Tendríamos una contradicción si la *colección* de conjuntos c_5, c_{25}, c_{125}, ... fuera la extensión de un conjunto, es decir, si existiera un conjunto, digamos c_r, cuyos elementos fueran exactamente los conjuntos $c_{5^{k+1}}$, es decir, si para algún r se cumpliera que $M(\in)(c_n, c_r)$ si y sólo si $n = 5^{k+1}$. En tal caso c_r sí que sería una biyección entre $\mathbb{N}$ y $\mathcal{P}\mathbb{N}$ y en M sería falso el teorema que afirma la no numerabilidad de $\mathcal{P}\mathbb{N}$. Pero es que no tenemos nada que justifique ha de existir tal conjunto c_r. Justo al contrario, como sabemos que M es un modelo de la teoría de conjuntos T, podemos asegurar que tal c_r no puede existir.

Más aún, notemos que la colección formada por los conjuntos $c_{5^{k+1}}$ no sólo no puede ser la extensión de un conjunto de M, sino que tampoco puede ser la extensión de una clase propia en M, es decir, no existe ninguna fórmula $\phi(x)$ del lenguaje $\mathcal{L}_{\mathrm{tc}}$ tal que los conjuntos a que cumplen $M \vDash \phi[v_x^a]$ sean precisamente los $c_{5^{k+1}}$. Si existiera, sería una subclase del conjunto[6] $\mathbb{N} \times \mathcal{P}\mathbb{N}$, luego sería un conjunto. Los matemáticos "ven" las clases propias en el sentido de que, aunque no pueden "encerrar" todos sus elementos en un conjunto, pueden hablar de ellas (nada les impide considerar la clase Ω de todos los ordinales, y definir una aplicación $F : \Omega \longrightarrow \Omega$, etc.). En cambio, la colección de los pares ordenados que, en un modelo, biyecta $\mathbb{N}$ como $\mathcal{P}\mathbb{N}$ es una colección "invisible" para los matemáticos, en el sentido de que ni siquiera pueden hacer referencia a sus elementos mediante una fórmula. Es una colección "no definible" en el modelo.

[6] En Z^* + AP se puede demostrar que el producto cartesiano de conjuntos es un conjunto, así como que toda subclase de un conjunto es un conjunto.

Otro ejemplo de esta situación lo proporcionan los modelos no estándar. Si M es un modelo no estándar de la teoría de conjuntos, podemos considerar la colección de todos los conjuntos que satisfacen la definición de número natural pero que son números no estándar, es decir, que no pueden obtenerse a partir del conjunto que satisface la definición de 0 aplicando un número finito de veces la definición de "siguiente". Tal colección no puede ser la extensión de ningún conjunto, pues un teorema elemental afirma que todo número natural no nulo tiene un inmediato anterior y, como los números no estándar son todos no nulos, resulta que todo número no estándar tiene un anterior. Así, si la colección de los números no estándar fuera la extensión de un conjunto, en M existiría un subconjunto no vacío de $\mathbb{N}$ sin un mínimo elemento, cuando hasta en la teoría básica T se demuestra que eso es imposible. Si en la teoría tenemos que $\mathbb{N}$ es un conjunto y que las subclases de los conjuntos son conjuntos, la colección de los números naturales no estándar tampoco puede ser una clase propia.

Esto no significa que los números no estándar no pertenezcan a ningún conjunto. Al contrario, dado un número natural, x, siempre podemos considerar el conjunto $\{n \in \mathbb{N} \mid n \leq x\}$. Si x no es estándar, entonces esta expresión determina un conjunto (un objeto de M) que satisface la definición de conjunto finito y cuya extensión es —pese a ello— infinita, pues contiene, entre otros, a todos los números naturales estándar.

En resumen, al rastrear hasta su base la paradoja de Skolem encontramos que surge de una confusión: la confusión entre una colección (metamatemática) de conjuntos, como es c_5, c_{25}, c_{125}, ... y un conjunto (matemático), es decir, un objeto de un modelo de la teoría de conjuntos.

4.4 Consideraciones finales

El lector no debe considerar anecdóticos o marginales los ejemplos de la sección anterior. Al contrario, contienen una parte importante de los hechos más profundos que vamos a estudiar en este libro. Probablemente, los irá asimilando cada vez mejor a medida que avancemos, pero para que así sea debería volver a meditar sobre ellos cada vez que encuentre nueva información relevante. La dificultad principal con la que se va a encontrar es que, a diferencia de lo que ocurre en contextos similares estrictamente matemáticos, lo necesario para comprenderlos cabalmente no es un mayor o menor grado de inteligencia, conocimientos o destreza, sino asimilar un determinado esquema conceptual mucho más rico que el que requiere la matemática formal.

Esta última sección pretende ser una ayuda para este fin. Afortunadamente, mientras la física moderna requiere pasar del esquema conceptual clásico a otro mucho más extraño, sutil y todavía no comprendido del todo, el esquema conceptual que requiere la lógica moderna —si bien distinto del que tradicionalmente han adoptado los matemáticos— no es extraño y novedoso, sino uno bien familiar y cotidiano.

Supongamos que hemos visto en el cine una película biográfica sobre Napoleón y al salir discutimos sobre ella. No tendremos ninguna dificultad en usar

correctamente la palabra "Napoleón" a pesar de que tiene tres significados diferentes según el contexto: Napoleón-histórico, Napoleón-personaje y Napoleón-actor. Las afirmaciones sobre el primero son objetivas y semánticamente completas: o bien Napoleón padecía de gota o no padecía, con independencia de que sepamos cuál era el caso. El segundo es una creación del guionista de la película. Debe tener un cierto parecido con el Napoleón-histórico para que merezca el mismo nombre, pero tampoco tiene por qué coincidir con él. Por ejemplo, podría ser que el Napoleón-histórico padeciera de gota y el Napoleón-personaje no, o viceversa. Más aún, podría ocurrir que en la película no se hiciera ninguna alusión a si el Napoleón-personaje padecía o no gota, y en tal caso carece de sentido preguntar si esta afirmación es verdadera o falsa. Una película es sintácticamente incompleta: lo que no se dice explícita o implícitamente en ella no es verdadero ni falso, es indecidible. Por último, un mismo guión puede ser interpretado de forma diferente por actores diferentes. Los actores pueden precisar aspectos de los personajes que no están determinados por el guión. Nadie tiene dificultad en distinguir una crítica al guionista de una película por la mejor o peor caracterización de un personaje con una crítica a un actor por su mejor o peor interpretación del mismo.

Pues bien, afirmamos que el esquema conceptual necesario para interpretar adecuadamente los ejemplos de la sección anterior es exactamente el mismo que el que espontáneamente empleamos al discutir sobre una película. Estrictamente hablando, una demostración formal no es más que una sucesión de signos en un papel, igual que una película no es más que una sucesión de cuadrículas de celuloide coloreado, pero cuando leemos una demostración formal —al igual que cuando vemos una película— no vemos eso. Vemos una historia sobre unos personajes, los cuales a su vez pueden ser réplicas de objetos reales.

Los números naturales metamatemáticos son como el Napoleón-histórico, son objetos de los que podemos hablar objetivamente, que cumplen o no cumplen ciertas propiedades con independencia de que sepamos o no cuál es el caso. Al trabajar metamatemáticamente con ellos estamos investigándolos igual que un historiador puede investigar a Napoleón: reunimos la información que tenemos a nuestro alcance y a partir de ella tratamos de inferir hechos nuevos. Cuando decidimos formalizar la teoría de los números naturales hacemos como el novelista que prepara una novela histórica, o como el guionista de cine: diseñamos un personaje que pretende ser lo más parecido posible al original. La aritmética de Peano es una película sobre los números naturales. Podemos pensar objetivamente en sus protagonistas, es decir, tratarlos como si fueran objetos reales, al igual que podemos pensar objetivamente en Sherlock Holmes o en el pato Donald, pero debemos pensar que sólo son determinaciones parciales.

Notemos que hay tres clases de personajes de película o de novela: los históricos, que se ciñen a las características de un ser real, los personajes históricos novelados, que se basan en un personaje histórico pero han sido distorsionados por el autor (una caricatura de Napoleón, por ejemplo) y los ficticios, como Sherlock Holmes, sin ninguna relación con la realidad. Sin embargo, esta distinción es externa a la propia película, en el sentido de que un espectador que no sepa más que lo que la película le muestra será incapaz de distinguir a qué

tipo pertenece cada personaje. Para hacer la distinción hemos de investigar la realidad y determinar si contiene objetos de características similares a los personajes.

Igualmente, podemos decir que los números naturales-matemáticos (= personajes) que aparecen en la aritmética de Peano son personajes históricos, porque todos los axiomas son afirmaciones verdaderas sobre los números naturales reales. Si extendemos la teoría para formar la aritmética no estándar obtenemos unos personajes históricos-novelados y, por último, una antigua discusión sobre la filosofía de las matemáticas puede enunciarse en estos términos como el dilema de si personajes como el conjunto de los números reales o los cardinales transfinitos son personajes históricos o ficticios. Después volveremos sobre este punto. Lo cierto es que, como meros espectadores, no podemos distinguirlos, pues podemos pensar con la misma objetividad y sentido de la realidad tanto acerca de Don Quijote como de Rodrigo Díaz de Vivar.

El modelo natural de la aritmética de Peano es la película perfecta: la película en la que cada personaje histórico se interpreta a sí mismo. No obstante, hemos visto que el mismo guión puede ser interpretado por actores esperpénticos, que se aprovechan de que el guión no dice explícitamente que no existen números no estándar. Lo peculiar de la situación es que, mientras es fácil exigir en el guión la existencia de naturales no estándar (añadiendo la constante c y los axiomas que dicen que es diferente de cualquier $0^{(n)}$), hemos probado que es imposible escribir un guión que exija la no existencia de números no estándar.

Un modelo numerable de la teoría de conjuntos es una película con efectos especiales. Tanto si queremos hacer una película sobre la llegada del hombre a la Luna (hecho histórico) como si queremos hacerla sobre la llegada del hombre a Júpiter (ciencia-ficción), no podemos permitirnos filmar escenas reales y, en ambos casos, tendremos que recurrir a los efectos especiales. Así pues, sin entrar en la discusión de si existe metamatemáticamente un conjunto no numerable como es $\mathcal{P}\mathbb{N}$, lo cierto es que podemos "simularlo" con efectos especiales.

Un técnico en efectos especiales puede hacer que una pequeña maqueta de plástico parezca una nave espacial, pero si por accidente se viera su mano en la escena, el espectador podría calcular el tamaño real de la "nave", y se daría cuenta de la farsa. Si M es un modelo de la teoría de conjuntos, podemos comparar a las colecciones de elementos de su universo con las personas que realizan la película, y las colecciones que constituyen la extensión de un conjunto con las personas que "se ven" en la pantalla. En el ejemplo de la sección anterior, c_0 es el actor que interpreta el papel de conjunto (= personaje) de los números naturales, mientras que la colección de los conjuntos $c_{5^{k+1}}$ es un técnico en efectos especiales. Está ahí, pero, si se viera en escena, el espectador se daría cuenta de que $\mathcal{P}\mathbb{N}$ es en realidad una pequeña colección numerable, y no el conjunto inmenso que pretende parecer.

Si a un matemático le enseñamos únicamente los "actores" de M, es decir, los conjuntos, las colecciones que aparecen en escena, creerá estar viendo el universo del que hablan todos los libros de matemáticas, con sus conjuntos no numerables incluidos, pero si llegara a ver colecciones como la de los conjuntos $c_{5^{k+1}}$ o la de los números naturales no estándar, si es que los hay, sería como

si el espectador sorprendiera a Napoleón en manos de un maquillador. Estas colecciones "no existen" exactamente en el mismo sentido en que los maquilladores "no existen" a ojos del espectador. Napoleón-actor necesita ser maquillado, Napoleón-personaje no.[7]

En resumen, la mayor dificultad que el lector se encontrará a la hora de interpretar los resultados que hemos visto y vamos a ver, es la de reconocer significados diversos según el contexto en conceptos que para el matemático suelen tener un único significado (p.ej. la terna *colección metamatemática–conjunto matemático como concepto axiomático–conjunto como objeto metamatemático de un modelo concreto).* La única finalidad del juego de analogías que hemos desplegado es la de ayudar al lector a advertir qué distinciones van a ser necesarias y en qué han de consistir. Sin embargo, es importante tener presente que ninguna de estas analogías es un argumento. Todas estas distinciones deben ser entendidas y justificadas directamente sobre los conceptos que estamos tratando: números naturales, conjuntos, signos, etc. Por otra parte, no es menos cierto que —aunque esto no quede justificado sino *a posteriori*— los esquemas conceptuales son idénticos: cualquier problema conceptual sobre la naturaleza de $\mathcal{P}\mathbb{N}$ puede trasladarse a un problema idéntico sobre Sherlock Holmes y viceversa, y esto puede ser de gran ayuda.

[7] En esta comparación, las clases que no son conjuntos equivalen a personajes de los que se habla en la película e intervienen en la trama, pero que nunca aparecen en escena y, por consiguiente, no son encarnados por ningún actor. Es como Tutank-Amon en una película de arqueólogos. Ciertamente, no es lo mismo Tutank-Amon que un maquillador. El matemático puede hablar de los cardinales aunque no vea ningún conjunto que los contenga a todos, pero no puede hablar de una biyección fantasma entre $\mathbb{N}$ y $\mathcal{P}\mathbb{N}$.

Segunda parte

Teorías aritméticas

Capítulo V

La aritmética de Peano

En el capítulo III hemos introducido la aritmética de Peano (AP), que es la más simple de todas las teorías aritméticas, la que nos permite hablar exclusivamente de números naturales, la teoría aritmética en la que "$\bigwedge x$" pretende significar "para todo número natural x", sin necesidad de dar una definición que especifique los objetos que son números naturales de entre una clase mayor de objetos. Ese "pretende significar" significa, precisamente, que AP admite el modelo natural $\mathbb{N}$ cuyo universo es el conjunto de los números naturales y en el que los funtores "siguiente", "suma" y "producto" se interpretan como las funciones usuales. Por consiguiente, todos los teoremas de AP pueden interpretarse como afirmaciones verdaderas sobre los números naturales y, en particular, sabemos que AP es consistente.

En este capítulo empezamos a explorar AP. En primer lugar demostraremos en su seno —incluso en subteorías más simples— los resultados más básicos de la aritmética de los números naturales y, en segundo lugar, demostraremos que en AP es posible formalizar y demostrar resultados conjuntistas que multiplicarán la capacidad expresiva que en un principio cabría esperar de esta teoría.

5.1 La aritmética de Robinson

La aritmética de Peano tiene infinitos axiomas, debido a que el principio de inducción es en realidad un esquema de axioma que da lugar a un axioma (un caso particular) para cada fórmula $\phi(x)$. Cuando más compleja sea $\phi(x)$ más complejo será el axioma correspondiente. Veremos más adelante que resulta útil determinar el grado de complejidad de los axiomas necesarios para demostrar cada resultado, y para determinarlo vamos a definir distintas subteorías de AP que admitan el principio de inducción para una clase de fórmulas de una complejidad determinada. El caso más simple es no admitir ninguna forma del principio de inducción, lo que nos lleva a la teoría siguiente:

Definición 5.1 Recordemos que el lenguaje de la aritmética $\mathcal{L}_a$ es el que tiene por signos eventuales una constante 0, un funtor monádico S (aunque escribimos

$t' \equiv St$) y dos funtores diádicos $+$ y $\cdot$. Llamaremos *Aritmética de Robinson* a la teoría axiomática Q sobre $\mathcal{L}_a$ cuyos axiomas son los siguientes:[1]

$$\begin{array}{ll}
\text{(Q1)} & \bigwedge x\ x' \neq 0 \\
\text{(Q2)} & \bigwedge xy(x' = y' \to x = y) \\
\text{(Q3)} & \bigwedge x(x \neq 0 \to \bigvee y\ x = y') \\
\text{(Q4)} & \bigwedge x\ x + 0 = x \\
\text{(Q5)} & \bigwedge xy\ x + y' = (x + y)' \\
\text{(Q6)} & \bigwedge x\ x \cdot 0 = 0 \\
\text{(Q7)} & \bigwedge xy\ x \cdot y' = xy + x
\end{array}$$

La Aritmética de Peano (AP) resulta de añadir a Q el principio de inducción:

$$\phi(0) \wedge \bigwedge x(\phi(x) \to \phi(x')) \to \bigwedge x\phi(x),$$

para toda fórmula $\phi(x)$, tal vez con más variables libres.

Notemos que el axioma Q3 no figura entre los axiomas de AP. Ello se debe a que se puede demostrar a partir del principio de inducción, por lo que, al añadir éste, Q3 se vuelve redundante. En efecto, basta considerar la fórmula

$$\phi(x) \equiv x = 0 \vee \bigvee y\, x = y'.$$

Teniendo esto en cuenta concluimos que todo teorema de Q es también un teorema de AP.

Observemos ahora que podemos definir $1 \equiv 0'$, y entonces los axiomas Q4 y Q5 implican que

$$x + 1 = x + 0' = (x + 0)' = x'.$$

Por ello, de aquí en adelante escribiremos $x+1$ en lugar de x'. En estos términos, los axiomas de Q se expresan así:

$$\begin{array}{ll}
\text{(Q1)} & \bigwedge x\ x + 1 \neq 0 \\
\text{(Q2)} & \bigwedge xy(x + 1 = y + 1 \to x = y) \\
\text{(Q3)} & \bigwedge x(x \neq 0 \to \bigvee y\ x = y + 1) \\
\text{(Q4)} & \bigwedge x\ x + 0 = x \\
\text{(Q5)} & \bigwedge xy(x + (y + 1) = (x + y) + 1) \\
\text{(Q6)} & \bigwedge x\ x \cdot 0 = 0 \\
\text{(Q7)} & \bigwedge xy\ x(y + 1) = xy + x
\end{array}$$

Definimos $x \leq y \equiv \bigvee z\ z + x = y$.

[1] También consideraremos a $0 = x|(x = x)$ como axioma de Q, lo que introduce en la teoría el convenio que toda descripción impropia (todo concepto mal definido) es, por definición, el 0. Así tenemos asegurado además que toda fórmula es equivalente en Q a otra sin descriptores.

Cuando trabajamos con teorías axiomáticas como Q o AP, que nos permiten hablar sobre los números naturales, hemos de distinguir entre los números naturales metamatemáticos y los números naturales de los que supuestamente estamos hablando a través de la teoría axiomática. Los razonamientos siguientes mostrarán la importancia de este punto.

En $\mathcal{L}_a$ podemos considerar la sucesión de designadores

$$0, \quad 0', \quad 0'', \quad 0''', \quad 0'''', \ldots$$

Conviene representarlos por $0^{(0)}$, $0^{(1)}$, $0^{(2)}$, $0^{(3)}$, $0^{(4)}$, ... de modo que, en general, para cada natural (metamatemático) n, representaremos por $0^{(n)}$ al designador formado por 0 seguido de n aplicaciones del funtor "siguiente". A estos designadores los llamaremos *numerales.*

Notemos que $0 \equiv 0^{(0)}$ y, que también hemos definido $1 \equiv 0^{(1)}$. En general, cuando escribamos fórmulas de $\mathcal{L}_a$ en las que aparezcan explícitamente estos designadores, no habrá ningún inconveniente en escribir, por ejemplo, 7 en lugar de $0^{(7)}$ (que es lo que hacen habitualmente los matemáticos). Sin embargo, la distinción entre un número natural metamatemático n y el numeral $0^{(n)}$ resulta fundamental cuando hablamos, como hacemos aquí mismo, de un caso genérico. Es crucial comprender que en "$0^{(n)}$" la "n" es una variable metamatemática (un pronombre indefinido castellano que se refiere a un número natural arbitrario), pero no es una variable del lenguaje formal de la aritmética. Del mismo modo que en 0, $0'$, $0''$, etc. no hay variables libres, ni aparecerá ninguna variable por más comitas que añadamos, en $0^{(n)}$ no hay ninguna variable libre. Lo que hay es una constante y n funtores, pero ninguna variable. En particular, es un sinsentido escribir

$$\bigwedge mn\ 0^{(m)} + 0^{(n)} = 0^{(n)} + 0^{(m)}.$$

Si tratamos de interpretar "eso", el cuantificador $\bigwedge$ nos obliga a sobrentender —como hemos hecho hasta ahora muchas veces— que "m" y "n" denotan dos variables de $\mathcal{L}_a$, como podrían ser $m \equiv x_5$ y $n \equiv x_8$, pero eso nos obligaría a interpretar el "término" $0^{(x_5)}$, y esto no está definido: sabemos lo que es 0, o 0 con una comita, o 0 con dos comitas, o, en general, 0 con n comitas, donde n es un número de comitas, pero nunca hemos definido 0 con x_5 comitas, donde x_5 no es un número, sino una variable.

Lo que sí tiene sentido es el metateorema siguiente:

Teorema *Para todo par de números naturales m y n, se cumple*

$$\underset{\mathrm{Q}}{\vdash} 0^{(m)} + 0^{(n)} = 0^{(n)} + 0^{(m)}.$$

Esto es un esquema teoremático, que afirma que las infinitas sentencias que se obtienen sustituyendo m y n por números naturales determinados son, todas ellas, teoremas de Q. Esto es consecuencia inmediata del (meta)teorema siguiente:

Teorema 5.2 *Sean m, n números naturales. Las fórmulas siguientes son demostrables en* Q:

1. $0^{(m)} + 0^{(n)} = 0^{(m+n)}$
2. $0^{(m)} \cdot 0^{(n)} = 0^{(mn)}$
3. $0^{(m)} \neq 0^{(n)}$ *(supuesto que $m \neq n$)*
4. $0^{(m)} \leq 0^{(n)}$ *(supuesto que $m \leq n$)*
5. $\neg 0^{(m)} \leq 0^{(n)}$ *(supuesto que $n < m$)*

DEMOSTRACIÓN: 1) por inducción sobre n. Para $n = 0$ se reduce a Q4: $0^{(m)} + 0 = 0^{(m)} = 0^{(m+0)}$. Si se puede demostrar para n, entonces, usando Q5 al final de la primera línea:

$$0^{(m)} + 0^{(n+1)} \equiv 0^{(m)} + 0^{(n)\prime} = 0^{(m)} + (0^{(n)} + 1) = (0^{(m)} + 0^{(n)}) + 1$$
$$= 0^{(m+n)} + 1 = 0^{(m+n+1)}.$$

La prueba de 2) es análoga. Para 3) no perdemos generalidad si suponemos que $m < n$. Razonando en Q, si suponemos $0^{(m)} = 0^{(n)}$, aplicando m veces Q2 llegamos a que $0 = 0^{(n-m-1)} + 1$, en contradicción con Q1.

4) es consecuencia de 1): si $m \leq n$, podemos escribir $r + m = n$, con lo que $0^{(r)} + 0^{(m)} = 0^{(n)}$, y esto implica $0^{(m)} \leq 0^{(n)}$.

La prueba de 5) la posponemos hasta haber probado el teorema 5.4, más abajo. ∎

Nota Observemos que hemos demostrado 1) y 2) por inducción sobre n, y eso es posible a pesar de de que en Q no contemos con el principio de inducción. Estamos probando por inducción que cualquier resultado de tipo $2+3 = 5$ puede probarse en Q (sin inducción).

Lo que dice en esencia el teorema anterior es que los cálculos concretos son formalizables en Q, es decir, que si se cumple que $6(2 + 3) = 30$, entonces esto mismo (interpretado como una sentencia de $\mathcal{L}_a$) es demostrable en Q. Ahora vamos a ver teoremas que relacionan numerales y números genéricos (representados por variables).

Teorema 5.3 *Sea n un número natural. Las fórmulas siguientes son demostrables en* Q:

1. $\bigwedge xy(x + y = 0 \to x = 0 \wedge y = 0)$
2. $\bigwedge xy(xy = 0 \to x = 0 \vee y = 0)$
3. $\bigwedge x\ 0 \leq x$
4. $\bigwedge x(x + 1 \leq 0^{(n+1)} \to x \leq 0^{(n)})$
5. $\bigwedge x((x + 1) + 0^{(n)} = x + 0^{(n+1)})$

DEMOSTRACIÓN: 1) Si $y \neq 0$ existe un z tal que $y = z + 1$, luego

$$x + y = x + (z + 1) = (x + z) + 1 \neq 0.$$

Por lo tanto $y = 0$, y esto implica a su vez que $x = 0$.

2) Si $x \neq 0 \wedge y \neq 0$, entonces $x = u + 1$, $y = v + 1$, luego

$$xy = x(v + 1) = xv + x = xv + (u + 1) = (xv + u) + 1 \neq 0.$$

3) Como $x + 0 = x$, tenemos que $0 \leq x$.

4) Si $x + 1 \leq 0^{(n+1)} = 0^{(n)} + 1$, existe un z tal que $z + (x + 1) = 0^{(n)} + 1$, lo que equivale a $(z + x) + 1 = 0^{(n)} + 1$, luego $z + x = 0^{(n)}$, luego $x \leq 0^{(n)}$.

5) Demostramos

$$\vdash_Q \bigwedge x((x + 1) + 0^{(n)} = x + 0^{(n+1)})$$

por inducción (metamatemática) sobre n. Para $n = 0$ tenemos

$$(x + 1) + 0 = x + 1 = x + 0^{(0+1)}.$$

Si es cierto para n, entonces

$$(x + 1) + 0^{(n)} = x + 0^{(n+1)},$$

luego

$$((x + 1) + 0^{(n)}) + 1 = (x + 0^{(n+1)}) + 1$$

lo que equivale a

$$(x + 1) + (0^{(n)} + 1) = x + (0^{(n+1)} + 1) = x + 0^{((n+1)+1)}.$$

■

Los dos últimos apartados del teorema anterior eran pasos técnicos previos para probar lo siguiente:

Teorema 5.4 *Sea n un número natural. Las fórmulas siguientes son demostrables en* Q:

1. $\bigwedge x(x \leq 0^{(n)} \leftrightarrow x = 0^{(0)} \vee x = 0^{(1)} \vee \cdots \vee x = 0^{(n)})$
2. $\bigwedge x(0^{(n)} \leq x \leftrightarrow 0^{(n)} = x \vee 0^{(n+1)} \leq x)$
3. $\bigwedge x(x \leq 0^{(n)} \vee 0^{(n)} \leq x)$

DEMOSTRACIÓN: 1) La implicación $\Leftarrow$ se reduce a que, como ya hemos visto, si $m \leq n$ entonces podemos probar que $x = 0^{(m)} \leq 0^{(n)}$. Probamos la contraria por inducción sobre n. Para $n = 0$ usamos el apartado a) del teorema anterior: $x \leq 0$ implica que existe un z tal que $z + x = 0$, luego $x = 0$. Supuesto para n, si $x \leq 0^{(n+1)}$, o bien $x = 0 \leq 0^{(n)}$, en cuyo caso ya hemos terminado, o bien $x = y+1$, luego $y+1 \leq 0^{(n+1)}$, luego por el teorema anterior $y \leq 0^{(n)}$, luego por hipótesis de inducción $y = 0 \vee \cdots \vee y = 0^{(n)}$, luego $x = 0^{(1)} \vee \cdots \vee x = 0^{(n+1)}$.

2) Si $0^{(n)} \leq x$, existe un z tal que $z + 0^{(n)} = x$. Si $z = 0$ tenemos que $0 + 0^{(n)} = 0^{(n)}$ por 5.2, luego $0^{(n)} = x$. En caso contrario existe un y tal que $z = y + 1$, luego $(y + 1) + 0^{(n)} = x$ y por el teorema anterior $y + 0^{(n+1)} = x$, luego $0^{(n+1)} \leq x$.

3) Por inducción sobre n. Para $n = 0$ sabemos que $0^{(0)} \leq x$. Supuesto cierto para n, tenemos que $x \leq 0^{(n)}$ o bien $0^{(n)} \leq x$. En el primer caso $x \leq 0^{(n+1)}$ por 1). En el segundo caso aplicamos 2), con lo que, o bien $0^{(n)} = x$, que a su vez implica $x \leq 0^{(n+1)}$, o bien $0^{(n+1)} \leq x$. ∎

Ahora ya podemos probar el apartado e) del teorema 5.2:

Por el teorema 5.4 tenemos que

$$\underset{\mathrm{Q}}{\vdash} 0^{(m)} \leq 0^{(n)} \to 0^{(m)} = 0 \vee \cdots \vee 0^{(m)} = 0^{(n)},$$

de donde

$$\underset{\mathrm{Q}}{\vdash} 0^{(m)} \neq 0 \wedge \cdots \wedge 0^{(m)} \neq 0^{(n)} \to \neg 0^{(m)} \leq 0^{(n)},$$

y si $n < m$ la hipótesis de la implicación se demuestra usando 5.2 3). ∎

Con todo esto podemos probar un resultado general. Para ello necesitamos una definición:

Definición 5.5 Diremos que una fórmula α de $\mathcal{L}_a$ es *(completamente) abierta* (resp. Δ_0) si existe una sucesión de fórmulas $\alpha_0, \ldots, \alpha_m \equiv \alpha$ de forma que cada α_i es de uno de los tipos siguientes:

1. $t_1 = t_2$ o $t_1 \leq t_2$, donde t_1, t_2 son términos sin descriptores (es decir, formados por variables, la constante 0, y los funtores).

2. $\neg\beta$ o $\beta \to \gamma$, donde β y γ son fórmulas anteriores de la sucesión.

3. (sólo para fórmulas Δ_0) $\bigwedge x \leq y\, \beta$ o $\bigvee x \leq y\, \beta$, donde β es una fórmula anterior de la sucesión y la variable y es distinta de x.

Aquí estamos adoptando como convenio las abreviaturas:

$$\bigwedge x \leq y\, \beta \equiv \bigwedge x(x \leq y \to \beta), \qquad \bigvee x \leq y\, \beta \equiv \bigvee x(x \leq y \wedge \beta).$$

Más llanamente, llamaremos fórmulas abiertas a las que no tienen descriptores ni más cuantificadores que los que aparecen en la definición de $\leq$, y las fórmulas Δ_0 son las que además admiten lo que llamaremos cuantificadores acotados.

De este modo, si α y β son fórmulas abiertas (resp. Δ_0) también lo son $\neg\alpha$, $\alpha \to \beta$, $\alpha \vee \beta$, $\alpha \wedge \beta$ y $\alpha \leftrightarrow \beta$ (y en el caso Δ_0 también $\bigwedge x \leq y\, \alpha$ y $\bigvee x \leq y\, \alpha$).

La característica más destacada de Q es la que muestra el teorema siguiente:

Teorema 5.6 (Σ_1-completitud de Q) *Sea $\alpha(x_0, x_1, \ldots, x_n)$ una fórmula de tipo Δ_0 cuyas variables libres estén entre las indicadas. Sea $\mathbb{N}$ el modelo natural de la aritmética. Si $\mathbb{N} \vDash \bigvee x_0\, \alpha(x_0, 0^{(a_1)}, \ldots, 0^{(a_n)})$, entonces*

$$\underset{\mathrm{Q}}{\vdash} \bigvee x_0\, \alpha(x_0, 0^{(a_1)}, \ldots, 0^{(a_n)}).$$

Demostración: En primer lugar probaremos que si $t(x_0, \ldots, x_n)$ es un término sin descriptores cuyas variables libres estén entre las indicadas, $k_0, \ldots, k_n$ son números naturales y $m \equiv \mathbb{N}(t(0^{(k_0)}, \ldots, 0^{(k_n)}))$ es el objeto denotado por el designador indicado (que no depende de la valoración que se considere) entonces

$$\underset{\mathrm{Q}}{\vdash} t(0^{(k_0)}, \ldots, 0^{(k_n)}) = 0^{(m)}.$$

En efecto, razonamos por inducción sobre la longitud de t. Si $t \equiv x_i$ es una variable, entonces basta tener en cuenta que $\mathbb{N}(0^{(k_i)}) \equiv k_i$.

Si $t \equiv 0$ es trivial. Si $t \equiv t_0'$ y el resultado es cierto para t_0, entonces, llamando $m = \mathbb{N}(t_0(0^{(k_0)}, \ldots, 0^{(k_n)}))$, tenemos que

$$\underset{\mathrm{Q}}{\vdash} t_0(0^{(k_0)}, \ldots, 0^{(k_n)}) = 0^{(m)},$$

luego

$$\underset{\mathrm{Q}}{\vdash} t_0'(0^{(k_0)}, \ldots, 0^{(k_n)}) = 0^{(m)} + 1,$$

y por otra parte

$$\mathbb{N}(t_0'(0^{(k_0)}, \ldots, 0^{(k_n)})) \equiv \mathbb{N}(t_0(0^{(k_0)}, \ldots, 0^{(k_n)})) + 1 = m + 1.$$

Ahora basta usar que, como sabemos, $\underset{\mathrm{Q}}{\vdash} 0^{(m)} + 0^{(1)} = 0^{(m+1)}$.

Si $t \equiv t_1 + t_2$ y suponemos el resultado para t_1, t_2, tenemos que

$$\underset{\mathrm{Q}}{\vdash} t_1(0^{(k_0)}, \ldots, 0^{(k_n)}) = 0^{(m_1)}, \qquad \underset{\mathrm{Q}}{\vdash} t_2(0^{(k_0)}, \ldots, 0^{(k_n)}) = 0^{(m_2)},$$

donde $m_1 \equiv \mathbb{N}(t_1(0^{(k_0)}, \ldots, 0^{(k_n)}))$ y $m_2 \equiv \mathbb{N}(t_2(0^{(k_0)}, \ldots, 0^{(k_n)}))$. Entonces obviamente

$$\underset{\mathrm{Q}}{\vdash} t(0^{(k_0)}, \ldots, 0^{(k_n)}) = 0^{(m_1)} + 0^{(m_2)},$$

y basta tener en cuenta que

$$\mathbb{N}(t(0^{(k_0)}, \ldots, 0^{(k_n)})) = m_1 + m_2, \qquad \underset{\mathrm{Q}}{\vdash} 0^{(m)} + 0^{(n)} = 0^{(m+n)}.$$

El caso $t = t_1 t_2$ se razona análogamente.

Ahora probamos que si $\alpha(x_0, \ldots, x_n)$ es Δ_0 y tiene sus variables libres entre las indicadas entonces

$$\mathbb{N} \vDash \alpha(0^{(k_0)}, \ldots, 0^{(k_n)}) \quad \text{implica que} \quad \underset{\mathrm{Q}}{\vdash} \alpha(0^{(k_0)}, \ldots, 0^{(k_n)})$$

y

$$\mathbb{N} \vDash \neg\alpha(0^{(k_1)}, \ldots, 0^{(k_n)}) \quad \text{implica que} \quad \underset{\mathrm{Q}}{\vdash} \neg\alpha(0^{(k_1)}, \ldots, 0^{(k_n)}).$$

En efecto, si $\alpha \equiv t_1 = t_2$, llamemos

$$m_1 \equiv \mathbb{N}(t_1(0^{(k_0)}, \ldots, 0^{(k_n)})), \quad m_2 \equiv \mathbb{N}(t_2(0^{(k_0)}, \ldots, 0^{(k_n)})).$$

Acabamos de probar que

$$\underset{\mathrm{Q}}{\vdash} t_1(0^{(k_0)}, \ldots, 0^{(k_n)}) = 0^{(m_1)}, \qquad \underset{\mathrm{Q}}{\vdash} t_2(0^{(k_0)}, \ldots, 0^{(k_n)}) = 0^{(m_2)}.$$

Por otra parte, $\mathbb{N} \vDash \alpha(0^{(k_0)}, \ldots, 0^{(k_n)})$ equivale a $m_1 \equiv m_2$, luego basta tener en cuenta que si $m_1 \equiv m_2$ trivialmente $\underset{\mathrm{Q}}{\vdash} 0^{(m_1)} = 0^{(m_2)}$ y si $m_1 \not\equiv m_2$ hemos probado que $\underset{\mathrm{Q}}{\vdash} 0^{(m_1)} \neq 0^{(m_2)}$.

El caso $\alpha \equiv t_1 \leq t_2$ es análogo, usando ahora que, por 5.2, si $m_1 \leq m_2$ entonces $\underset{\mathrm{Q}}{\vdash} 0^{(m_1)} \leq 0^{(m_2)}$, mientras que si $m_2 < m_1$ entonces $\underset{\mathrm{Q}}{\vdash} \neg 0^{(m_1)} \leq 0^{(m_2)}$.

Por el planteamiento de la inducción, el caso $\alpha \equiv \neg\beta$ es trivial. Supongamos que $\alpha \equiv \beta \to \gamma$. Si $\mathbb{N} \vDash (\alpha(0^{(k_0)}, \ldots, 0^{(k_n)}) \to \beta(0^{(k_0)}, \ldots, 0^{(k_n)}))$, entonces

$$\text{no } \mathbb{N} \vDash \alpha(0^{(k_0)}, \ldots, 0^{(k_n)}) \quad \text{o } \mathbb{N} \vDash \beta(0^{(k_0)}, \ldots, 0^{(k_n)}).$$

Por lo tanto

$$\underset{\mathrm{Q}}{\vdash} \neg\alpha(0^{(k_0)}, \ldots, 0^{(k_n)}) \quad \text{o} \quad \underset{\mathrm{Q}}{\vdash} \beta(0^{(k_0)}, \ldots, 0^{(k_n)})$$

y en ambos casos en Q se prueba la implicación. El caso para $\neg(\alpha \to \beta)$ es análogo. Supongamos ahora que $\alpha \equiv \bigwedge x \leq x_i\, \beta(x, x_0, \ldots, x_n)$. Si se cumple

$$\mathbb{N} \vDash \bigwedge x \leq 0^{(k_i)}\, \beta(x, 0^{(k_0)}, \ldots, 0^{(k_n)}),$$

entonces para todo $r \leq k_i$ se cumple $\mathbb{N} \vDash \beta(0^{(r)}, 0^{(k_0)}, \ldots, 0^{(k_n)})$, luego por hipótesis de inducción

$$\underset{\mathrm{Q}}{\vdash} \beta(0^{(r)}, 0^{(k_0)}, \ldots, 0^{(k_n)}).$$

Ahora usamos que

$$\underset{\mathrm{Q}}{\vdash} (x \leq 0^{(k_i)} \to x = 0^{(0)} \vee \cdots \vee x = 0^{(k_i)}).$$

Como $\underset{\mathrm{Q}}{\vdash} x = 0^{(r)} \to \beta(x, 0^{(k_0)}, \ldots, 0^{(k_n)})$ para todo $r \leq k_i$, de aquí se sigue que $\underset{\mathrm{Q}}{\vdash} \bigwedge x(x \leq 0^{(k_i)} \to \beta(x, 0^{(k_0)}, \ldots, 0^{(k_n)}))$.

El caso de la negación es más sencillo, y los casos cuando se cumple que $\alpha \equiv \bigvee x \leq x_i\, \beta(x, x_0, \ldots, x_n)$ son los mismos que los correspondientes al generalizador.

Finalmente, si $\mathbb{N} \vDash \bigvee x_0\, \alpha(x_0, 0^{(a_1)} \ldots, 0^{(a_n)})$, entonces existe un a_0 tal que $\mathbb{N} \vDash \alpha(0^{(a_0)}, \ldots, 0^{(a_n)})$, luego $\underset{\mathrm{Q}}{\vdash} \alpha(0^{(a_0)}, \ldots, 0^{(a_n)})$, de donde claramente $\underset{\mathrm{Q}}{\vdash} \bigvee x_0\, \alpha(x_0, 0^{(a_1)}, \ldots, 0^{(a_n)})$. ∎

5.2 La aritmética con inducción abierta

Para demostrar las propiedades básicas de la suma, el producto y la relación de orden es suficiente añadir a Q el principio de inducción para fórmulas abiertas (definición 5.5). Llamaremos IA a esta teoría axiomática, y todas las demostraciones de este apartado se hacen en IA.

El teorema siguiente demuestra entre otras cosas que la suma y el producto tienen las propiedades asociativa y conmutativa, por lo que en lo sucesivo escribiremos expresiones de la forma $x_1 + \cdots + x_n$ y $x_1 \cdots x_n$ sin preocuparnos del orden o de la forma en que se asocian los sumandos o factores.

Teorema 5.7 *Se cumple:*

1. $\bigwedge xyz\ x + (y + z) = (x + y) + z$
2. $\bigwedge xy\ x + y = y + x$
3. $\bigwedge xy\ xy = yx$
4. $\bigwedge xyz\ (x + y)z = xz + yz$
5. $\bigwedge xyz\ x(yz) = (xy)z$
6. $\bigwedge xyz(x + z = y + z \to x = y)$

Demostración: 1) Por inducción con[2] $\phi(z) \equiv x + (y + z) \equiv (x + y) + z$. Para $z = 0$ es inmediato y, si vale para z, entonces

$$x + (y + (z + 1)) = x + ((y + z) + 1) = (x + (y + z)) + 1$$

$$= ((x + y) + z) + 1 = (x + y) + (z + 1).$$

2) En primer lugar demostramos que $\bigwedge x(0 + x = x)$. Para ello aplicamos el esquema de inducción a la fórmula $\phi(x) \equiv 0 + x = x$. Ciertamente se cumple para 0 y, si $0 + x = x$, entonces

$$0 + (x + 1) = (0 + x) + 1 = x + 1.$$

Ahora razonamos por inducción con la fórmula $\phi(y) \equiv (x+1)+y = (x+y)+1$. Para $y = 0$ se reduce a $x + 1 = x + 1$. Si se cumple para y tenemos que

$$(x + 1) + (y + 1) = ((x + 1) + y) + 1 = ((x + y) + 1) + 1 = (x + (y + 1)) + 1.$$

Por último consideramos $\phi(y) \equiv x + y = y + x$. Para $y = 0$ es lo primero que hemos probado. Si se cumple $x + y = y + x$, entonces

$$x + (y + 1) = (x + y) + 1 = (y + x) + 1 = (y + 1) + x,$$

por el resultado precedente.

[2]Es inmediato (pero debe ser comprobado) que todas las fórmulas a las que aplicamos el principio de inducción son abiertas.

3) Primero aplicamos inducción a la fórmula $\phi(x) \equiv 0 \cdot x = 0$, seguidamente a $\phi(y) \equiv (x+1)y = xy + y$. Para 0 la comprobación es trivial y

$$(x+1)(y+1) = (x+1)y + x + 1 = xy + y + x + 1 = xy + x + y + 1$$

$$= x(y+1) + (y+1).$$

Por último usamos $\phi(x) \equiv xy = yx$. Para 0 es trivial y

$$(x+1)y = xy + y = yx + y = y(x+1).$$

4) Por inducción con $\phi(z) \equiv (x+y)z = xz + yz$

5) Por inducción con $\phi(z) \equiv x(yz) = (xy)z$.

6) Por inducción con $\phi(z) \equiv (x + z = y + z \to x = y)$. ∎

Nos ocupamos ahora de la relación de orden de los números naturales:

Teorema 5.8 *Se cumple:*

1. $\bigwedge xy(x \leq y \vee y \leq x)$
2. $\bigwedge x\ x \leq x$
3. $\bigwedge xy(x \leq y \wedge y \leq x \to x = y)$
4. $\bigwedge xyz(x \leq y \wedge y \leq z \to x \leq z)$
5. $\bigwedge xyz(x \leq y \leftrightarrow x + z \leq y + z)$
6. $\bigwedge xyz(z \neq 0 \wedge xz = yz \to x = y)$
7. $\bigwedge xyz(z \neq 0 \to (x \leq y \leftrightarrow xz \leq yz)$

DEMOSTRACIÓN: 1) Por inducción con $\phi(x) \equiv (x \leq y \vee y \leq x)$.

Para $x = 0$ sabemos por 5.3 que $0 \leq y$. Supuesto cierto para x, si $y \leq x$, entonces existe un z tal que $y + z = x$, luego $y + z + 1 = x + 1$, luego $y \leq x + 1$. Si $x \leq y$, entonces $x + z = y$. Si $z = 0$ tenemos que $y = x \leq x + 1$, y si $z \neq 0$ entonces $z = u + 1$, luego $u + x + 1 = y$, luego $x + 1 \leq y$.

2) es trivial, pues $x + 0 = x$.

3) Tenemos que $u + x = y$ y $v + y = x$, luego $u + v + y = 0 + y$, luego $u + v = 0$, luego $u = 0$ por 5.3, luego $x = y$.

4) Tenemos que $x + u = y$ y $y + v = z$, luego $x + u + v = z$, luego $x \leq z$.

5) Se cumple $x \leq y$ si y sólo si existe un u tal que $x + u = y$, si y sólo si $x + z + u = y + z$ si y sólo si $x + z \leq y + z$.

6) Si $xz = yz$, no perdemos generalidad si suponemos $x \leq y$, es decir, $x + u = y$. Entonces $xz + uz = yz$, luego $uz = 0$, luego $u = 0$, luego $x = y$.

7) Si $x \leq y$, entonces $x + u = y$, luego $xz + uz = yz$, luego $xz \leq yz$. Si $xz \leq yz$, o bien $x \leq y$, como queremos probar, o bien $y \leq x$, en cuyo caso $yz \leq xz$, luego $yz = xz$, luego $x = y$, luego $x \leq y$. ∎

Tenemos, pues, que $\leq$ es una relación de orden total cuyo mínimo es 0. Escribiremos

$$x < y \equiv x \leq y \wedge x \neq y \leftrightarrow \bigvee z(z \neq 0 \wedge z + x = y).$$

Observemos que claramente $x < y \to x + 1 \leq y$, por lo que no hay números entre x y $x + 1$ o, en otros términos, $x + 1$ es el menor número mayor que x. Es inmediato que

$$\bigwedge xyz(x < y \leftrightarrow x + z < y + z), \qquad \bigwedge xyz(z \neq 0 \to (x < y \leftrightarrow xz < yz)).$$

Si $x \leq y$, tenemos que existe un z tal que $x + z = y$, y por 5.7 6) este z resulta estar unívocamente determinado. Esto justifica la definición siguiente:

Definición 5.9 $y - x \equiv z \mid (y = z + x)$.

Según acabamos de observar, $y-x$ es una descripción propia si y sólo si $x \leq y$, pero, por el convenio que hemos adoptado sobre las descripciones impropias tenemos que $y < x \to y - x = 0$. El teorema siguiente se demuestra con facilidad:

Teorema 5.10 *Se cumple*

1. $\bigwedge xyz(x \geq y \to x - y = (x + z) - (y + z))$,
2. $\bigwedge xyz(x \geq y \to (x - y) + z = (x + z) - y)$,
3. $\bigwedge xyz\ (x - y)z = xz - yz$.

Ahora demostramos el teorema sobre la división euclídea:

Teorema 5.11 $\bigwedge xy(y \neq 0 \to \overset{1}{\bigvee} cr(r < y \wedge x = yc + r)$

Demostración: Sea $\phi(c) \equiv yc \leq x$. Se cumple $\phi(0)$ y existe un c tal que $\neg\phi(c)$, por ejemplo, $c = x + 1$, pues $y(x + 1) \geq 1(x + 1) > x$. Por lo tanto, no puede ser cierto $\bigwedge c(\phi(c) \to \phi(c+1))$, ya que entonces por inducción tendríamos $\bigwedge c\, \phi(c)$. Así pues, existe un c tal que $\phi(c) \wedge \neg\phi(c + 1)$. Explícitamente, esto significa que $yc \leq x < yc + y$. En particular $c \leq yc \leq x$. Sea $r = x - yc$, de modo que $x = yc + r < yc + y$, luego $r < y$.

Para probar la unicidad suponemos $x = yc + r = yc' + r'$ con $r, r' < y$. No perdemos generalidad si suponemos $r \leq r'$. Entonces $yc = yc' + (r' - r)$, luego $yc' \leq yc$, luego $y(c - c') = yc - yc' = r' - r \leq r' < y$. Si $r \neq r'$ entonces tiene que ser $c = c'$, porque en caso contrario $y \leq y(c - c') < y$. Pero entonces $yc + r = yc + r'$ y $r = r'$. Concluimos que $r = r'$, y entonces $yc = yc'$, luego $c = c'$. ∎

Por último demostramos las propiedades básicas de la divisibilidad:

Definición 5.12 $x \mid y \equiv \bigvee z\ y = xz$.

Teorema 5.13 *Se cumple:*

1. $\bigwedge x(1 \mid x \wedge x \mid x \wedge x \mid 0)$,
2. $\bigwedge xy(x \mid y \wedge y \neq 0 \to x \leq y)$,
3. $\bigwedge xyz(x \mid y \wedge y \mid z \to x \mid z)$,
4. $\bigwedge xy(x \mid y \wedge y \mid x \to x = y)$,
5. $\bigwedge xyz(x \mid y \to x \mid yz)$,
6. $\bigwedge xyz(x \mid y \wedge x \mid z \to x \mid (y \pm z))$.

DEMOSTRACIÓN: 1) es inmediato. Para probar 2) observamos que si $y = xz$ con $y \neq 0$, entonces $z \neq 0$, luego $1 \leq z$, luego $x \leq xz = y$.

Para 4) observamos que si $y = ux \wedge x = vy$ entonces, descartando el caso trivial en que $x = y = 0$, tenemos que $y = uvy$, luego $uv = 1$ y esto implica $u = 1$ por 2). Todo lo demás es sencillo. ∎

Pares ordenados Ahora estamos en condiciones de definir en IA un concepto fundamental para poner de manifiesto la capacidad expresiva de las teorías aritméticas. Vamos a ver que es posible definir un criterio para identificar los números naturales con los pares ordenados de números naturales. La idea subyacente la muestra esta figura:

⋮						
4	10					
3	6	11				
2	3	7	12			
1	1	4	8	13		
0	0	2	5	9	14	
	0	1	2	3	4	⋯

Vamos a definir el par $\langle x, y\rangle$ como el número natural situado en la fila x y la columna y de la figura. La diagonal que contiene, por ejemplo, al $13 = \langle 3, 1\rangle$ contiene todos los pares cuyas componentes suman $x + y = 4$. Para llegar a ella hay que pasar antes por las diagonales anteriores, que contienen $1+2+3+4 = 10$ números, pero como empezamos en el 0 resulta que el $10 = \langle 0, 4\rangle$ es ya el primero de dicha diagonal. Para llegar a $\langle 3, 1\rangle$ hemos de avanzar $x = 3$ posiciones. En general, el par $\langle x, y\rangle$ se alcanza en la posición

$$z = 1 + 2 + \cdots + (x + y) + x = \frac{(x + y)(x + y + 1)}{2} + x.$$

Razonando en IA, diremos que un número x es *par* si $2 \mid x$, y en caso contrario es *impar*. Notemos que $\bigwedge x(2 \mid x \vee 2 \mid x + 1)$. Basta expresar $x = 2u + r$, con $r = 0, 1$.

En particular, o bien $2 \mid (x+y)$ o bien $2 \mid (x+y+1)$, luego en cualquier caso $2 \mid (x+y)(x+y+1)$, y también $2 \mid (x+y)(x+y+1)+2x$, luego existe un (único) z tal que

$$2z = (x+y)(x+y+1)+2x.$$

Definición 5.14 $\langle x,y\rangle \equiv z \mid 2z = (x+y)(x+y+1)+2x$.

Ahora demostramos el teorema básico:

Teorema 5.15 $\bigwedge z \overset{1}{\bigvee} xy\ z = \langle x,y\rangle$.

Demostración: Sea $\phi(r) \equiv r(r+1) \leq 2z$. Claramente $\phi(0) \wedge \neg\phi(z+1)$, luego no puede realizarse la inducción sobre ϕ, es decir, existe un r tal que $\phi(r) \wedge \neg\phi(r+1)$. Explícitamente, $r(r+1) \leq 2z < (r+1)(r+2)$. Es fácil ver que r es único, pues si $r' \leq r$ se cumple $\phi(r')$ y si $r' > r$ se cumple $\neg\phi(r')$.

Como $2z - r(r+1)$ es par, existe un x tal que $2x = 2z - r(r+1)$. Se cumple que $x \leq r$, pues en caso contrario $2r < 2z - r^2 - r$, luego $r^2 + 3r < 2z$. Es fácil ver que la suma y el producto de pares es par, que el producto de impares es impar y que la suma de impares es par. De aquí se sigue (distinguiendo casos según que r sea par o impar) que r^2+3r es par en cualquier caso, luego $r^2+3r+1 \leq 2z$ no puede cumplirse con igualdad, luego $r^2+3r+2 \leq 2z$, es decir, $(r+1)(r+2) \leq 2z$, contradicción.

Así pues, $x \leq r$ y existe un y tal que $x+y = r$. En definitiva llegamos a que $2z = 2x + (x+y)(x+y+1)$, que es lo mismo que $z = \langle x,y\rangle$.

Si $\langle x,y\rangle = \langle x',y'\rangle$, llamamos $r' = x'+y'$, de modo que $2z = r'(r'+1)+2x'$, con $x' \leq r'$, luego

$$r'(r'+1) \leq 2z \leq r'^2 + r' + 2r' < r'^2 + 3r' + 2 = (r'+1)(r'+2).$$

Por la unicidad de r resulta que $r' = r$, luego $2x = 2x'$, luego $x = x'$, luego $y = y'$. ∎

Más explícitamente, la unicidad que hemos probado equivale a que

$$\bigwedge xyx'y'(\langle x,y\rangle = \langle x',y'\rangle \to x = x' \wedge y = y').$$

Así, a partir de ahora, podemos ver al número $13 = \langle 3,1\rangle$ como el par de números naturales cuya primera componente es 3 y cuya segunda componente es 1. En general, podemos definir:

Definición 5.16 $z_0 \equiv x \mid \bigvee y\ z = \langle x,y\rangle, \qquad z_1 \equiv y \mid \bigvee x\ z = \langle x,y\rangle$.

Así, para todo z se cumple que $z = \langle z_0, z_1\rangle$.

Notemos que $x, y \leq \langle x,y\rangle$. En efecto, podemos suponer que $x+y \geq 1$, y entonces

$$2\langle x,y\rangle \geq (x+y)(x+y+1) = (x+y)^2 + x + y \geq x+y+x+y = 2x+2y,$$

luego $\langle x,y\rangle \geq x+y$. ∎

La tabla siguiente muestra que los pocos conceptos aritméticos que hemos definido hasta ahora son Δ_0:

$x \le y$	$\bigvee z \le y\, z = x$	$z = y - x$	$y = x + z \vee (y \le x \wedge z = 0)$
$x \mid y$	$\bigvee z \le y\, y = xz$	$z = \langle x, y\rangle$	$2z = (x + y)(x + y + 1) + 2x$
$x = z_0$	$\bigvee y \le z\, z = \langle x, y\rangle$	$y = z_1$	$\bigvee x \le z\, z = \langle x, y\rangle$

Esto es relevante debido a la teoría que vamos a introducir en la sección siguiente:

5.3 La aritmética con inducción Σ_1

Entre la aritmética con inducción abierta y la aritmética de Peano (en la que admitimos la inducción para fórmulas aritméticas arbitrarias) hay un caso intermedio que conviene estudiar separadamente y que definimos a continuación:

Definición 5.17 Diremos que una fórmula de $\mathcal{L}_a$ es de tipo Σ_1 (resp. Π_1) si es de la forma $\bigvee x\, \alpha$, (resp. $\bigwedge x\, \alpha$) donde α es una fórmula Δ_0.

Más en general, si T es una teoría axiomática en la que pueden probarse los axiomas de IA, diremos que una fórmula cualquiera es de tipo Σ_1^T o Π_1^T si es equivalente en T a una fórmula del tipo correspondiente. Una fórmula es de tipo Δ_1^T si es a la vez Σ_1^T y Π_1^T. Normalmente omitiremos el superíndice T si está claro por el contexto en qué teoría estamos trabajando.

Notemos que toda fórmula α de tipo Δ_0 es también Δ_1, pues si x es cualquier variable que no aparezca en ella, es equivalente a $\bigvee x\, \alpha$ y a $\bigwedge x\, \alpha$.

El teorema siguiente ayuda a identificar las fórmulas de tipo Σ_1 y Π_1:

Teorema 5.18 *Sea T una teoría axiomática que contenga a* IA *y sean α y β fórmulas de $\mathcal{L}_a$. Entonces:*[3]

1. *Si α, β son Σ_1, Π_1, lo mismo vale para $\alpha \wedge \beta$ y $\alpha \vee \beta$.*
2. *Si α es Π_1 (resp. Σ_1) y β es Σ_1 (resp. Π_1), $\alpha \to \beta$ es Σ_1 (resp. Π_n).*
3. *Si α es Σ_1 entonces $\neg\alpha$ es Π_1, y viceversa.*
4. *Si α es Σ_1, también lo es $\bigvee x\alpha$.*
5. *Si α es Π_1, también lo es $\bigwedge x\alpha$.*

Demostración: Notemos que 2) es consecuencia inmediata de 1), pues $\alpha \to \beta$ equivale a $\neg\alpha \vee \beta$, y 3) es inmediata, pues al anteponer un negador a cuantificador podemos pasarlo a la derecha invirtiendo el cuantificador.

[3] Véase además el teorema 5.23, más abajo.

4) Consideremos una fórmula α de clase Σ_1, es decir que es equivalente a otra de la forma $\bigvee y\,\beta$, donde β es Δ_0. Entonces $\bigvee x\alpha$ es equivalente a $\bigvee xy\,\beta$, y basta probar que esta fórmula es Σ_1. Ahora bien:

$$\bigvee xy\,\beta \leftrightarrow \bigvee u\bigwedge xy(u = \langle x, y\rangle \to \beta) \leftrightarrow \bigvee u\bigwedge x \le u\bigwedge y \le u(u = \langle x, y\rangle \to \beta).$$

Como $u = \langle x, y\rangle$ es Δ_0, es claro que todo lo que hay tras el $\bigvee u$ es Δ_0, luego la fórmula es Σ_1 como había que probar.

La prueba de 5) es análoga.

1) Tomemos fórmulas α y β equivalentes a fórmulas $\bigvee x\,\alpha'$ y $\bigvee y\,\beta'$, respectivamente, con α' y β' de tipo Δ_0. Podemos suponer que x no aparece en $\bigvee y\,\beta'$ y que y no aparece en $\bigvee x\,\alpha'$. Entonces $\alpha \wedge \beta$ es equivalente a

$$\bigvee x\,\alpha' \wedge \bigvee y\,\beta' \leftrightarrow \bigvee xy(\alpha' \wedge \beta').$$

Como $\alpha' \wedge \beta'$ es Δ_0, la fórmula completa es Σ_1 por 4).

El caso de $\alpha \vee \beta$ es idéntico. ∎

Conviene observar que la clave en la demostración del teorema anterior es la posibilidad de contraer cuantificadores usando los pares ordenados que hemos definido.

El concepto siguiente también ayuda a identificar fórmulas Σ_1, Π_1 y Δ_1:

Definición 5.19 Diremos que un término t de $\mathcal{L}_a$ es Δ_0, Σ_1, Π_1 o Δ_1 (en una teoría T) si y sólo si lo es la fórmula $x = t$, donde x es una variable que no esté en t.

Notemos que si un término t es Σ_1 (en una teoría que extienda a IA) entonces es Δ_1, pues

$$x = t \leftrightarrow \bigwedge u(u = t \to x = u).$$

Por otra parte, si $\phi(x_1, \ldots, x_n)$ es una fórmula Σ_1, Π_1 o Δ_1 y $t_1, \ldots, t_n$ son términos Δ_1 (en una teoría que extienda a IA), entonces la fórmula $\phi(t_1, \ldots, t_n)$ es del mismo tipo que ϕ. En efecto (suponiendo sin pérdida de generalidad que $x_1, \ldots, x_n$ no están en $t_1, \ldots, t_n$):

$$\phi(t_1, \ldots, t_n) \leftrightarrow \bigvee x_1 \cdots x_n(x_1 = t_1 \wedge \cdots \wedge x_n = t_n \wedge \phi(x_1, \ldots, x_n))$$
$$\leftrightarrow \bigwedge x_1 \cdots x_n(x_1 = t_1 \wedge \cdots \wedge x_n = t_n \to \phi(x_1, \ldots, x_n)).$$

Definición 5.20 Llamaremos $\mathrm{I}\Sigma_1$ a la teoría que resulta de añadir a la teoría básica Q el principio de inducción restringido a fórmulas de tipo Σ_1.

En principio, podemos definir $\mathrm{I}\,\Pi_1$ como la teoría que resulta de añadir a Q el principio de inducción para fórmulas de tipo Π_1, pero el teorema siguiente demuestra que $\mathrm{I}\Sigma_1$ e $\mathrm{I}\,\Pi_1$ son la misma teoría. Notemos antes que, aunque los axiomas de inducción de $\mathrm{I}\Sigma_1$ (resp. $\mathrm{I}\,\Pi_1$) son los que corresponden a fórmulas del tipo considerado en sentido estricto, es claro que en ella se pueden probar los casos correspondientes a fórmulas equivalentes en la teoría a las fórmulas del tipo correspondiente.

Teorema 5.21 *En* $\mathrm{I}\Sigma_1$ *se puede probar el principio de inducción para fórmulas de tipo* Π_1*, mientras que en* $\mathrm{I}\,\Pi_1$ *se puede probar el principio de inducción para fórmulas de tipo* Σ_1*.*

DEMOSTRACIÓN: Tomemos una fórmula ϕ de tipo Π_1 y supongamos

$$\phi(0) \wedge \bigwedge x(\phi(x) \to \phi(x+1)).$$

Queremos probar que $\bigwedge x\,\phi(x)$. Supongamos, por reducción al absurdo, que existe un a tal que $\neg\phi(a)$. Aplicamos inducción a la fórmula

$$\psi(z) \equiv z \le a \to \bigvee u(u + z = a \wedge \neg\phi(u)),$$

que es de tipo Σ_1 (en $\mathrm{I}\Sigma_1$). Obviamente se cumple $\psi(0)$ y, supuesto $\psi(z)$, si $z+1 \le a$, entonces $z \le a$, luego por $\psi(z)$ existe un u tal que $u+z = a \wedge \neg\phi(u)$. No puede ser $u = 0$, luego $u = v + 1$, con $z + 1 + v = a$, y tiene que ser $\neg\phi(v)$, pues $\phi(v) \to \phi(u)$. Así pues, $\bigvee v(v + z + 1 = a \wedge \neg\phi(v))$, y esto es $\psi(z+1)$.

Por Σ_n-inducción tenemos $\bigwedge x\,\psi(x)$. En particular $\psi(a)$, que implica $\neg\phi(0)$, contradicción.

El recíproco se prueba análogamente: si partimos de una fórmula $\phi(x)$ de tipo Σ_1 y suponemos que cumple las hipótesis del principio de inducción pero existe un a tal que $\neg\phi(a)$, aplicamos Π_1-inducción a la fórmula

$$\psi(z) \equiv \bigwedge u(z + u = a \to \neg\phi(u)).$$

El argumento es muy similar. ∎

Tras la definición 5.1 observamos que el axioma Q3 es redundante en AP, y la prueba usa el principio de inducción con una fórmula Σ_1, por lo que podemos decir, más concretamente, que Q3 es redundante en $\mathrm{I}\Sigma_1$ (y en cualquier teoría que la extienda). En particular, la teoría $\mathrm{I}\Sigma_1$ se obtiene también de restringir el principio de inducción en AP a fórmulas Σ_1.

Claramente tenemos $\mathrm{Q} \subset \mathrm{IA} \subset \mathrm{I}\Sigma_1 \subset \mathrm{AP}$, donde la inclusión entre dos teorías quiere decir que los axiomas de la primera son demostrables en la segunda.

Un resultado muy importante es que las fórmulas Σ_1 y Π_1 son cerradas para cuantificadores acotados. Para probarlo necesitamos un hecho previo:

Teorema 5.22 (Principio de Recolección) *Para toda fórmula* $\phi(x, y)$ *(posiblemente con más variables libres, pero distintas de* v*) la fórmula siguiente es un teorema de* AP *(y de* $\mathrm{I}\Sigma_1$ *si* ϕ *es* Δ_0*):*

$$\bigwedge x \le u \bigvee y\,\phi(x, y) \to \bigvee v \bigwedge x \le u \bigvee y \le v\,\phi(x, y).$$

DEMOSTRACIÓN: Supongamos $\bigwedge x \le u \bigvee y\,\phi(x, y)$. Vamos a aplicar el principio de inducción a la fórmula

$$\psi(w) \equiv w \le u + 1 \to \bigvee v \bigwedge x < w \bigvee y \le v\,\phi(x, y).$$

Notemos que si ϕ es de tipo Δ_0 entonces ψ es de tipo Σ_1 y la inducción es válida en $\mathrm{I}\Sigma_1$.

Para $w = 0$ es trivial. Si vale para w, suponemos que $w + 1 \leq u + 1$. Por hipótesis existe un y_0 tal que $\phi(w, y_0)$ y sea $v' = \text{máx}\{v, y_0\}$. Entonces

$$\bigwedge x < w + 1 \bigvee y \leq v'\, \phi(x, y).$$

Concluimos que se cumple $\psi(u + 1)$, y eso es lo que queríamos probar. ∎

El teorema siguiente vale en cualquier teoría que extienda a $\mathrm{I}\Sigma_1$:

Teorema 5.23 *Si α es una fórmula Σ_1, Π_1 o Δ_1, también lo son las fórmulas $\bigwedge x \leq u\, \alpha$ y $\bigvee x \leq u\, \alpha$.*

Demostración: Si α es Σ_1 es equivalente a una de la forma $\bigvee y\, \phi(x, y)$, con ϕ de tipo Δ_0. Por el teorema anterior tenemos la equivalencia:

$$\bigwedge x \leq u \bigvee y\, \phi(x, y) \leftrightarrow \bigvee v \bigwedge x \leq u \bigvee y \leq v\, \phi(x, y).$$

La fórmula tras el $\bigvee v$ es Δ_0, luego la fórmula completa es Σ_1, como había que probar.

La clausura respecto a $\bigvee x \leq u$ se sigue de 5.18. Si α es Π_1 razonamos igual, pero aplicando la equivalencia a la fórmula $\neg\phi$ y negando ambas partes, con lo cual nos queda que

$$\bigvee x \leq u \bigwedge y\, \phi(x, y) \leftrightarrow \bigwedge v \bigvee x \leq u \bigwedge y \leq v\, \phi(x, y).$$

∎

Seguidamente demostramos la variante fuerte del principio de inducción:

Teorema 5.24 *Si $\phi(x)$ es cualquier fórmula, la fórmula siguiente es un teorema de* AP *(y se demuestra en $\mathrm{I}\Sigma_1$ para fórmulas Σ_1 o Π_1):*

$$\bigwedge x(\bigwedge y < x\ \phi(y) \to \phi(x)) \to \bigwedge x\, \phi(x).$$

Demostración: Por inducción sobre $\psi(x) \equiv \bigwedge y < x\, \phi(y)$. Para $x = 0$ se cumple trivialmente. Supongamos que $\bigwedge y < x\, \phi(y)$. Entonces por hipótesis $\phi(x)$. Veamos que $\bigwedge y < x+1\, \phi(y)$. En efecto, si $y < x+1$, entonces $y \leq x$, luego o bien $y < x$ (en cuyo caso $\phi(y)$ por hipótesis de inducción) o bien $y = x$ (en cuyo caso ya hemos observado que se cumple $\phi(y)$). Concluimos que $\bigwedge x \psi(x)$, luego, para todo x se cumple $\psi(x + 1)$, lo cual implica $\phi(x)$. ∎

El teorema siguiente afirma que si existe un número natural que cumple una propiedad entonces existe un mínimo número que la cumple.

Teorema 5.25 *Si $\phi(x)$ es cualquier fórmula, la fórmula siguiente es un teorema de* AP *(y se demuestra en $\mathrm{I}\Sigma_1$ para fórmulas Σ_1 o Π_1):*

$$\bigvee x \phi(x) \to \overset{1}{\bigvee} x(\phi(x) \wedge \bigwedge y < x\, \neg\phi(y)).$$

DEMOSTRACIÓN: Supongamos $\bigvee x\phi(x)$. Si la existencia es falsa, tenemos que $\bigwedge x(\phi(x) \to \bigvee y < x\ \phi(y))$. Ahora razonamos por inducción con la fórmula $\psi(x) \equiv \bigwedge y < x\, \neg\phi(y)$. Obviamente se cumple para 0. Si vale para x, tomemos un $y < x + 1$, es decir, $y \leq x$. Si $y < x$ se cumple $\neg\phi(y)$ por hipótesis de inducción, mientras que si $y = x$ también tiene que ser $\neg\phi(x)$, ya que en caso contrario tendría que existir un $y < x$ que cumpliera $\phi(y)$, y no existe.

Esto prueba $\bigwedge x\psi(x)$. Por lo tanto, para todo x tenemos $\psi(x + 1)$, luego $\neg\phi(x)$, contradicción. La unicidad es clara. ∎

Un enunciado claramente equivalente del teorema anterior es

$$\bigvee x\phi(x) \to \overset{1}{\bigvee} x(\phi(x) \wedge \bigwedge y(\phi(y) \to x \leq y)).$$

Bajo la hipótesis obvia de acotación también podemos justificar la existencia de máximo:

Teorema 5.26 *Si $\phi(x)$ es cualquier fórmula, la fórmula siguiente es un teorema de* AP *(y se demuestra en $\mathrm{I}\Sigma_1$ para fórmulas Σ_1 o Π_1):*

$$\bigvee x\, \phi(x) \wedge \bigvee y \bigwedge u(\phi(u) \to u \leq y) \to \overset{1}{\bigvee} x(\phi(x) \wedge \bigwedge u(\phi(u) \to u \leq x)).$$

DEMOSTRACIÓN: Tomemos y según la hipótesis. Basta aplicar el teorema anterior a la fórmula

$$\psi(z) \equiv z \leq y \wedge \phi(y - z) \leftrightarrow \bigvee u \leq z(u + z = y \wedge \phi(u)).$$

Si z es el mínimo que que cumple $\psi(z)$, tomamos $x = y - z$ y claramente cumple lo pedido. ∎

Definición 5.27

$$\text{mín}\, x|\phi(x) \equiv x \mid (\phi(x) \wedge \bigwedge y(\phi(y) \to x \leq y)),$$

$$\text{máx}\, x|\phi(x) \equiv x \mid (\phi(x) \wedge \bigwedge y(\phi(y) \to y \leq x)),$$

Los dos teoremas anteriores dan condiciones suficientes (y de hecho necesarias) para que estas descripciones sean propias.

5.4 Relaciones y funciones aritméticas

En general, una cosa es que una sentencia aritmética sea cierta (en el modelo natural) y otra cosa que sea demostrable en una teoría aritmética dada. La razón por la que hemos destacado las fórmulas de tipo Σ_1, Π_1 y Δ_1 es que para ellas la relación entre que sean verdaderas y que sean demostrables (en una teoría tan elemental como Q) es mucho más estrecha, como vamos a ver en esta sección.

Definición 5.28 Toda fórmula $\phi(x_1, \ldots, x_n)$ de $\mathcal{L}_a$ define una relación en $\mathbb{N}$, la dada por

$$R(m_1, \ldots, m_n) \quad \text{si y sólo si} \quad \mathbb{N} \vDash \phi(0^{(m_1)}, \ldots, 0^{(m_n)}).$$

y cada término $t(x_1, \ldots, x_n)$ define una función en $\mathbb{N}$, la dada por

$$f(a_1, \ldots, a_n) = a \quad \text{syss} \quad \mathbb{N} \vDash t(0^{(a_1)}, \ldots, 0^{(a_n)}) = 0^{(a)}.$$

Las relaciones y funciones definidas de este modo (o que pueden definirse de este modo) se llaman *aritméticas*. Un poco más en general, diremos que una función n-ádica f es aritmética si existe una fórmula ϕ tal que

$$f(a_1, \ldots, a_n) = a \quad \text{syss} \quad \mathbb{N} \vDash \phi(0^{(a_1)}, \ldots, 0^{(a_n)}, 0^{(a)}).$$

Así la función definida por un término t es la misma que la definida por la fórmula $t(x_1, \ldots, x_n) = y$, donde y es una variable que no aparezca libre en t.

Más concretamente, diremos que una relación o función aritmética es Σ_1 o Π_1 si la fórmula ϕ puede tomarse[4] Σ_1 o Π_1. Diremos que la relación o la función es Δ_1 si es a la vez Σ_1 y Π_1.

Notemos que si una función es Σ_1 entonces es de hecho Δ_1. En efecto, si cumple la definición anterior con ϕ de tipo Σ_1, entonces

$$f(a_1, \ldots, a_n) = a \quad \text{syss} \quad \mathbb{N} \vDash \bigwedge x(\phi(0^{(a_1)}, \ldots, 0^{(a_n)}, x) \to x = 0^{(a)}),$$

y esta última fórmula es Π_1 (en IA).

En general, usaremos la misma notación para representar una fórmula o un término y la relación o función que definen sin que ello deba inducir a error. De hecho, ya lo venimos haciendo en los casos más simples: una cosa es el término $x + y$ de $\mathcal{L}_a$ y otra la función suma en $\mathbb{N}$, aunque usemos el mismo signo $+$ en ambos casos. La relación entre ambos es que

$$r = m + n \quad \text{syss} \quad \mathbb{N} \vDash 0^{(r)} = 0^{(m)} + 0^{(n)}.$$

Por ejemplo, en 5.12 definimos la fórmula $x \mid y$, y es fácil ver que la relación

$$m \mid n \quad \text{syss} \quad \mathbb{N} \vDash 0^{(m)} \mid 0^{(n)}$$

no es más que la divisibilidad, es decir, que $m \mid n$ syss existe un r tal que $n = mr$.

Del mismo modo, el término definido en 5.14 nos determina una función Δ_0 en $\mathbb{N}$ que, si desarrollamos la definición $\mathbb{N} \vDash 0^{(r)} = \langle 0^{(m)}, 0^{(n)} \rangle$ vemos que es

$$\langle m, n \rangle = \frac{(m+n)(m+n+1)}{2} + m.$$

[4]En principio, aquí nos referimos a que la fórmula es Σ_1 o Π_1 en sentido estricto, no equivalente a una fórmula de este tipo en alguna teoría. Ahora bien, si ϕ es una fórmula Σ_1^T o Π_1^T para una cierta teoría T sobre $\mathcal{L}_a$ tal que $\mathbb{N} \vDash$ T, entonces la relación definida por ϕ es la misma que la definida por la correspondiente fórmula Σ_1 o Π_1 equivalente, por lo que la relación es también Σ_1 o Π_1.

Como el teorema 5.15 es verdadero en $\mathbb{N}$, podemos afirmar que todo número natural n se expresa de forma única como $n = \langle n_0, n_1\rangle$, para ciertos números n_0, n_1. En general, cualquier teorema que demostremos en una teoría que admita a $\mathbb{N}$ como modelo se traduce en afirmaciones verdaderas sobre las relaciones o funciones que intervengan en él.

También tenemos las funciones inversas n_0 y n_1 definidas por los términos de la definición 5.16. Todas estas relaciones y funciones que estamos considerando son claramente[5] Δ_0.

Ahora bien, como recordábamos al principio de esta sección, debemos tener presente que no es lo mismo

$$\mathbb{N} \vDash \phi(0^{(a_1)}, \dots, 0^{(a_n)}) \quad \text{que} \quad \vdash_T \phi(0^{(a_1)}, \dots, 0^{(a_n)}),$$

para una teoría T. Si $\mathbb{N} \vDash T$ entonces se cumple la implicación $\Leftarrow$, pero, incluso si $T = \mathrm{AP}$, nada nos garantiza que si una sentencia es verdadera en $\mathbb{N}$ tenga que ser demostrable en T. En este sentido debemos recordar el teorema 5.6, que ahora podemos reformular así:

Teorema 5.29 (Σ_1-completitud de Q) *Si $\phi(x_1, \dots, x_n)$ es una fórmula Σ_1, entonces, para todos los números naturales $a_1, \dots, a_n$, se cumple*

$$\mathbb{N} \vDash \phi(0^{(a_1)}, \dots, 0^{(a_n)}) \quad \textit{syss} \quad \vdash_{\mathrm{Q}} \phi(0^{(a_1)}, \dots, 0^{(a_n)}).$$

Así pues, para una sentencia Σ_1, "ser verdadera" es lo mismo que "ser demostrable en Q", pero veremos que "ser falsa" no equivale necesariamente a "ser refutable en Q" (ni siquiera a "ser refutable en AP"). Lo segundo lo cumplen obviamente las fórmulas Π_1, mientras que las fórmulas Δ_1 cumplen ambas cosas. En particular, toda sentencia Δ_1 es demostrable o refutable en Q, pero podemos afinar un poco más:

Teorema 5.30 *Si R es una relación n-ádica Δ_1, entonces existe una fórmula $\phi(x_1, \dots, x_n)$ de tipo Σ_1 tal que*

$$R(a_1, \dots, a_n) \quad \textit{syss} \quad \vdash_{\mathrm{Q}} \phi(0^{(a_1)}, \dots, 0^{(a_n)}),$$

$$\neg R(a_1, \dots, a_n) \quad \textit{syss} \quad \vdash_{\mathrm{Q}} \neg\phi(0^{(a_1)}, \dots, 0^{(a_n)}).$$

Demostración: En principio, la hipótesis significa que existen fórmulas σ y π de tipo Δ_0 tales que

$$R(a_1, \dots, a_n) \quad \text{syss} \quad \mathbb{N} \vDash \bigvee x\, \sigma(x, 0^{(a_1)}, \dots, 0^{(a_n)})$$

$$\text{syss} \quad \mathbb{N} \vDash \bigwedge x\, \pi(x, 0^{(a_1)}, \dots, 0^{(a_n)}).$$

[5]Notemos que la fórmula $x = z_0$ es Δ_0, pues equivale a $\bigvee y \le z\ z = \langle x, y\rangle$, e igualmente con $y = z_1$.

Definimos entonces

$$\phi(x_1,\dots,x_n) \equiv \bigvee x(\sigma(x,x_1,\dots,x_n) \wedge \bigwedge y \le x\ \pi(y,x_1,\dots,x_n)).$$

Ciertamente se trata de una fórmula Σ_1. Veamos que cumple lo pedido. Si se cumple $R(a_1,\dots,a_n)$, existe un m tal que $\mathbb{N} \vDash \sigma(0^{(m)},0^{(a_1)},\dots,0^{(a_n)})$ y, como también $\mathbb{N} \vDash \bigwedge x\,\pi(x,0^{(a_1)},\dots,0^{(a_n)})$, en particular

$$\mathbb{N} \vDash \bigwedge y \le 0^{(m)}\,\pi(y,0^{(a_1)},\dots,0^{(a_n)}),$$

de donde $\mathbb{N} \vDash \phi(0^{(a_1)},\dots,0^{(a_n)})$ y, por el teorema anterior $\underset{\mathrm{Q}}{\vdash} \phi(0^{(a_1)},\dots,0^{(a_n)})$.

Supongamos ahora $\neg R(a_1,\dots,a_n)$, de modo que existe un natural m tal que $\mathbb{N} \vDash \neg\pi(0^{(m)},0^{(a_1)},\dots,0^{(a_n)})$, luego $\underset{\mathrm{Q}}{\vdash} \neg\pi(0^{(m)},0^{(a_1)},\dots,0^{(a_n)})$.

Vamos a probar $\neg\phi(0^{(a_1)},\dots,0^{(a_n)})$ en Q por reducción al absurdo. Suponemos $\phi(0^{(a_1)},\dots,0^{(a_n)})$, es decir, que existe un x tal que

$$\sigma(x,0^{(a_1)},\dots,0^{(a_n)}) \wedge \bigwedge y \le x\ \pi(y,0^{(a_1)},\dots,0^{(a_n)}).$$

Por 5.4 tenemos que $0^{(m)} \le x \vee x \le 0^{(m)}$. El primer caso es imposible, pues nos daría $\pi(0^{(m)},0^{(a_1)},\dots,0^{(a_n)})$, y podemos probar lo contrario. Si se cumple $x \le 0^{(m)}$, entonces, de nuevo por 5.4, podemos probar $x = 0^{(0)} \vee \cdots \vee x = 0^{(m)}$, luego

$$\sigma(0^{(0)},0^{(a_1)},\dots,0^{(a_n)}) \vee \cdots \vee \sigma(0^{(m)},0^{(a_1)},\dots,0^{(a_n)}),$$

pero, como ningún número natural k cumple $\mathbb{N} \vDash \sigma(0^{(k)},0^{(a_1)},\dots,0^{(a_n)})$, en Q podemos probar

$$\neg\sigma(0^{(0)},0^{(a_1)},\dots,0^{(a_n)}) \wedge \cdots \wedge \neg\sigma(0^{(m)},0^{(a_1)},\dots,0^{(a_n)}),$$

y así tenemos una contradicción. ∎

Observemos que si la relación n-ádica R está definida por una fórmula ϕ_0 que es Δ_1 en una teoría T sobre $\mathcal{L}_a$, entonces la fórmula ϕ dada por el teorema anterior puede tomarse equivalente a ϕ_0 en T. En efecto, lo que tenemos es que

$$\underset{T}{\vdash}(\phi_0 \leftrightarrow \bigvee x\,\sigma) \qquad \text{y} \qquad \underset{T}{\vdash}(\phi_0 \leftrightarrow \bigwedge x\,\pi),$$

y es claro entonces que la fórmula ϕ construida en el teorema es equivalente en T a ϕ_0.

Para funciones podemos probar un resultado más simple. Conviene enunciarlo para funciones parciales en el sentido siguiente:

Diremos que una fórmula $\phi(x_1,\dots,x_n,y)$ *define una función n-ádica parcial* F si para todos los números naturales $a_1,\dots,a_n$ existe a lo sumo un m tal que $\mathbb{N} \vDash \phi(0^{(a_1)},\dots,0^{(a_n)},0^{(m)})$, y en tal caso escribiremos $F(a_1,\dots,a_n) = m$. Diremos que F es de tipo Σ_1 o Π_1 si lo es la fórmula ϕ.

Teorema 5.31 *Si F es una función n-ádica parcial de tipo Σ_1, existe una fórmula $\psi(x_1,\dots,x_n,y)$ de $\mathcal{L}_a$ de tipo Σ_1 tal que, si $F(a_1,\dots,a_n)$ está definido, entonces*

$$\underset{\mathrm{Q}}{\vdash} \bigwedge y(\psi(0^{(a_1)},\dots,0^{(a_n)},y) \leftrightarrow y = 0^{(F(a_1,\dots,a_n))}).$$

Esto significa que si $F(a_1,\dots,a_n) = a$ está definida, en Q no sólo se demuestra $\psi(0^{(a_1)},\dots,0^{(a_n)},0^{(a)})$, sino que $0^{(a)}$ es el único número y que cumple $\psi(0^{(a_1)},\dots,0^{(a_n)},y)$.

Demostración: Por abreviar la notación supondremos que $n = 1$, es decir, escribiremos x en lugar de $x_1,\dots,x_n$, pero es claro que la prueba vale para cualquier número de argumentos sin cambio alguno. Sea $\phi(x,y) \equiv \bigvee z\,\sigma(x,y,z)$ una fórmula que defina a F, donde σ es una fórmula Δ_0. Definimos

$$\psi_0(x,y,z) \equiv \sigma(x,y,z) \wedge \bigwedge uv \le y(u \ne y \to \neg\sigma(x,u,v))$$

$$\wedge \bigwedge uv \le z(u \ne y \to \neg\sigma(x,u,v))$$

y a su vez llamamos $\psi(x,y) \equiv \bigvee z\,\psi_0(x,y,z)$.

Si $F(m) = k$, entonces existe un q tal que $\mathbb{N} \vDash \sigma(0^{(m)},0^{(k)},0^{(q)})$. Es claro que $\mathbb{N} \vDash \psi_0(0^{(m)},0^{(k)},0^{(q)})$, luego $\underset{\mathrm{Q}}{\vdash} \psi_0(0^{(m)},0^{(k)},0^{(q)})$, y esto implica $\underset{\mathrm{Q}}{\vdash} \psi(0^{(m)},0^{(k)})$. Por consiguiente, $\underset{\mathrm{Q}}{\vdash} \bigwedge y(y = 0^{(k)} \to \psi(0^{(m)},y))$.

Para probar la implicación contraria, razonando en Q, suponemos

$$y \ne 0^{(k)} \wedge \psi_0(0^{(m)},y,z)$$

para llegar a una contradicción. Sea $h = \text{máx}\{k,q\}$. En Q podemos probar

$$(0^{(h)} \le y \vee y \le 0^{(h)}) \wedge (0^{(h)} \le z \vee z \le 0^{(h)}).$$

Supongamos en primer lugar que $0^{(h)} \le y \vee 0^{(h)} \le z$. Entonces, o bien $0^{(k)} \le y \wedge 0^{(q)} \le y$ o bien $0^{(k)} \le z \wedge 0^{(q)} \le z$. En cualquiera de los dos casos $\psi_0(0^{(m)},y,z)$ implica $\neg\sigma(0^{(m)},0^{(k)},0^{(q)})$, lo que nos da una contradicción, pues en Q puede probarse $\sigma(0^{(m)},0^{(k)},0^{(q)})$.

Por consiguiente tiene que ser $y \le 0^{(h)} \wedge z \le 0^{(h)}$. En Q se demuestra $\psi_0(0^{(m)},0^{(k)},0^{(q)})$, y esto implica $\neg\sigma(0^{(m)},y,z)$, en contradicción con la hipótesis $\psi_0(0^{(m)},y,z)$. ∎

5.5 Conjuntos en $\mathrm{I}\Sigma_1$

Vamos a probar que en la teoría $\mathrm{I}\Sigma_1$ es posible definir una relación de pertenencia que haga verdaderos los axiomas de una teoría de conjuntos más fuerte que la teoría básica B que estudiamos en en el capítulo III. Esto nos permitirá a su vez hablar de sucesiones finitas de números naturales.

5.5.1 Máximo común divisor

Empezamos con algunos resultados que, de hecho, se demuestran en $\mathrm{I}\Delta_0$. Por ejemplo, la fórmula $d \mid x \equiv \bigvee u \le x\, y = du$ es Δ_0, al igual que $d \mid x \wedge d \mid y$, lo que, en virtud del teorema 5.26 nos permite definir el máximo común divisor de dos números en $\mathrm{I}\Delta_0$:

Definición 5.32 Definimos el *máximo común divisor* de dos números como[6]

$$(x,y) \equiv d \mid (x \neq 0 \wedge y \neq 0 \wedge d \mid x \wedge d \mid y \wedge \bigwedge e(e \mid x \wedge e \mid y \to e \le d)$$

Es obvio que $(x,y) = (y,x)$, y que $(x,x) = x$. Notemos que, por el convenio sobre las descripciones impropias, $(x,0) = 0$.

Seguidamente demostramos lo que en álgebra se conoce como relación de Bezout:

Teorema 5.33 $\bigwedge xy(0 < y \le x \to \bigvee u \le x \bigvee v \le x\ (x,y) = xu - yv)$.

Demostración: Observemos que $x = x \cdot 1 - y \cdot 0 > 0$, con $1 \le x \wedge 0 \le x$. Por lo tanto, podemos tomar el mínimo número z que cumple

$$\phi(z) \equiv \bigvee u \le x \bigvee v \le x(z = xu - yv \wedge z > 0).$$

Veamos que $z = xu - yv = (x,y)$. Por construcción $z \le x$, y también se cumple que $z \le y$ porque $y = xy - y(x-1)$, con $y \le x \wedge x - 1 \le x$.

Para probar que $z \mid x$ dividimos $x = zs + t$ con $t < z$ (teorema 5.11). Si fuera $t \neq 0$ entonces $t = x - zs$. No puede ser $s = 0$, pues entonces sería $x < z$. Entonces

$$t = x - zs = x - (xu - yv)s = x + qyx - (qyx + xus - yvs)$$

$$= x + xqy - xus - (yqx - yvs) = x(1 + qy - us) - y(qx - vs),$$

donde q es cualquier número natural suficientemente grande como para que $1 + qy - us > 0$ y $qx - vs > 0$. Tomemos[7] concretamente el mínimo q que cumpla $qx > vs$ (notemos que $q = vs + 1$ cumple la propiedad, luego hay un mínimo). Vamos a ver que también cumple el segundo requisito. Ante todo, tiene que ser $q > 0$ y además $qx - vs \le x$, pues en caso contrario $(q-1)x > vs$, en contra de la minimalidad de q. Operando de forma similar a la anterior obtenemos que

$$t + xus + y(qx - vs) = x(1 + qy),$$

de donde $xus < x(1 + qy)$, luego $1 + qy - us > 0$, como queríamos probar. Más aún, se cumple que $1 + qy - us \le y \le x$, pues si fuera $1 + qy - us > y$,

$$t = x(1 + (q-1)y - us) + y(vs - (q-1)x) \ge x,$$

donde usamos que $(q-1)x \le vs$ por la elección de q. Pero $t < z \le x$, luego tenemos una contradicción que prueba que $1 + qy - us \le x$, y entonces t contradice la minimalidad de z. Esto prueba que $t = 0$ y que $z \mid x$.

[6] Notemos que la fórmula $d = (x,y)$ es Δ_0, pues $\bigwedge e$ puede cambiarse por $\bigwedge e \le x$.

[7] Notemos que las dos veces que hemos tomado un mínimo, lo hemos hecho respecto una fórmula Δ_0.

Similarmente, dividimos $y = zs + t$ con $t < z$, razonamos como antes que $s \neq 0$ y, en el supuesto de que $t \neq 0$, expresamos

$$t = y - zs = y - (xu - yv)s = x(yq - us) - y(xq - vs - 1).$$

Para que esto tenga sentido tomamos el mínimo q tal que $yq > us$, con lo que $yq - us \leq y$, y en tal caso tiene que ser $vs + 1 \leq xq$, pues en caso contrario podríamos escribir

$$t = x(yq - us) + y(vs + 1 - xq) \geq x \geq z,$$

y tendríamos una contradicción. Por otra parte, $xq - vs - 1 \leq x$, pues en caso contrario

$$xy < y(xq - vs - 1) = x(yq - us) - t < x(yq - us),$$

luego $y < yq - us$ y $us < y(q - 1)$, en contradicción con la elección de q. Nuevamente tenemos que t contradice la minimalidad de z.

En definitiva, $z \mid x \wedge z \mid y$, y la expresión $z = xu - yv$ implica que cualquier otro divisor común de x e y divide a z. ∎

Notemos que del teorema anterior se sigue (y, de hecho, se ha puesto de manifiesto al final de la demostración) que

$$\bigwedge xye(x \neq 0 \wedge y \neq 0 \wedge e \mid x \wedge e \mid y \to e \mid (x, y)),$$

que es algo ligeramente más fuerte que lo que exige la definición.

Teorema 5.34 $\bigwedge xyz((x, y) = 1 \wedge x \mid yz \to x \mid z)$.

Demostración: Por el teorema anterior, $1 = xu - yv$, y si $x \mid yz$ entonces $x \mid xuz - yvz = z$. ∎

Con esto podemos definir el concepto de número primo y demostrar sus propiedades básicas, aunque todavía no estamos en condiciones de hablar de descomposiciones en factores primos.

Definición 5.35 $\text{Primo}(x) \equiv x > 1 \wedge \bigwedge u \leq x(u \mid x \to u = 1 \vee u = x)$.

Nota Hemos puesto un $u \leq x$ redundante en la definición de primo para poner de manifiesto que la fórmula $\text{Primo}(x)$ es Δ_0, al igual que $u \mid x$, o que $d = (x, y)$ o $z = \langle x, y\rangle$. ∎

Observemos que las relaciones definidas en $\mathbb{N}$ (en el sentido de 5.28) por todos los conceptos aritméticos que estamos introduciendo son las usuales, en el sentido de que, por ejemplo, la relación $P(n)$ dada por $\mathbb{N} \vDash \text{Primo}(0^{(n)})$ es la que se cumple exactamente cuando el número n es primo en el sentido usual. Normalmente dejaremos que el lector se persuada por sí mismo de estos hechos sin más comentarios.

Ejemplo Ahora podemos enunciar en $\mathcal{L}_a$ la conjetura de Goldbach:

$$\bigwedge x(2 \mid x \wedge x \geq 4 \to \bigvee pq \leq x(\mathrm{Primo}(p) \wedge \mathrm{Primo}(q) \wedge x = p + q))$$

y observamos que es una sentencia de tipo Π_1, pues la fórmula que sigue al $\bigwedge x$ es de tipo Δ_0. De acuerdo con el teorema 5.29, si la conjetura de Goldbach fuera falsa, esto podría ser demostrado en Q (es decir, si hay un número par n que no es suma de dos primos, en Q podríamos probar que $0^{(n)}$ no es suma de dos primos), pero si es verdadera no tenemos garantía de que pueda probarse en Q (ni en AP). ∎

Teorema 5.36 *Se cumple:*

1. $\bigwedge x(x > 1 \to \bigvee p\,(\mathrm{Primo}(p) \wedge p \mid x)$.
2. $\bigwedge pxy(\mathrm{Primo}(p) \wedge p \mid xy \to p \mid x \vee p \mid y)$.

Demostración: 1) Tenemos que $u = x$ cumple $u > 1 \wedge u \mid x$. Sea p el mínimo que cumple esto, es decir, $p > 1 \wedge p \mid x \wedge \bigwedge v(v > 1 \wedge v \mid x \to p \leq v)$. Basta probar que p es primo, pero si $v \mid p \wedge v > 1$ entonces $v \mid x$, luego $p \leq v \leq p$, luego $v = p$.

2) Si $p \nmid y$, entonces $(p, y) = 1$, pues se trata de un divisor de p, que tiene que ser 1 o p, y no puede ser p. Por el teorema anterior $p \mid x$. ∎

De momento no podemos ir más lejos. Por ejemplo, no podemos demostrar que todo número se descompone en producto de primos porque no tenemos forma de expresarlo, porque en AP, en principio, sólo podemos hablar de números, y no de sucesiones de números $\{p_i\}_{i=0}^n$ como sería necesario para afirmar que un número m se expresa como $m = p_1 \cdots p_n$. Sin embargo, esta limitación es sólo aparente, como veremos en breve.

5.5.2 Exponenciación

Ahora vamos a dar el primer paso para demostrar que en AP es posible hablar de conjuntos. Se trata de hacer algo parecido a lo que ya hemos hecho con los pares ordenados. Hemos visto que cada número codifica (o puede verse como) un par de números, e igualmente veremos que cada número codifica (o puede verse como) un conjunto. La idea es que los elementos de un número n serán los números correspondientes a las posiciones de los unos en la expresión binaria de n. Por ejemplo, si $n = 37 = 100101_2$ codifica el conjunto $\{0, 2, 5\}$.

El problema es que para definir esta codificación necesitamos definir al menos la exponencial 2^n, y para definir la exponencial necesitaríamos tener ya definida la codificación de conjuntos. Para resolver este inconveniente, introduciremos ahora una codificación de conjuntos rudimentaria, pero suficiente para definir la exponencial en base 2. Con ella definiremos luego una codificación de conjuntos mucho más potente.

No estamos todavía en condiciones de definir el factorial de un número natural, pero de momento nos basta con lo siguiente:

Teorema 5.37 $\bigwedge x \bigvee y \bigwedge u(0 < u \leq x \to u \mid y)$.

DEMOSTRACIÓN: Por inducción[8] sobre x. Para $x = 0$ es trivial y si vale para x, tomamos un y divisible entre todo $0 < u \leq x$ y observamos que $y(x+1)$ cumple el teorema para $x + 1$. ∎

Definición 5.38 Llamamos[9] $\mathrm{fc}(x) \equiv \min y \mid \bigwedge u \leq x(u > 0 \to u \mid y)$.

Con esto estamos en condiciones de definir una relación de pertenencia rudimentaria (más adelante la sustituiremos por otra mejor):

Definición 5.39 Abreviaremos[10] $u \in_0 (y, z) \equiv (1 + (1 + u)z) \mid y$.

Así, cada par de números naturales determina un conjunto, de tal forma que, como mostramos a continuación, todo conjunto finito definido por una fórmula está determinado por un par de números:

Teorema 5.40 *Sea $\phi(u)$ una fórmula, que puede tener otras variables libres. Entonces la fórmula siguiente es un teorema de* AP *(de $\mathrm{I}\Sigma_1$ si ϕ es Δ_0):*

$$\bigwedge x \bigvee yz \bigwedge u < x(u \in_0 (y, z) \leftrightarrow \phi(u))$$

DEMOSTRACIÓN: Si $x = 0$ el resultado es trivial, y si $x = 1$, basta tomar $y = z = 1$ si $\neg\phi(0)$ o bien $y = z = 0$ si $\phi(0)$. Por lo tanto, podemos suponer que $x > 1$. Sea $z = \mathrm{fc}(x)$.

Veamos ahora que si $u < v < x$, entonces $(1+(1+u)z, 1+(1+v)z) = 1$. En efecto, sea c el máximo común divisor. Abreviaremos $u_1 = 1 + u$, $v_1 = 1 + v$, de modo que $1 + u_1 z = ac$, $1 + v_1 z = bc$, para ciertos a, b. Notemos que $0 < v - u < x$, luego $v - u \mid z$. Por otra parte

$$1 + u_1 z \mid abc = a(1 + v_1 z), \qquad 1 + u_1 z \mid a(1 + u_1 z),$$

luego $1 + u_1 z$ también divide a la resta:

$$1 + u_1 z \mid a(v_1 - u_1)z = a(v - u)z.$$

Como claramente $(1 + u_1 z, z) = 1$, por 5.34 concluimos que

$$1 + u_1 z \mid a(v - u) \mid az.$$

Por el mismo argumento, $1 + u_1 z \mid a$, pero $1 + u_1 z = ac$, luego $1 + u_1 z = a$ y $c = 1$, como había que probar.

Ahora probamos lo siguiente por inducción[11] sobre t:

$$\bigwedge t \leq x \bigvee y(\bigwedge u < x((u < t \wedge \phi(u) \to u \in_0 (y, z)$$
$$\wedge\, (u \geq t \vee \neg\phi(u) \to (y, 1 + (1 + u)z) = 1)).$$

[8] Notemos que la propiedad es Σ_1.
[9] Notemos que la fórmula $y = \mathrm{fc}(x)$ es Δ_0.
[10] Esta definición es Δ_0, pues equivale a $\bigvee st \leq y(y = st \wedge s = 1 + (1 + u)z)$.
[11] Notemos que la fórmula es Σ_1 si ϕ es Δ_0

Para $t = 0$ basta tomar $y = 1$. Supongamos el resultado cierto para t y sea y el número correspondiente. Podemos suponer que $t+1 \le x$, o de lo contrario no hay nada que probar. Si $\neg\phi(t)$, entonces el mismo y cumple el resultado para $t+1$. Supongamos, pues, $\phi(t)$ y llamemos $y' = y(1+(1+t)z)$. Entonces $t \in_0 (y', z)$, y si $u < t \wedge \phi(u)$ entonces $u \in_0 (y', z)$.

Tomamos por último un $u < x$ tal que $u \ge t+1 \vee \neg\phi(u)$. Sabemos entonces que $(y, 1+(1+u)z) = 1$, y queremos probar lo mismo con y'. También sabemos que $(y, 1+(1+t)z) = 1$ y hemos probado que $1+(1+u)z$ y $1+(1+t)z$ son primos entre sí.

Entonces, si $(y', 1+(1+u)t) \neq 1$, tomemos un divisor primo p. Como $p \mid y'$, tiene que ser $p \mid y$ o bien $p \mid (1+(1+t)z)$, lo que contradice que $1+(1+u)z)$ sea primo con estos dos números.

Con esto termina la inducción, y tomando $t = x$ resulta la fórmula del enunciado. ∎

Definición 5.41 $\mathrm{Suc}(y,z,x) \equiv \bigwedge u < x \bigvee v < y\, \langle u,v\rangle \in_0 (y,z)$. Si se cumple esto, para cada $u < x$ llamaremos $(y,z)_u \equiv \min v < y \mid \langle u,v\rangle \in_0 (y,z)$.

$$\mathrm{SucExp}(y,z,x) \equiv x \ge 1 \wedge \mathrm{Suc}(y,z,x) \wedge (y,z)_0 = 1$$

$$\wedge \bigwedge u < x-1\ (y,z)_{u+1} = 2(y,z)_u.$$

$$\exp(x,v) \equiv \bigvee yz(\mathrm{SucExp}(y,z,x+1) \wedge (y,z)_x = v).$$

La fórmula[12] $\mathrm{Suc}(y,z,x)$ significa que los números y, z codifican una sucesión de longitud x, la que a cada $u < x$ le asigna $(y,z)_u$. La fórmula $\mathrm{SucExp}(y,z,x)$ significa que y, z codifican una sucesión cuyo primer término es 1 y que cada cual se obtiene del anterior multiplicándolo por 2, de modo que se trata de la sucesión $1, 2, 4, 8, \ldots, 2^x$. Por lo tanto, $\exp(x,v)$ significa que $v = 2^x$.

Teorema 5.42 $\bigwedge x \overset{1}{\bigvee} v \exp(x,v)$.

Demostración: Si $\mathrm{SucExp}(y,z,x) \wedge \mathrm{SucExp}(y',z',x') \wedge x \le x'$, entonces $\bigwedge u < x\ (y,z)_u = (y',z')_u$. Esto se prueba fácilmente por inducción[13] sobre u. Si $u = 0$, entonces $(y,z)_0 = 1 = (y',z')_0$. Si es cierto para u y $u+1 < x$, entonces

$$(y,z)_{u+1} = 2(y,z)_u = 2(y',z')_u = (y',z')_{u+1}.$$

[12] Observemos que

$$\mathrm{Suc}(y,z,x) \leftrightarrow \bigwedge u < x \bigvee v w < y(w = \langle u,v\rangle \wedge w \in_0 (y,z)),$$

$$v = (y,z)_u \leftrightarrow \bigvee w \le y(w = \langle u,v\rangle \wedge w \in_0 (y,z) \wedge \bigwedge v' < v(\neg \bigvee w \le y \cdots)),$$

$$\mathrm{SucExp}(y,z,x) \equiv x \ge 1 \wedge \mathrm{Suc}(y,z,x) \wedge \bigvee w \le y(w = (y,z)_0 \wedge w = 1) \wedge$$

$$\bigwedge u < x-1 \bigvee w w' \le y(w = (y,z)_{u+1} \wedge w' = (y,z)_u \wedge w = 2w').$$

Vemos así que las tres fórmulas son Δ_0.

[13] la fórmula es Δ_0

Esto nos da la unicidad. Veamos ahora que si $\mathrm{SucExp}(y,z,x)$ entonces $u<x$ implica $u<(y,z)_u$. Por inducción[14] sobre u. Para $u=0$ es $0<1=(y,z)_0$. Si es cierto para u, es decir, si $u<(y,z)_u$ y $u+1<x$, entonces

$$u+1\le 1\cdot(y,z)_u<2(y,z)_u=(y,z)_{u+1}.$$

Similarmente, si $u<v<x$, entonces $(y,z)_u<(y,z)_v$. Por inducción[15] sobre v. Si $v=0$ no hay nada que probar. Si vale para v y se cumple $u<v+1$, entonces $u\le v$. Si $u=v$ tenemos que

$$(y,z)_u<2(y,z)_u=(y,z)_{v+1}.$$

Si $u<v$ tenemos $(y,z)_u<(y,z)_v<2(y,z)_v=(y,z)_{v+1}$.

Pasemos ya a probar la existencia. Para ello probamos por inducción[16] sobre x que $\bigwedge x\ge 1\bigvee yz\ \mathrm{SucExp}(y,z,x)$.

En efecto, para $x=0$ es trivial. Si vale para x, tratamos aparte el caso $x=0$ (es decir, probamos aparte que el resultado vale para $x=1$). Para ello aplicamos 5.40, que nos da:

$$\bigvee yz\bigwedge u<2(u\in_0(y,z)\leftrightarrow u=1).$$

Teniendo en cuenta que $\langle 0,0\rangle=0$ y $\langle 0,1\rangle=1$, esto significa que si tomamos y, z que cumplan esto, se cumple $\mathrm{Suc}(y,z,1)$ y $(y,z)_0=1$, luego $\mathrm{SucExp}(y,z,1)$.

Supongamos ahora que $x\ge 1$ y que existen y, z tales que $\mathrm{SucExp}(y,z,x)$. Sea $q=(y,z)_{x-1}$ y sea $q'=\langle x,2q\rangle$. Por 5.40 existen y',z' tales que

$$\bigwedge u\le q'(u\in_0(y',z')\leftrightarrow(u\in_0(y,z)\wedge\bigvee i<x\bigvee v\ x=\langle i,v\rangle)\vee u=\langle x,2q\rangle).$$

Se cumple que $\mathrm{Suc}(y',z',x+1)$. En efecto, si $i<x+1$, o bien $i<x$, en cuyo caso, por la monotonía que hemos probado,

$$u=\langle i,(y,z)_i\rangle\le\langle x,q\rangle\le q',$$

luego $u\in_0(y',z')$, o bien $i=x$, en cuyo caso $u=\langle x,2q\rangle=q'$ también cumple $u\in_0(y',z')$, luego en cualquier caso $\bigwedge i<x+1\bigvee v\,\langle i,v\rangle\in_0(y',z')$. Más aún, si $\langle i,v\rangle\in_0(y',z')$ con $i<x$, tiene que ser $v=(y,z)_i$, pues en caso contrario $v<(y,z)_i$, pero entonces $\langle i,v\rangle\le\langle i,(y,z)_i\rangle\le q'$, luego $\langle i,v\rangle\in_0(y,z)$, pero entonces $(y,z)_i\le v$, contradicción. En otros términos,

$$\bigwedge i<x\ (y',z')_i=(y,z)_i.$$

Similarmente se ve que $(y',z')_x=2q=2(y,z)_{x-1}$, y de aquí se sigue inmediatamente que $\mathrm{SucExp}(y',z',x+1)$. Esto completa la inducción y, para cada x, tenemos que existen y,z tales que $\mathrm{SucExp}(y,z,x+1)$, luego $\bigvee v\exp(x,v)$. ∎

[14] La fórmula es $u<x\to\bigvee v<y(u<v\wedge v=(y.z)_u)$, luego es Δ_0.
[15] La fórmula es $\bigvee ww'\le y(w=(y,z)_u\wedge w'=(y,z)_v\wedge w<w')$, luego es Δ_0.
[16] La fórmula es Σ_1.

Definición 5.43 Definimos[17] $2^x \equiv v \mid \exp(x, v)$.

Claramente se cumple:

$$2^0 = 1 \wedge \bigwedge x(2^{x+1} = 2^x \cdot 2).$$

Más aún, en la prueba del teorema anterior hemos visto también las dos primeras de las tres propiedades siguientes (y la tercera se prueba sin problemas por inducción[18] sobre y):

$$\bigwedge x\ x < 2^x, \qquad \bigwedge uv(u < v \to 2^u < 2^v), \qquad \bigwedge xy\ 2^{x+y} = 2^x \cdot 2^y.$$

Teorema 5.44 $\bigwedge xy \overset{1}{\bigvee} uvw(v \leq 1 \wedge w < 2^x \wedge y = 2^{x+1} \cdot u + 2^x \cdot v + w)$.

Demostración: Por la división euclídea, existen u y $q < 2^{x+1}$ tales que $y = 2^{x+1} \cdot u + q$. A su vez, existen v y $w < 2^x$ tales que $q = 2^x v + w$, pero tiene que ser $v < 2$, pues en caso contrario $q \geq 2^{x+1}$. Esto nos da la existencia. Si tenemos

$$2^{x+1} \cdot u + 2^x \cdot v + w = 2^{x+1} \cdot u' + 2^x \cdot v' + w',$$

con $v, v' \leq 1$, entonces $w = w'$, pues ambos son el resto de la división euclídea entre 2^x, luego

$$2^{x+1} \cdot u + 2^x \cdot v = 2^{x+1} \cdot u' + 2^x \cdot v'$$

y de aquí $2u + v = 2u' + v'$, pero entonces $v = v'$, pues ambos son el resto de la división euclídea entre 2, luego $u = u'$. ∎

El número v dado por el teorema anterior es la cifra que ocupa la posición x en la expresión binaria de y o, simplemente, el bit x-ésimo de y:

Definición 5.45 Abreviaremos:[19]

$$\mathrm{bit}(x, y) \equiv v \mid \bigvee uw(v \leq 1 \wedge w < 2^x \wedge y = 2^{x+1} \cdot u + 2^x \cdot v + w).$$

5.5.3 La relación de pertenencia

Ahora estamos en condiciones de definir conjuntos en $\mathrm{I}\Sigma_1$ a través de una función de pertenencia mucho más adecuada que la que habíamos definido en la sección anterior. Consideramos la fórmula siguiente:[20]

Definición 5.46 $x \in y \equiv \mathrm{bit}(x, y) = 1, \qquad x \notin y \equiv \mathrm{bit}(x, y) = 0.$

Tal y como indicábamos más arriba, $x \in y$ significa que la cifra de orden x del desarrollo binario de y es 1.

[17] Tenemos que el término 2^x es Σ_1, porque $v = 2^x$ equivale a $\exp(x, v)$, que es claramente Σ_1, luego 2^x es Δ_1 por la observación tras la definición 5.19.

[18] La fórmula es Δ_1, pues resulta de introducir los términos 2^{x+y}, 2^x, 2^y (de tipo Δ_1) en la fórmula $u = vw$ (trivialmente Δ_1).

[19] Se trata de un término Δ_1, pues $v = \mathrm{bit}(x, y)$ resulta de introducir los términos 2^{x+1} y 2^x (de tipo Δ_1) en una fórmula Σ_1.

[20] Claramente es Δ_1.

Ejemplo Como $10 = 2^3 + 2$, tenemos que los únicos elementos de 10 son los números $1, 3$, de modo que podemos ver al número natural 10 como el conjunto $\{1, 3\}$. A su vez, $1 = 2^0$ es $1 = \{0\}$ y $3 = 2 + 1$ es $3 = \{0, 1\}$ y a su vez $0 = \varnothing$, luego, en total:

$$10 = \{1, 3\} = \{\{0\}, \{0, 1\}\} = \{\{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}.$$

■

Vamos a comprobar que esta relación de pertenencia cumple los axiomas necesarios para poder identificar a cada número natural y con el conjunto de los números naturales x que cumplen $x \in y$.

La primera propiedad del teorema siguiente implica que los cuantificadores de la forma $\bigwedge x \in y$ o $\bigvee x \in y$ están acotados, pues equivalen, respectivamente, a $\bigwedge x < y(x \in y \to \cdots)$ y $\bigvee x < y(x \in y \wedge \cdots)$, luego conservan el tipo Σ_1 o Π_1 de la fórmula que les sigue.

Teorema 5.47 *Se cumple:*

1. $\bigwedge xy(x \in y \to x < y)$
2. $\bigwedge y(\bigwedge x\ x \notin y \leftrightarrow y = 0)$
3. $\bigwedge yu(y < 2^u \leftrightarrow \bigwedge x(x \in y \to x < u))$

Demostración: 1) Si $x \in y$ entonces $\mathrm{bit}(x, y) = 1$, luego $y \geq 2^x > x$. Esto implica que $\bigwedge x\ x \notin 0$, que es una implicación de 2). Para probar la otra, si $y \neq 0$, se cumple que $2^0 = 1 \leq y$ y si $2^x \leq y$ entonces $x < y$, luego existe el máximo[21] x tal que $2^x \leq y$. Sea $w < 2^x$ tal que $y = 2^x \cdot v + w$. Tiene que ser $v \leq 1$, pues en caso contrario $2^{x+1} \leq y$. Resulta entonces que $\mathrm{bit}(x, y) = 1$, luego $x \in y$.

3) Si $y < 2^u$ y $x \in y$, entonces $2^x \leq y < 2^u$, luego $x < u$. Si $y \geq 2^u$, sea $x \geq u$ el máximo tal que $2^x \leq y$. En la prueba de 2) hemos visto que $x \in y$.

■

Necesitamos varios resultados técnicos. El primero (cuya interpretación es clara) lo usaremos varias veces en los razonamientos posteriores:

Teorema 5.48 $\bigwedge wz(w < 2^z \to \bigwedge x(x \in 2^z + w \leftrightarrow x \in w \vee x = z))$.

Demostración: Suponemos $w < 2^z$. Si $x = z$ es claro que $x \in 2^z + w$, por definición. Si $x \in w$, entonces $2^x \leq w < 2^z$, luego $x < z$ y $w = 2^{x+1}s + 2^x + t$, con $t < 2^x$, luego

$$2^z + w = 2^z + 2^{x+1}s + 2^x + t = 2^{x+1}(2^{z-x-1} + s) + 2^x + t,$$

luego $x \in 2^z + w$.

[21] La fórmula $2^x \leq y$ es Δ_1, pues se obtiene introduciendo el término 2^x (de tipo Δ_1) en la fórmula $u \leq y$.

Supongamos ahora que $x \in 2^z + w$, de modo que $2^z + w = 2^{x+1}s + 2^x + t$, con $t < 2^x$. Como $2^z + w < 2^z + 2^z = 2^{z+1}$, el teorema anterior nos da que $x < z+1$, o equivalentemente, $x \leq z$. Si $x = z$ hemos terminado. Supongamos que $x < z$ y veamos que $x \in w$. Para ello probamos que $2^{x+1}s \geq 2^z$. En caso contrario $2^{x+1}s < 2^z = 2^{x+1}2^{z-x-1}$, de donde $s < 2^{z-x-1}$, luego $s + 1 \leq 2^{z-x-1}$, luego $2^{x+1}(s+1) \leq 2^z$, luego $2^{x+1}s \leq 2^z - 2^{x+1}$, luego

$$2^z + w = 2^{x+1}s + 2^x + t < 2^z - 2^{x+1} + 2^x + 2^x = 2^z,$$

contradicción. Por lo tanto $s \geq 2^{z-x-1}$. Por otro lado, $s < 2^{z-x}$, ya que en otro caso $2^{x+1}s \geq 2^{z+1}$ y

$$2^{z+1} < 2^{z+1} + 2^x + t \leq 2^{x+1}s + 2^x + t = 2^z + w < 2^{z+1},$$

contradicción. Así pues, $2^{z-x-1} \leq s < 2^{z-x}$. Esto significa que en la división euclídea $s = 2^{z-x-1}u + q$, con $q < 2^{z-x-1}$, no puede ser ni $u = 0$ y $u > 1$, luego $u = 1$ y $s = 2^{z-x-1} + q$. Por lo tanto

$$2^z + w = 2^{x+1}s + 2^x + t = 2^z + 2^{x+1}q + 2^x + t$$

y $w = 2^{x+1}q + t$, luego $x \in w$. ∎

Definición 5.49 Abreviaremos[22] $x \subset y \equiv \bigwedge u(u \in x \to u \in y)$.

Observemos que $2^5 - 1 = 11111_2$, luego, como conjunto, contiene a todos los números menores que 5. Por lo tanto:

Teorema 5.50 $\bigwedge uy(y < 2^u \leftrightarrow y \subset 2^u - 1)$

Demostración: Teniendo en cuenta 5.47 3), basta probar que

$$\mathrm{bit}(x, 2^u - 1) = \begin{cases} 1 & \text{si } x < u, \\ 0 & \text{si } x \geq u. \end{cases}$$

Ciertamente, si $\mathrm{bit}(x, 2^u - 1) = 1$, entonces $2^u - 1 = 2^{x+1}s + 2^x + t$, luego $2^x \leq 2^u - 1 < 2^u$, luego $x < u$. Falta probar que

$$\bigwedge u \bigwedge x < u \ \mathrm{bit}(x, 2^u - 1) = 1.$$

Lo probamos por inducción sobre u. Para $u = 0$ es trivial. Supuesto cierto para u, observamos que $2^{u+1} - 1 = 2^u + (2^u - 1)$ y basta aplicar el teorema anterior. ∎

Teorema 5.51 *Sea $\phi(u)$ una fórmula con tal vez otras variables libres (de tipo Δ_1 para que la prueba valga en $\mathrm{I}\Sigma_1$). Entonces*

$$\bigwedge x \bigvee y < 2^x \bigwedge u < x(u \in y \leftrightarrow \phi(u)).$$

[22] Por 5.47 tenemos que $x \subset y \leftrightarrow \bigwedge u < x(u \in x \to u \in y)$, luego es Δ_1.

DEMOSTRACIÓN: Por inducción[23] sobre x. El caso $x = 0$ es trivial. Supongamos que el teorema vale para x y consideremos el y correspondiente. Si se cumple $\phi(x)$ tomamos $y' = 2^x + y$ y aplicamos 5.48. Si $u < x + 1$

$$u \in y' \leftrightarrow u \in y \vee u = x \leftrightarrow \phi(u).$$

Si $\neg\phi(x)$, basta tomar $y' = y$. ∎

Teorema 5.52 $\bigwedge uv(u \subset v \to u \leq v)$.

DEMOSTRACIÓN: Diremos que y es una *restricción* de z a x si[24]

$$R(y, z, x) \equiv y < 2^x \wedge \bigwedge u < x(u \in y \leftrightarrow u \in z).$$

Aplicando el teorema anterior a $\phi(u) \equiv u \in z$ concluimos que existen restricciones de z a todo x. Supongamos $u \subset v$ y veamos por inducción[25] sobre x que

$$\bigwedge u'v' < 2^x(R(u', u, x) \wedge R(v', v, x) \to u' \leq v').$$

Si $x = 0$ es trivial porque necesariamente $u' = v' = 0$. Si es cierto para x, tomemos ahora $u'', v'' < 2^{x+1}$ tales que $R(u'', u, x + 1) \wedge R(v'', v, x + 1)$. Sea

$$u' = \begin{cases} u'' - 2^z & \text{si } x \in u, \\ u'' & \text{si } x \notin u, \end{cases} \qquad v' = \begin{cases} v'' - 2^z & \text{si } x \in v, \\ v'' & \text{si } x \notin v. \end{cases}$$

Se cumple entonces que $R(u', u, x) \wedge R(v', v, x)$. En efecto, si $x \in u$ entonces $x \in u''$ y, como $u'' < 2^{x+1}$, tiene que ser $u'' = 2^x + u'$, mientras que si $x \notin u$ entonces $x \notin u''$ y $u'' < 2^x$. Usando 5.48 en el primer caso, se sigue fácilmente $R(u', u, x)$, y lo mismo vale para v'. Por hipótesis de inducción $u' \leq v'$. Ahora, si $x \notin u$, entonces $u'' = u' \leq v' \leq v''$, mientras que si $x \in u$ entonces $x \in v$, luego $u'' = 2^x + u' \leq 2^x + v' = v''$.

Finalmente observamos que u es una restricción de u a u, e igualmente con v, luego tenemos que $u \leq v$. ∎

Como consecuencia inmediata:

Teorema 5.53 (Extensionalidad) $\bigwedge uv(\bigwedge i(i \in u \leftrightarrow i \in v) \to u = v)$.

Teorema 5.54 (Especificación) *Si $\phi(u)$ es cualquier fórmula, en* AP *(o en* $\mathrm{I}\Sigma_1$ *si ϕ es Δ_1) se demuestra que*

$$\bigwedge x \overset{1}{\bigvee} y \bigwedge u(u \in y \leftrightarrow u \in x \wedge \phi(u)).$$

DEMOSTRACIÓN: La existencia se sigue inmediatamente de 5.51, y la unicidad del teorema anterior. ∎

[23] Notemos que sólo podemos asegurar que la coimplicación es Σ_1 si ϕ es Δ_1.
[24] La fórmula es claramente Δ_1, pues resulta de introducir 2^x en una fórmula Δ_1.
[25] La fórmula de la inducción es claramente Δ_1.

Definición 5.55 $\{u \in x \mid \phi(u)\} \equiv y \mid \bigwedge u(u \in y \leftrightarrow u \in x \wedge \phi(u))$.

Hemos probado que se trata siempre de una descripción propia[26] en AP, y en $\mathrm{I}\Sigma_1$ si ϕ es Δ_1

Si aplicamos 5.51 a la fórmula $u = u$ obtenemos $\bigwedge x \bigvee y < 2^x \bigwedge u < x\ u \in y$ o, equivalentemente,

$$\bigwedge x \bigvee y \bigwedge u(u \in y \leftrightarrow u < x).$$

Por la extensionalidad tenemos la unicidad, luego podemos definir:[27]

$$I_x \equiv y \mid \bigwedge u(u \in y \leftrightarrow u < x),$$

de modo que I_x es el conjunto de los números naturales menores que x.

Con esto estamos en condiciones de probar que la relación de pertenencia cumple los axiomas de una teoría de conjuntos bastante más rica que la teoría (básica) de Zermelo que estudiamos en el capítulo III:

Teorema 5.56 *Se cumple:*[28]

1) **Extensionalidad** $\bigwedge xy(\bigwedge u(u \in x \leftrightarrow u \in y) \to x = y)$
2) **Par** $\bigwedge xy \bigvee z \bigwedge u(u \in z \leftrightarrow u = x \vee u = y)$
3) **Unión** $\bigwedge x \bigvee y \bigwedge u(u \in y \leftrightarrow \bigvee v(u \in v \wedge v \in x))$
4) **Reemplazo** *Para toda fórmula $\phi(x, y)$, tal vez con más variables libres,*

$$\bigwedge xyz(\phi(x, y) \wedge \phi(x, z) \to y = z) \to \bigwedge a \bigvee b \bigwedge y(y \in b \leftrightarrow \bigvee x \in a\, \phi(x, y))$$

5) **Partes** $\bigwedge x \bigvee y \bigwedge u(u \in y \leftrightarrow u \subset x)$
6) **Regularidad** $\bigwedge x(\bigvee y\, y \in x) \to \bigvee y(y \in x \wedge \neg \bigvee u(u \in y \wedge u \in x))$

Demostración: 1) ya lo hemos probado. Para 2) tomamos $z = 2^x + 2^y$ si $x \neq y$, o bien $z = 2^x$ si $x = y$.

Para 3) basta considerar $y = \{u \in I_x \mid \bigvee v(u \in v \wedge v \in x)\}$ y observar que si $u \in v \wedge v \in x$ entonces $u < v < x$, luego $u \in I_x$ y $u \in y$.

4) Consideramos la fórmula $\psi(x, y) \equiv \phi(x, y) \vee (\neg \bigvee z\, \phi(x, z) \wedge y = 0)$. Es claro que $\bigwedge x < a \overset{1}{\bigvee} y\, \psi(x, y)$, luego por el principio de recolección 5.22 existe un v tal que $\bigwedge x < a \bigvee y < v\, \psi(x, y)$. A continuación tomamos

$$b = \{y \in I_v \mid \bigvee x \in a\, \phi(x, y)\}.$$

[26] Pero notemos que un término $\{u \in t \mid \phi(u)\}$ es una descripción propia en $\mathrm{I}\Sigma_1$ para cualquier término t (no necesariamente Δ_1). Ello se debe a que en la fórmula del teorema de especificación siempre podemos eliminar el generalizador $\bigwedge x$ sustituyendo a x por t.

[27] El término I_x es Δ_1, pues $y = I_x$ equivale a

$$\bigwedge u < y(u \in y \to u < x) \wedge \bigwedge u < x\ u \in y.$$

[28] Todos los apartados se demuestran de hecho en $\mathrm{I}\Sigma_1$, salvo 4).

De este modo, si $\bigvee x \in a\, \phi(x,y)$, existe un $y' < v$ tal que $\psi(x,y')$, pero por definición de ψ tiene que ser $\phi(x,y')$, y por la unicidad $y = y' \in I_v$, luego $y \in b$.

5) Basta tomar $y = \{u \in I_{x+1} \mid u \subset x\}$, pues si $u \subset x$ entonces $u \leq x$, luego $u \in I_{x+1}$, luego $u \in y$.

6) Basta tomar $y = \min x|x \in y$. ∎

Nota Las fórmulas del teorema anterior (consideradas como fórmulas del lenguaje $\mathcal{L}_{\mathrm{tc}}$) son los axiomas de la teoría de conjuntos de Zermelo-Fraenkel (ZF), que estudiaremos en el capítulo X. En realidad dicha teoría tiene un axioma más, el axioma de infinitud (AI), que postula la existencia de conjuntos infinitos.

Notemos que el esquema de especificación no figura entre los axiomas de ZF, pero ello se debe a que es demostrable a partir del esquema de reemplazo —la fórmula 5) del teorema anterior—, pero aquí no necesitamos comprobar esto porque ya hemos probado que el esquema de especificación es demostrable en AP (teorema 5.54).

Por otro lado, la teoría de conjuntos ZFC, de la que hemos hablado en la introducción, se obtiene de ZF añadiendo un axioma más, el axioma de elección, y veremos[29] que, visto como sentencia de $\mathcal{L}_a$, también es demostrable en AP o incluso en $\mathrm{I}\Sigma_1$.

La conclusión es que si α es un teorema de ZFC−AI, entonces, visto como fórmula de $\mathcal{L}_a$ (es decir, interpretando $x \in y$ como la fórmula definida en 5.46), es un teorema de AP. Simplemente, una demostración en ZFC−AI se interpreta sin cambio alguno como una demostración en AP, pues los axiomas de ZFC−AI son teoremas de AP. En particular, la consistencia de AP implica la consistencia de ZFC−AI.

En términos semánticos, tenemos que $\mathbb{N} \vDash \mathrm{ZFC} - \mathrm{AI}$ sin más que tomar como interpretación del relator $\in$ la relación aritmética definida por la fórmula $x \in y$ de $\mathcal{L}_a$. ∎

5.5.4 Un teorema de recursión

Como primera aplicación de los resultados precedentes demostramos en $\mathrm{I}\Sigma_1$ un resultado similar al teorema 3.26 con el que definiremos expresiones como las potencias m^n o el factorial $n!$ de un número natural:

Teorema 5.57 *Sean $x \in X$ y $\psi(n,x,y)$ dos fórmulas Σ_1 (tal vez con más variables libres) tales que*

$$\bigwedge nx(x \in X \to \overset{1}{\bigvee} y(y \in X \wedge \psi(n,x,y))),$$

sea a tal que $a \in X$. Entonces existe una fórmula $\chi(a,n,x)$ (con las mismas variables adicionales que tengan las fórmulas dadas) de tipo Σ_1 tal que $\bigwedge n \overset{1}{\bigvee} x\chi(a,n,x)$ y, si llamamos $F(n) \equiv x|\chi(a,n,x)$ y $G(n,x) \equiv y|\psi(n,x,y)$, entonces

$$\bigwedge n\ F(n) \in X \wedge F(0) = a \wedge \bigwedge n\ F(n+1) = G(n,F(n)).$$

[29] Véase la nota tras el teorema 6.28.

DEMOSTRACIÓN: Definimos

$$\chi(a,n,x) \equiv \bigvee s(s : I_{n+1} \longrightarrow X \wedge s(n) = x \wedge s(0) = a$$
$$\wedge \bigwedge i < n\ \psi(i, s(i), s(i+1))).$$

Se trata de una fórmula Σ_1, pues es equivalente a[30]

$$\bigvee s(\bigwedge u \le n \bigvee y \le s(y \in X \wedge \langle u, y\rangle \in s)$$
$$\wedge \bigwedge u \le n \bigwedge yz \le s(\langle u, y\rangle \in s \wedge \langle u, z\rangle \in s \to y = z)$$
$$\wedge \bigwedge w \le s(w \in s \to \bigvee u \le n \bigvee y \le s\ w = \langle u, y\rangle) \wedge \langle n, x\rangle \in s \wedge \langle 0, a\rangle \in s$$
$$\wedge \bigwedge i < n \bigvee yz \le s(\langle i, y\rangle \in s \wedge \langle i+1, z\rangle \in s \wedge \psi(i, y, z))).$$

Una simple inducción[31] prueba que $\bigwedge n \bigvee x\ (x \in X \wedge \chi(a, n, x))$. En efecto, para $n = 0$ basta tomar $s = \{\langle 0, a\rangle\}$ y se cumple $\chi(a, 0, a)$. Si $x \in X \wedge \chi(a, n, x)$, tomamos s según la definición de χ. Como $x \in X$, existe un único $y \in X$ tal que $\psi(n, x, y)$, con lo que $s^* = s \cup \{\langle n+1, y\rangle\}$ cumple la definición de $\chi(a, n+1, y)$.

En segundo lugar probamos que $\chi(n, x) \wedge \chi(n, y) \to x = y$. En efecto, sean s y s^* según la definición de χ, de modo que $s(n) = x \wedge s^*(n) = y$. Veamos por inducción[32] sobre i que $\bigwedge i(i \le n \to s(i) = s^*(i))$. Para $i = 0$ es $s(i) = a = s^*(i)$, y si vale para i, entonces $\psi(i, s(i), s(i+1)) \wedge \psi(i, s^*(i), s^*(i+1))$, luego por la unicidad $s(i+1) = s^*(i+1)$. En particular $x = s(n) = s^*(n) = y$.

Con esto tenemos probada la unicidad de χ y que $\bigwedge n\ F(n) \in X$. También hemos visto que $\chi(a, 0, a)$, luego $F(0) = a$. Por último, si s prueba que se cumple $\chi(a, n+1, F(n+1))$, sea $s^* = s \setminus \{\langle (n+1, F(n+1))\rangle\}$. Es fácil ver que s^* cumple la definición de $\chi(a, n, s(n))$, por lo que $s(n) = F(n)$ y, como sabemos que $\psi(n, s(n), s(n+1)$, resulta que $\psi(n, F(n), F(n+1))$, es decir, se cumple que $F(n+1) = G(n, F(n))$. ∎

Si aplicamos esto a $\phi(x) \equiv x = x$ y $\psi(m, n, x) \equiv x \cdot m$ obtenemos un término $F(n)$ de tipo Σ_1, luego Δ_1, tal que, si lo representamos por m^n, cumple

$$\bigwedge m\ m^0 = 1 \wedge \bigwedge mn\ m^{n+1} = m^n \cdot m.$$

A partir de aquí se demuestran sin dificultad las propiedades de las potencias:

$$\bigwedge mnp\ m^{n+p} = m^n \cdot m^p, \quad \bigwedge mnp\ (mn)^p = m^p \cdot n^p, \quad \bigwedge mnp\ (m^n)^p = m^{np},$$
$$\bigwedge n \ge 1\ 0^n = 0, \qquad \bigwedge n\ 1^n = 1,$$
$$\bigwedge mnp(m < n \wedge p \ge 1 \to m^p < n^p), \quad \bigwedge mnp(n < p \wedge m > 1 \to m^n < m^p),$$

etc. Notemos que la exponencial 2^n que ya teníamos definida es un caso particular de la que acabamos de definir.

Si aplicamos el teorema anterior a la fórmula $\psi(n, x) \equiv x \cdot (n+1)$ obtenemos un término $n!$ de tipo Δ_1 que satisface las propiedades siguientes:

$$0! = 1 \wedge \bigwedge n\ (n+1)! = n! \cdot (n+1).$$

[30] En el capítulo siguiente probaremos resultados generales para agilizar el cálculo de la complejidad de fórmulas como ésta.

[31] La fórmula es claramente Σ_1.

[32] La fórmula equivale a $i \le n \to \bigvee u \le s(\langle i, u\rangle \in s \wedge \langle i, u\rangle \in s^*)$, luego es Σ_1.

5.6 Sucesiones finitas

En esta sección veremos que en $\mathrm{I}\Sigma_1$ es posible identificar cada sucesión finita de números naturales con un número natural. En 5.14 hemos visto cómo identificar cada par de números naturales x, y con un único número natural $\langle x, y\rangle$, y en 5.16 hemos definido las componentes de cada número natural visto como par. Similarmente, podemos definir

$$\langle x, y, z\rangle_3 \equiv \langle\langle x, y\rangle_2, z\rangle_2$$

y las proyecciones $z_1^3 \equiv (z_1)_1$, $z_2^3 \equiv (z_1)_2$, $z_3^3 \equiv z_2$, de modo que

$$z = \langle z_1, z_2, z_3\rangle, \qquad z = \langle u, v, w\rangle_3 \to u = z_1 \wedge v = z_2 \wedge w = z_3.$$

Con esto podemos identificar los números naturales con las ternas de números naturales. En general, podemos definir

$$\langle x_1, \ldots, x_n\rangle_n \equiv \big\langle\langle x_1, \ldots, x_{n-1}\rangle_{n-1}, x_n\big\rangle_2$$

y las proyecciones[33]

$$\begin{aligned} p_i^n(z) &= p_i^{n-1}(z_1), \qquad \text{para } i < n\\ p_n^n(z) &= z_2. \end{aligned}$$

Si definimos $\langle x\rangle_1 = x$, estas definiciones coinciden con las que ya teníamos en el caso $n = 2$.

Razonando por inducción (metamatemática) sobre n, se prueba sin dificultad que

$$z = \langle p_1^n(z), \ldots, p_n^n(z)\rangle_n, \quad z = \langle x_1, \ldots, x_n\rangle_n \to x_1 = z_1 \wedge \cdots \wedge x_n = z_n.$$

Ejemplo Un simple cálculo nos da que

$$\langle 3, 1\rangle_2 = 13, \quad \langle 3, 1, 5\rangle_3 = \langle 13, 5\rangle_2 = 184,$$

de modo que el número 184 puede verse como él mismo, o como el par $\langle 13, 5\rangle_2$ o como la terna $\langle 3, 1, 5\rangle_3$. ∎

En general, cada número natural n puede verse indistintamente como un número (él mismo), como un par de números, o una terna, o una cuádrupla, etc. Por ejemplo, el número $n = 17\,337\,210$ puede verse como

$$17\,337\,210, \quad (5\,882, 5), \quad (104, 3, 5), \quad (13, 0, 3, 5), \quad (3, 1, 0, 3, 5),$$

$$(0, 2, 1, 0, 3, 5), \quad (0, 0, 2, 1, 0, 3, 5), \quad (0, 0, 0, 2, 1, 0, 3, 5), \ldots$$

[33] Notemos que no hemos definido un único término $\langle x_1, \ldots, x_n\rangle_n$ con $n+1$ variables libres, sino infinitos términos con n variables libres. Igualmente, la definición de las proyecciones no es la definición de un único término $p_i^n(z)$ con 3 variables libres, sino que estamos definiendo infinitos términos con una variable libre, cada uno con su propia definición.

Para evitar que un mismo número pueda verse como muchas cosas a la vez, podemos definir una nueva familia de infinitos términos:

$$\langle x_1, \ldots, x_n\rangle_\infty^n = \langle n-1, \langle x_1, \ldots, x_n\rangle_n\rangle_2 + 1.$$

Así, cada número natural no nulo s codifica una única sucesión. Para calcularla, pasamos a $s-1$, interpretamos el resultado como un par y la primera componente $+1$ nos indica la longitud n de la sucesión codificada, y ésta es la segunda componente interpretada precisamente como sucesión de longitud n. Podemos considerar que el número natural 0 codifica la sucesión vacía, que tiene 0 términos.

Definición 5.58 Definimos la *longitud* de un número natural como

$$\ell(s) = (1-(1-s))((s-1)_1 + 1).$$

Aquí hay que tener presente que hemos definido $x-y$ como 0 cuando $x \leq y$. Por lo tanto, vemos que $\ell(0) = 0$, mientras que si $s \geq 1$ es $\ell(s) = (s-1)_1 + 1$.

Para definir las proyecciones definimos primero un término[34] Δ_1 mediante el teorema 5.57:

$$R(s,0) = (s-1)_2, \qquad R(s,i+1) = R(s,i)_1.$$

Así podemos definir a su vez el término Δ_1:

$$p_i^\infty(s) = (1-i)R(s,(s-1)_1) + (1-(1-i))R(s,(s-1)_1 - i)_2.$$

Notemos que $p_i^\infty(s)$ es un término con dos variables libres, es decir, que ahora no tenemos infinitos términos con una variable libre, sino que tanto i como s son variables.

Ejemplo Si $s = \langle 3,2,7\rangle_\infty = \big\langle 2, \langle\langle 3,2\rangle_2, 7\rangle_2\big\rangle_2 + 1$, tenemos que

$$R(s,0) = \langle\langle 3,2\rangle_2, 7\rangle_2, \quad R(s,1) = \langle 3,2\rangle_2, \quad R(s,2) = 3,$$

luego $p_0^\infty(s) = R(s,2) = 3$, $p_1^\infty(s) = R(s,1)_2 = 2$, $p_2^\infty(s) = R(s,0)_2 = 7$. ∎

Nota A partir de ahora escribiremos $s_i \equiv p_i^\infty(s)$. ∎

Veamos que todo número natural, visto como sucesión, está unívocamente determinado por sus proyecciones.

Teorema 5.59 $\ell(s) = \ell(t) \wedge (\bigwedge i < \ell(s)\, s_i = t_i) \to s = t$.

Demostración: Si $\ell(s) = \ell(t) = 0$, entonces $s = t = 0$. Supongamos que $\ell(s) = \ell(t) > 0$. Llamamos

$$l = (s-1)_1 = \ell(s) - 1 = \ell(t) - 1 = (t-1)_1.$$

[34]El teorema 5.57 nos da que $y = R(s,i)$ es una fórmula Σ_1, luego $R(s,i)$ es un término Δ_1, por la observación tras la definición 5.19.

Veamos por inducción que

$$i \leq l \to R(s, l-i) = R(t, l-i).$$

Para $i = 0$ tenemos que

$$R(s, l) = s_0 = t_0 = R(t, l).$$

Si es cierto para $i < l$, entonces $i + 1 \leq l$, luego $i + 1 + (l - (i + 1)) = l$, de donde $(l - (i + 1)) + 1 = l - i$. Por lo tanto,

$$R(s, l-(i+1))_1 = R(s, (l-(i+1))+1) = R(s, l-i) = R(t, l-i) = R(t, l-(i+1))_1.$$

Por otra parte,

$$R(s, l - (i + 1))_2 = s_{i+1} = t_{i+1} = R(t, l - (i + 1))_2,$$

luego $R(s, l-(i+1)) = R(t, l-(i+1))$. Esto completa la inducción y, aplicándolo a $i = l$ obtenemos

$$(s - 1)_2 = R(s, 0) = R(t, 0) = (t - 1)_2,$$

pero $(s - 1)_1 = l = (t - 1)_1$, luego de hecho $s - 1 = t - 1$, luego $s = t$. ∎

Ejemplo La tabla siguiente muestra la interpretación como sucesiones de los 40 primeros números naturales:

0		10	$0,0,0,0$	20	$0,0,0,0,1$	30	$0,3$
1	0	11	4	21	$0,0,0,0,0,0$	31	$1,0,0$
2	1	12	$0,2$	22	6	32	$0,0,1,1$
3	$0,0$	13	$0,1,0$	23	$2,0$	33	$0,0,0,0,2$
4	2	14	$0,0,0,1$	24	$0,1,1$	34	$0,0,0,0,1,0$
5	$0,1$	15	$0,0,0,0,0$	25	$0,0,0,2$	35	$0,0,0,0,0,0,1$
6	$0,0,0$	16	5	26	$0,0,0,1,0$	36	$0,0,0,0,0,0,0,0$
7	3	17	$1,1$	27	$0,0,0,0,0,1$	37	8
8	$1,0$	18	$0,0,2$	28	$0,0,0,0,0,0,0$	38	$1,2$
9	$0,0,1$	19	$0,0,1,0$	29	7	39	$0,0,3$

Por ejemplo, para calcular la sucesión asociada al número 352 889 465, como no es 0, le restamos 1 y lo interpretamos como par:

$$352\,889\,464 = \langle 3, 26\,563\rangle_2\,,$$

lo que nos indica que debemos interpretar el número 26 563 como una sucesión de longitud 4, la cual resulta ser $3, 2, 2, 1$. ∎

Una comprobación rutinaria muestra que

$$p_i^\infty(\langle x_0, \ldots, x_{n-1}\rangle_\infty^n) = x_i,$$

lo que prueba que toda sucesión finita de números naturales está codificada por un número natural. En la práctica podemos omitir el superíndice n en el término $\langle x_0, \ldots, x_{n-1}\rangle_\infty^n$, puesto que éste se deduce del número de variables.

No debemos confundir un número natural n con la sucesión de longitud 1 que lo tiene como único término. Ésta viene dada por $\langle n\rangle_\infty = \langle 0, n\rangle_2 + 1$.

Un último hecho general de interés sobre las proyecciones es el siguiente:

Teorema 5.60 $\ell(s) \le s$ *y* $\bigwedge i < \ell(s)\, s_i \le s$.

DEMOSTRACIÓN: Claramente $\ell(s) = (s-1)+1 \le s$. Para probar la segunda parte vemos primero por inducción que $s > 0 \to R(s,i) < s$. En efecto, para $i = 0$ es $R(s,0) = (s-1)_2 \le s - 1 < s$. Si es cierto para i, entonces

$$R(s, i+1) = R(s,i)_1 \le R(s,i) < s.$$

Ahora probamos por inducción que $s > 0 \wedge i < \ell(s) \to s_i < s$. Para $i = 0$ es

$$s_0 = p_0^\infty(s) = R(s, (s-1)_1) < s.$$

Si vale para i y $i + 1 < \ell(s)$, entonces

$$s_{i+1} = p_{i+1}^\infty(s) = R(s, (s-1)_1 - (i+1))_2 \le R(s, (s-1)_1 - (i+1)) < s.$$

Si $s = 0$ la conclusión es trivial. ∎

Definición 5.61 Diremos que una sucesión t *extiende* a otra s si

$$s \sqsubseteq t \equiv \ell(s) \le \ell(t) \wedge \bigwedge i < \ell(s)\, s_i = t_i.$$

Es fácil ver que:

1. $s \sqsubseteq s$,
2. $s \sqsubseteq t \wedge t \sqsubseteq s \to s = t$,
3. $s \sqsubseteq t \wedge t \sqsubseteq u \to s \sqsubseteq u$,
4. $0 \sqsubseteq s$.

Finalmente observamos que las manipulaciones básicas que podemos hacer con sucesiones pueden definirse aritméticamente:

Definimos el término Δ_1 dado por

$$s^\frown \langle n\rangle = (1-s)\langle n\rangle_\infty + (1-(1-s))(\langle \ell(s), \langle (s-1)_2, n\rangle_2\rangle_2 + 1).$$

Una comprobación rutinaria muestra que

$$\ell(s^\frown \langle n\rangle) = \ell(s) + 1, \quad s \sqsubseteq s^\frown \langle n\rangle, \quad (s^\frown \langle n\rangle)_{\ell(s)} = n.$$

Vemos así que $s^\frown \langle n\rangle$ no es sino la sucesión que resulta de añadir n como último término a la sucesión representada por s. Más en general, definimos

$$s^\frown t = F(s, t, \ell(t)),$$

donde F es la función definida con el teorema 5.57 mediante

$$F(s,t,0) = s, \qquad F(s,t,n+1) = F(s,t,n)^\frown \langle t_n \rangle .$$

De nuevo una comprobación rutinaria muestra que

$$\ell(s^\frown t) = \ell(s) + \ell(t), \quad \bigwedge i < \ell(s)(s^\frown t)_i = s_i, \quad \bigwedge i < \ell(t)\,(s^\frown t)_{\ell(s)+i} = t_i.$$

En particular, $s \sqsubseteq s^\frown t$.

Definimos la *restricción* de una sucesión mediante el funtor dado por

$$s|_0 = 0, \qquad s|_{i+1} = s|_i^\frown \langle s_i \rangle .$$

Es fácil probar que $i \le \ell(s) \to \ell(s|_i) = i \wedge s|_i \sqsubseteq s$. Más aún, si $t \sqsubseteq s$, entonces $t = s|_{\ell(t)}$.

Similarmente definimos $s|^i = F(s,i,\ell(s)-i)$, donde

$$F(s,i,0) = s_i, \qquad F(s,i,n+1) = F(s,i,n)^\frown \langle s_{i+n} \rangle .$$

Así, si $i \le \ell(s)$, se cumple que $\ell(s|^i) = \ell(s) - i$ y $s = s|_i^\frown s|^i$.

A partir de este momento ya no usaremos más los términos $\langle x_1, \ldots, x_n \rangle_n$ ni sus proyecciones correspondientes, sino que siempre que hablemos de sucesiones de números naturales consideraremos los términos $\langle x_1, \ldots, x_n \rangle_\infty$ y las proyecciones dadas por el término $p_i^\infty(s)$, con dos variables libres. Todos estos términos son de tipo Δ_1.

Capítulo VI

La teoría de Kripke-Platek

En el capítulo III introdujimos la teoría (restringida) de Zermelo (Z^*) y probamos que es una teoría aritmética, de modo que todo teorema de AP es demostrable en Z^*. Más precisamente, para cada fórmula γ de $\mathcal{L}_a$ podemos asociarle una "traducción" (que en 3.28 hemos llamado $\bar{\gamma}$, pero que para el caso de $\mathcal{L}_{\mathrm{tc}}$ representaremos más precisamente por γ_{tc}), de modo que si $\underset{\mathrm{AP}}{\vdash} \gamma$, entonces

$$\underset{\mathrm{Z}^*}{\vdash} \bigwedge x_1 \cdots x_n \in \omega\, \gamma_{\mathrm{tc}}.$$

La "traducción" consiste simplemente en sustituir el 0 y los funtores de $\mathcal{L}_a$ por sus definiciones en Z^* y cambiar los cuantificadores $\bigwedge x$ y $\bigvee x$ de $\mathcal{L}_a$ por $\bigwedge x \in \omega$ y $\bigvee x \in \omega$, respectivamente.

Por ejemplo, la traducción de $\bigwedge x\ x + 0 = x$ es $\bigwedge x \in \omega\ x + 0 = x$, donde $+$ es un funtor en la primera fórmula, mientras que en la segunda $x + y$ es la fórmula definida en Z^* por recurrencia, y 0 es una constante en la primera fórmula, mientras que en la segunda es $0 \equiv \varnothing \equiv x|\bigwedge u\ u \notin x$.

Recíprocamente, en el capítulo anterior hemos visto que en Z^* (de hecho en $\mathrm{I}\Sigma_1$) es posible definir una relación de pertenencia que satisface todos los axiomas de ZFC menos el axioma de infinitud (aunque hemos dejado pendiente definir —y demostrar— el axioma de elección).

En realidad aquí debemos hacer una precisión: todos los axiomas de ZFC$-$AI los hemos demostrado en $\mathrm{I}\Sigma_1$ excepto el axioma de reemplazo (y, en particular, el de especificación).[1] En este capítulo vamos a presentar una teoría de conjuntos que es a ZFC$-$AI lo que $\mathrm{I}\Sigma_1$ es a AP. Se trata de la teoría de conjuntos de Kripke-Platek (KP), en la que se pueden demostrar los axiomas de $\mathrm{I}\Sigma_1$ y, recíprocamente, en $\mathrm{I}\Sigma_1$ se pueden demostrar los axiomas de KP.

Para los propósitos de este libro podríamos limitarnos a trabajar con AP y con ZFC$-$AI, pero KP aparece en varios contextos cuando se profundiza en la teoría de conjuntos, por lo que es interesante conocerla.

[1] Ambos los hemos probado en AP, y el segundo lo hemos probado en $\mathrm{I}\Sigma_1$ para fórmulas Δ_0. En realidad, refinando los argumentos, podríamos probar ambos en $\mathrm{I}\Sigma_1$ para fórmulas Δ_1, y de hecho lo probaremos en este capítulo.

6.1 La jerarquía de Lévy

En 5.17 hemos definido las fórmulas Σ_1, Π_1 y Δ_1 de $\mathcal{L}_a$, que en realidad son únicamente los primeros niveles de una jerarquía de fórmulas aritméticas conocida como la jerarquía de Kleene. No la hemos introducido porque no la vamos a necesitar,[2] pero aquí vamos a introducir una jerarquía análoga para las fórmulas de $\mathcal{L}_{\mathrm{tc}}$.

Definición 6.1 Diremos que una fórmula α de $\mathcal{L}_{\mathrm{tc}}$ es Δ_0 si existe una sucesión de fórmulas $\alpha_0, \ldots, \alpha_m \equiv \alpha$ tal que cada α_i es de uno de los tipos siguientes:

1. $x = y$ o $x \in y$,
2. $\neg\beta$ o $\beta \to \gamma$, donde β y γ son fórmulas anteriores de la sucesión.
3. $\bigwedge x \in y\, \beta$, donde β es una fórmula anterior de la sucesión y la variable y es distinta de x.

Más llanamente, llamaremos fórmulas Δ_0 a las que no tienen descriptores y cuyos cuantificadores están todos acotados de la forma indicada.

De este modo, si α y β son fórmulas Δ_0 también lo son[3] $\neg\alpha$, $\alpha \to \beta$, $\alpha \vee \beta$, $\alpha \wedge \beta$, $\alpha \leftrightarrow \beta$, $\bigwedge x \in y\, \alpha$ y $\bigvee x \in y\, \alpha$.

Diremos que una fórmula de $\mathcal{L}_{\mathrm{tc}}$ es de tipo Σ_n (resp. Π_n), con $n \geq 1$, si es de la forma $\bigvee x_1 \bigwedge x_2 \cdots x_n \alpha$ (resp. $\bigwedge x_1 \bigvee x_2 \cdots x_n \alpha$), con n cuantificadores alternados, donde α es Δ_0.

Más en general, si T es una teoría axiomática en la que pueden probarse los axiomas de la teoría básica B, diremos que una fórmula cualquiera es de tipo Σ_n^T o Π_n^T si es equivalente en T a una fórmula del tipo correspondiente. Una fórmula es de tipo Δ_n^T si es a la vez Σ_n^T y Π_n^T. Normalmente omitiremos el superíndice T si está claro por el contexto en qué teoría estamos trabajando.

Siempre podemos añadir cuantificadores delante o detrás del prefijo de una fórmula con variables que no aparezcan en ella, y la fórmula resultante es equivalente, por lo que toda fórmula Σ_n o Π_n es Δ_{n+1}.

Observemos que la fórmula $u = (x, y)$ es Δ_0^{B}. En efecto:

$$u = \{x, y\} \leftrightarrow \bigwedge v \in u\ (v = x \vee v = y) \wedge x \in u \wedge y \in v,$$

$$u = (x, y) \leftrightarrow \bigvee vw \in u\ (v = \{x\} \wedge w = \{x, y\}) \wedge \bigwedge v \in u (u = \{x\} \vee u = \{x, y\}).$$

Notemos también que

$$\bigvee xy\ u = (x, y) \leftrightarrow \bigvee v \in u \bigvee xy \in v\ u = (x, y).$$

Teniendo esto en cuenta, demostramos:

[2] Véase la sección 4.3 de [LF].

[3] En realidad $\bigvee x \in y\, \alpha$ no es Δ_0, sino lógicamente equivalente a una fórmula Δ_0, pero esta sutileza será irrelevante.

Teorema 6.2 *Sea T una teoría axiomática que contenga a* B. *Para cada número natural $n \geq 1$ y para fórmulas α y β cualesquiera se cumple:*[4]

1. *Si α, β son Σ_n, lo mismo vale para $\bigvee x\alpha, \alpha \wedge \beta$ y $\alpha \vee \beta$.*
2. *Si α, β son Π_n, lo mismo vale para $\bigwedge x\alpha, \alpha \wedge \beta$ y $\alpha \vee \beta$.*
3. *Si α es Σ_n entonces $\neg\alpha$ es Π_n, y viceversa.*
4. *Si α es Π_n (resp. Σ_n) y β es Σ_n (resp. Π_n), $\alpha \to \beta$ es Σ_n (resp. Π_n).*
5. *Si α y β son Δ_n, también lo son*

$$\neg\alpha, \quad \alpha \wedge \beta, \quad \alpha \vee \beta, \quad \alpha \to \beta, \quad \alpha \leftrightarrow \beta.$$

DEMOSTRACIÓN: 1) Por simplicidad supondremos $n = 3$. El caso general es formalmente idéntico. Tenemos que $\alpha \leftrightarrow \bigvee x_1 \bigwedge x_2 \bigvee x_3\, \phi$, donde ϕ es Δ_0. Así

$$\bigvee x\alpha \leftrightarrow \bigvee x x_1 \bigwedge x_2 \bigvee x_3\, \phi \leftrightarrow \bigvee w \bigvee z \in w \bigvee x x_1 \in z(w = (x, x_1) \wedge \bigwedge x_2 \bigvee x_3\, \phi)$$

$$\leftrightarrow \bigvee w \bigwedge x_2 \bigvee x_3 \bigvee z \in w \bigvee x x_1 \in z(w = (x, x_1) \wedge \phi).$$

Si $\beta \leftrightarrow \bigvee y_1 \bigwedge y_2 \bigvee y_3\, \psi$, donde ψ es Δ_0 y las variables y_1, y_2, y_3 son distintas de x_1, x_2, x_3, entonces

$$\alpha \wedge \beta \leftrightarrow \bigvee x_1 y_1 \bigwedge x_2 y_2 \bigvee x_3 y_3(\phi \wedge \psi).$$

Ahora basta aplicar tres veces el caso ya probado y el correspondiente de 2), que se sigue del que hemos probado aplicando 3). Notemos que 3) es inmediato. El caso de $\alpha \vee \beta$ es idéntico.

Como ya hemos señalado, 3) es inmediato, y 2) se sigue de 1) por 3).

4) se sigue de los apartados anteriores porque $\alpha \to \beta$ equivale a $\neg\alpha \vee \beta$ y 5) es evidente. ∎

En cualquier teoría que contenga a B podemos hablar de clases propias en el sentido explicado en la subsección 3.2.2, de modo que la notación

$$A = \{x \mid \phi(x, x_1, \ldots, x_n)\}$$

significará que $x \in A$ es una abreviatura por $\phi(x, x_1, \ldots, x_n)$. En particular podemos hablar de clases Σ_n, Π_n o Δ_n según que la fórmula $x \in A$ sea del tipo correspondiente. En particular, diremos que una fórmula $\phi(x_1, \ldots, x_n, x)$ define una función $F : A_1 \times \cdots \times A_n \longrightarrow A$ si cumple

$$\bigwedge x_1 \in A_1 \cdots \bigwedge x_n \in A_n \overset{1}{\bigvee} x \in A\ \phi(x_1, \ldots, x_n, x).$$

En tal caso abreviaremos

$$F(x_1, \ldots, x_n) \equiv x|(x_1 \in A_1 \wedge \cdots \wedge x_n \in A_n \wedge \phi(x_1, \ldots, x_n, x)).$$

[4] Véase además el teorema 6.3, más abajo.

Se cumple que si las clases A_i son Δ_n y F es Σ_n (en el sentido de que lo es ϕ), con $n \geq 1$, entonces el término $F(x_1, \ldots, x_n)$ es Δ_n pues[5]

$$x = F(x_1, \ldots, x_n) \leftrightarrow (x_1 \in A_1 \wedge \cdots \wedge x_n \in A_n \wedge \phi(x_1, \ldots, x_n))$$

$$\vee ((x_1 \notin A_1 \vee \cdots x_n \notin A_n) \wedge x = \varnothing)$$

$$\leftrightarrow (x_1 \in A_1 \wedge \cdots \wedge x_n \in A_n \wedge \bigwedge y(\phi(x_1, \ldots, x_n, y) \to y = x)$$

$$\vee ((x_1 \notin A_1 \vee \cdots x_n \notin A_n) \wedge x = \varnothing).$$

Es claro que al sustituir términos Δ_n en fórmulas Σ_n o Π_n la fórmula resultante sigue siendo Σ_n o Π_n (por el mismo argumento empleado tras la definición 5.19).

Es pura rutina comprobar que los conceptos conjuntistas básicos definibles en la teoría básica B son Δ_0. Veamos algunos ejemplos:[6]

$$w = \{u, v\} \leftrightarrow v \in w \wedge v \in w \wedge \bigwedge x \in w\, (x = u \vee x = v),$$

$$w = (u, v) \leftrightarrow \bigvee rs \in w(r = \{u\} \wedge s = \{u, v\}) \wedge \bigwedge x \in w(x = \{u\} \vee x = \{u, v\}).$$

$$z = x \cup y \leftrightarrow \bigwedge u \in z\, (u \in x \vee u \in y) \wedge \bigwedge u \in x\, u \in z \wedge \bigwedge u \in y\, u \in z$$

$$z = x \cap y \leftrightarrow \bigwedge u \in z\, (u \in x \wedge u \in y) \wedge \bigwedge u \in x(u \in y \to u \in z),$$

$$z = x \setminus y \leftrightarrow \bigwedge u \in z\, (u \in x \wedge u \notin y) \wedge \bigwedge u \in x(u \notin y \to u \in z),$$

$$z = \varnothing \leftrightarrow \bigwedge u \in z\ u \neq u.$$

6.2 La teoría KP

La *teoría de conjuntos de Kripke-Platek* (KP) es la teoría axiomática sobre el lenguaje $\mathcal{L}_{\mathrm{tc}}$ cuyos axiomas son los siguientes:

Extensionalidad	$\bigwedge xy(\bigwedge u(u \in x \leftrightarrow u \in y) \to x = y)$
Par	$\bigwedge xy \bigvee z(x \in z \wedge y \in z)$
Unión	$\bigwedge x \bigvee y \bigwedge u \in x \bigwedge v \in u\ v \in y$
Δ_0-especificación	$\bigwedge x \bigvee y \bigwedge u(u \in y \leftrightarrow (u \in x \wedge \phi(u)))$ $(*)$
Δ_0-recolección	$\bigwedge u \bigvee v\ \phi(u, v) \to \bigwedge a \bigvee b \bigwedge u \in a \bigvee v \in b\ \phi(u, v)$ $(*)$
Π_1-regularidad	$\bigvee u\, \phi(u) \to \bigvee u\, (\phi(u) \wedge \bigwedge v \in u\, \neg\phi(v))$ $(**)$

$(*)$ Para toda fórmula ϕ (tal vez con más variables libres) de tipo Δ_0,
$(**)$ Para toda fórmula ϕ (tal vez con más variables libres) de tipo Π_1.

Veamos el papel que desempeña en esta teoría cada uno de sus tres esquemas de axioma. Empezamos por el primero: El esquema de Δ_0-especificación es una versión restringida del esquema de especificación de Z^*. Como en esta teoría, el

[5] Adoptamos el convenio de que $(x|x = x) = \varnothing$.

[6] En la sección A.3 del apéndice A hemos recogido las comprobaciones de que los principales conceptos conjuntistas son Δ_0 o Δ_1 en la jerarquía de Lévy.

axioma de extensionalidad nos da que el y cuya existencia postula es único, por lo que podemos considerar el conjunto

$$\{u \in x \mid \phi(u)\} \equiv y | \bigwedge u(u \in y \leftrightarrow (u \in x \wedge \phi(u)))$$

La única diferencia con Z^* es que ahora sólo tenemos garantizado que este término es una descripción propia cuando la fórmula ϕ es Δ_0. No obstante, esto basta para demostrar los axiomas del conjunto vacío y de la diferencia de la teoría B, pues, tomando un conjunto cualquiera x, el conjunto $\{u \in x \mid u \neq u\}$ es un conjunto vacío (donde tenemos en cuenta que la fórmula $u \neq u$ es Δ_0) y el conjunto $\{u \in y \mid u \notin x\}$ es la diferencia de los conjuntos x e y. Por lo tanto, todos los teoremas de B son también teoremas de KP. En particular tenemos definida la clase ω de los números naturales (pero no la suma y el producto, que hemos definido en Z^* y luego veremos que también son definibles en KP).

Del esquema de Π_1-regularidad extraemos dos consecuencias principales. La primera resulta de aplicarlo a la fórmula $\phi(u) \equiv u \in x$, lo que nos da que

$$\bigwedge x(x \neq \varnothing \to \bigvee u(u \in x \wedge u \cap x = \varnothing).$$

Como esto vale para todo conjunto, concluimos que todo conjunto está bien fundado en el sentido de 3.9, luego la definición de ordinal dada allí puede simplificarse en KP hasta

$$x \in \Omega \leftrightarrow \bigwedge u \in x\ u \subset x \wedge \bigwedge uv \in x(u \in v \vee v \in u \vee u = v).$$

Así, $x \in \Omega$ es una fórmula Δ_0 en KP (mientras que en B no lo es). De aquí se sigue que la fórmula $x \in \omega$ también es Δ_0.

Observemos que no podemos adaptar la prueba del teorema 3.25 para demostrar en KP el principio de Σ_1-inducción para los números naturales, porque el argumento define un subconjunto de ω mediante una fórmula que sería Π_1, cuando necesitaríamos que fuera Δ_0. Sin embargo, el esquema de Π_1-regularidad nos da un argumento alternativo. Es inmediato que equivale al resultado siguiente:

Principio de $\boldsymbol{\Sigma_1^\epsilon}$-inducción *Si $\phi(u)$ es una fórmula de clase Σ_1 (con posibles parámetros), entonces*

$$\bigwedge x(\bigwedge u \in x\ \phi(u) \to \phi(x)) \to \bigwedge x\, \phi(x).$$

Así, para demostrar que todo conjunto tiene una propiedad Σ_1 basta demostrar que un conjunto x la tiene bajo la hipótesis de inducción de que todos sus elementos la tienen. Si aplicamos esto a la fórmula

$$\Phi(x) \equiv (x \in \Omega \wedge \phi(x)) \vee x \notin \Omega$$

obtenemos:

Principio de $\boldsymbol{\Sigma_1}$-inducción transfinita *Si $\phi(u)$ es una fórmula Σ_1 (con posibles parámetros), entonces*

$$\bigwedge \alpha \in \Omega(\bigwedge \delta < \alpha\, \phi(\delta) \to \phi(\alpha)) \to \bigwedge \alpha \in \Omega\, \phi(\alpha).$$

Lo mismo vale (con la misma prueba) si cambiamos Ω por ω, de modo que para probar que todo número natural tiene una propiedad Σ_1 basta probar que uno la tiene supuesto que la tienen todos los menores que él. Finalmente, de aquí obtenemos la Σ_1-inducción usual:

Principio de Σ_1-inducción *Si $\phi(u)$ es una fórmula Σ_1 (con posibles parámetros), entonces*

$$\phi(0) \wedge \bigwedge n \in \omega(\phi(n) \to \phi(n')) \to \bigwedge n \in \omega\, \phi(n).$$

DEMOSTRACIÓN: Supongamos $\phi(0) \wedge \bigwedge n \in \omega(\phi(n) \to \phi(n'))$ pero que, a pesar de ello, $\bigvee n \in \omega \neg\phi(n)$. Como $n \in \omega \wedge \neg\phi(n)$ es Π_1, por el axioma de regularidad existe un $n \in \omega$ tal que $\neg\phi(n) \wedge \bigwedge u \in n\, \phi(n))$. No puede ser $n = 0$, porque $\phi(0)$, luego $n = m'$, para cierto $m < n$, pero entonces $\phi(m)$, y por hipótesis también $\phi(n)$, contradicción. ∎

Nota El mismo argumento empleado en 5.21 muestra que el principio de inducción anterior es válido también para fórmulas Π_1. ∎

Pasamos ahora al esquema de Δ_0-recolección. Conviene observar que es equivalente a la siguiente versión "local":

$$\bigwedge a(\bigwedge u \in a \bigvee v\, \phi(u,v) \to \bigvee b \bigwedge u \in a \bigvee v \in b\, \phi(u,v))$$

Es obvio que la versión "local" implica la "global". Para el recíproco, si suponemos $\bigwedge u \in a \bigvee v\, \phi(u,v)$, aplicamos la versión global a la fórmula Δ_0

$$\Phi(u,v) \equiv (u \in a \wedge \phi(u,v)) \vee (u \notin a \wedge v = u).$$

La consecuencia que más nos va a interesar es el análogo del teorema 5.23:

Teorema 6.3 *Si T es una extensión de* KP, *las clases de fórmulas Σ_1^T y Π_1^T son cerradas para $\bigwedge x \in y, \bigvee x \in y$.*

DEMOSTRACIÓN: Supongamos que ϕ_0 es equivalente a $\bigvee z\, \phi$, donde ϕ es Δ_0, entonces, usando el axioma de Δ_0-recolección:

$$\bigwedge x \in y\, \phi_0(x) \leftrightarrow \bigwedge x \in y \bigvee z\, \phi(x,z) \leftrightarrow \bigvee y' \bigwedge x \in y \bigvee z \in y'\, \phi(x,z),$$

y la última fórmula es Σ_1. La clausura de una fórmula Π_1 respecto de $\bigvee x \in y$ se obtiene aplicando la parte ya probada a su negación. ∎

Ahora vamos a ver que en KP se puede probar, entre otras cosas, la Δ_1-especificación y la Σ_1-recolección. Para ello demostraremos en primer lugar una versión más precisa del teorema anterior. Para ello necesitamos la definición siguiente:

Definición 6.4 Llamaremos fórmulas $\tilde{\Sigma}_1$ en $\mathcal{L}_{\mathrm{tc}}$ a las que cumplen los criterios siguientes:

1. Toda fórmula Δ_0 es $\tilde{\Sigma}_1$.
2. Si α y β son $\tilde{\Sigma}_1$, también lo son

$$\alpha \wedge \beta, \quad \alpha \vee \beta, \quad \bigwedge u \in x\, \alpha, \quad \bigvee u \in x\, \alpha, \quad \bigvee u\, \alpha.$$

Las fórmulas $\tilde{\Pi}_1$ son las que se construyen del mismo modo pero cambiando $\bigvee u\, \alpha$ por $\bigwedge u\, \alpha$.

El teorema anterior muestra que las fórmulas $\tilde{\Sigma}_1$ y $\tilde{\Pi}_1$ son Σ_1^{KP} y Π_1^{KP} en sentido amplio, respectivamente, pero vamos a dar una prueba alternativa que nos dará una fórmula explícita equivalente de tipo Σ_1 o Π_1.

Si ϕ es una fórmula $\tilde{\Sigma}_1$ y x es una variable que no esté en ϕ, llamamos $\phi^{(x)}$ a la fórmula Δ_0 que resulta de sustituir en ϕ cada cuantificador no acotado $\bigvee u$ por $\bigvee u \in x$.

El teorema siguiente se demuestra trivialmente por inducción sobre la longitud de la fórmula:

Teorema 6.5 *Si ϕ es una fórmula de clase $\tilde{\Sigma}_1$ y x, y son variables que no están en ϕ, entonces se cumple:*

$$\phi^{(x)} \wedge x \subset y \to \phi^{(y)}, \qquad \phi^{(x)} \to \phi.$$

La versión explícita de los teoremas 6.2 y 6.3 para fórmulas $\tilde{\Sigma}_1$ es el teorema siguiente, del que obtendremos muchas consecuencias:

Teorema 6.6 ($\tilde{\boldsymbol{\Sigma}}_\mathbf{1}$-reflexión) *Si ϕ es una fórmula de clase $\tilde{\Sigma}_1$ y x es una variable que no está en ϕ, la fórmula $\phi \leftrightarrow \bigvee u\, \phi^{(u)}$ es un teorema de* KP.

Demostración: Una implicación es inmediata por el teorema anterior. Probamos la contraria por inducción sobre la longitud de ϕ. Si ϕ es de clase Δ_0 entonces $\phi \equiv \phi^{(x)}$ y el resultado es trivial.

Si $\phi \equiv \psi_1 \wedge \psi_2$, entonces $\phi^{(x)} \equiv \psi_1^{(x)} \wedge \psi_2^{(x)}$. Por hipótesis de inducción tenemos que

$$\psi_1 \leftrightarrow \bigvee u\, \psi_1^{(u)}, \qquad \psi_2 \leftrightarrow \bigvee u\, \psi_2^{(u)}.$$

Si se cumple $\phi \equiv \psi_1 \wedge \psi_2$, entonces existen x_1 y x_2 tales que $\psi_1^{(x_1)}$ y $\psi_2^{(x_2)}$. Si llamamos $x = x_1 \cup x_2$ entonces tenemos $\psi_1^{(x)} \wedge \psi_2^{(x)}$ por el teorema anterior, luego $\bigvee u\, \phi^{(u)}$.

Si $\phi \equiv \psi_1 \vee \psi_2$ la equivalencia es trivial:

$$\phi \leftrightarrow \bigvee u\, \psi_1^{(u)} \vee \bigvee u\, \psi_2^{(u)} \leftrightarrow \bigvee u(\psi_1^{(u)} \vee \psi_2^{(u)}) \leftrightarrow \bigvee u\, \phi^{(u)}.$$

Si $\phi \equiv \bigvee v \in y\, \psi$, entonces

$$\phi \leftrightarrow \bigvee v \in y \bigvee u\, \psi^{(u)} \leftrightarrow \bigvee u \bigvee v \in y\, \psi^{(u)} \equiv \bigvee u\, \phi^{(u)}.$$

Si $\phi \equiv \bigwedge v \in y\, \psi$, entonces, por hipótesis de inducción, $\phi \leftrightarrow \bigwedge v \in y \bigvee u\, \psi^{(u)}$. Suponiendo ϕ, por Δ_0-recolección existe un w tal que $\bigwedge v \in y \bigvee u \in w\, \psi^{(u)}$. Sea $w' = \bigcup w$. Así $\bigwedge v \in y\, \psi^{(w')}$, luego $\bigvee u \bigwedge v \in y\, \psi^{(u)} \equiv \bigvee u\, \phi^{(u)}$.

Si $\phi \equiv \bigvee v\, \psi$ y suponemos ϕ, entonces sea x tal que $\psi(x)$. Por hipótesis de inducción existe un w tal que $\psi^{(w)}(x)$. Sea $w' = w \cup \{x\}$. Entonces también $\psi^{(w')}$, luego $\bigvee v \in w'\, \psi^{(w')} \equiv \phi^{(w')}$, luego $\bigvee u\, \phi^{(u)}$. ∎

Ahora ya podemos "mejorar" los axiomas de KP:

Teorema 6.7 ($\boldsymbol{\Sigma_1}$-recolección) *Para toda fórmula ϕ de clase Σ_1 (tal vez con más variables libres), la fórmula siguiente es un teorema de* KP:

$$\bigwedge u \in x \bigvee v\, \phi(u,v) \to \bigvee y(\bigwedge u \in x \bigvee v \in y\, \phi(u,v) \wedge \bigwedge v \in y \bigvee u \in x\, \phi(u,v)).$$

Demostración: Basta probarlo para fórmulas de clase $\tilde{\Sigma}_1$, pues en particular éstas incluyen a todas las fórmulas Σ_1 en sentido estricto, y a su vez, si el teorema se cumple para ellas, se cumple también para todas las fórmulas Σ_1 en sentido amplio (las equivalentes a fórmulas Σ_1 en sentido estricto).

Por $\tilde{\Sigma}_1$-reflexión, supuesto $\bigwedge u \in x \bigvee v\, \phi(u,v)$, existe un z tal que

$$\bigwedge u \in x \bigvee v \in z\, \phi^{(z)}(u,v).$$

Por Δ_0-especificación existe el conjunto

$$y = \{v \in z \mid \bigvee u \in x\, \phi^{(z)}(u,v)\},$$

y claramente cumple lo pedido. ∎

Teorema 6.8 ($\boldsymbol{\Sigma_1}$-reducción) *Si ϕ es una fórmula Π_1 y ψ es Σ_1, la fórmula siguiente es un teorema de* KP:

$$\bigwedge u \in x(\phi(u) \to \psi(u)) \to$$

$$\bigvee y(\bigwedge u \in x(\phi(u) \to u \in y) \wedge \bigwedge u \in y(u \in x \wedge \psi(u))).$$

Demostración: No perdemos generalidad si suponemos que ϕ es Π_1 en sentido estricto y ψ es Σ_1 en sentido estricto. Pongamos, concretamente, que $\phi \equiv \bigwedge v\, \phi'$. Supongamos

$$\bigwedge u \in x(\phi(u) \to \psi(u))$$

o, lo que es lo mismo, $\bigwedge u \in x(\bigvee v \neg\phi'(u) \vee \psi(u))$. Esta fórmula es $\tilde{\Sigma}_1$, luego por $\tilde{\Sigma}_1$-reflexión existe un z tal que $\bigwedge u \in x(\phi^{(z)}(u) \to \psi^{(z)}(u))$. Basta tomar $y = \{u \in x \mid \psi^{(z)}(u)\}$, que existe por Δ_0-especificación. Así, si $u \in x$ cumple $\phi(u)$, pero $u \notin y$, entonces $\neg\psi^{(z)}(u)$, pero esto implica $\neg\phi^{(z)}(u)$, luego $\neg\phi(u)$, y tenemos una contradicción, luego $u \in y$. ∎

De aquí se sigue inmediatamente:

Teorema 6.9 ($\boldsymbol{\Delta_1}$-especificación) *Si ϕ es una fórmula Π_1 y ψ es Σ_1, la fórmula siguiente es un teorema de* KP:

$$\bigwedge u \in x(\phi(u) \leftrightarrow \psi(u)) \to \bigvee y \bigwedge u(u \in y \leftrightarrow u \in x \wedge \psi(u)).$$

Observemos que lo que afirma el teorema anterior es que las fórmulas Δ_1^T, para cualquier teoría T que extienda a KP, también definen subconjuntos de un conjunto dado. Así pues,

$$\{u \in x \mid \phi(u)\}$$

es una descripción propia siempre que ϕ es una fórmula Δ_1.

De aquí podemos deducir un último resultado general de interés, pero para ello necesitamos probar antes algunos hechos básicos, empezando por la existencia de productos cartesianos:

Teorema 6.10 $\overset{1}{\bigvee} z \bigwedge w(w \in z \leftrightarrow \bigvee u \in x \bigvee v \in y\ w = (u,v))$.

Demostración: La unicidad es consecuencia del axioma de extensionalidad. Basta probar la existencia. Aplicamos Δ_0-recolección a la fórmula $\phi(u,w) \equiv w = (u,v)$, y obtenemos un b tal que $\bigwedge u \in x\ (u,v) \in b$. Ahora consideramos la fórmula Δ_0

$$\phi(v,t) \equiv \bigwedge u \in x \bigvee w \in t\ w = (u,v).$$

Así, $\phi(v,t)$ afirma que t contiene todos los pares (u,v) con $u \in x$. Acabamos de probar que para todo v existe un tal t, luego podemos aplicar de nuevo el axioma de recolección, que nos da un b tal que

$$\bigwedge v \in y \bigvee t \in b \bigwedge u \in x\,(u,v) \in t.$$

Ahora tomamos $a = \bigcup b$, de modo que si $u \in x \wedge v \in y$, entonces existe un $t \in b$ tal que $(u,v) \in t$, luego $(u,v) \in a$. Basta tomar

$$z = \{w \in a \mid \bigvee u \in x \bigvee v \in y\ w = (u,v)\}.$$

∎

Por lo tanto:

$$x \times y \equiv z \mid \bigwedge w(w \in z \leftrightarrow \bigvee u \in x \bigvee v \in y\, w = (u,v))$$

es una descripción propia. Si $\phi(u,v)$ es una fórmula Δ_1, podemos considerar el conjunto

$$\{(u,v) \in x \times y \mid \phi(u,v)\} \equiv \{w \in x \times y \mid \bigvee u \in x \bigvee v \in y\ (w = (u,v) \wedge \phi(u,v))\},$$

pues la fórmula que lo especifica es también Δ_1.

A partir de aquí ya es fácil probar que todos los conceptos conjuntistas básicos definen conjuntos. Conviene observar en general que

$$\bigwedge x \in \bigcup y\, \alpha \leftrightarrow \bigwedge u \in y \bigwedge x \in u\ \alpha, \qquad \bigvee x \in \bigcup y\, \alpha \leftrightarrow \bigvee u \in y \bigvee x \in u\ \alpha,$$

por lo que estas acotaciones de cuantificadores no aumentan el tipo de una fórmula. Esto es útil porque $u, v \in \{u, v\} \in (u, v)$, luego $u, v \in \bigcup(u, v)$. Teniendo esto en cuenta vemos que podemos definir como sigue el dominio y el rango de un conjunto:

$$\mathcal{D}f = \{u \in \bigcup\bigcup f \mid \bigvee w \in f \bigvee v \in \bigcup w\ w = (u, v)\},$$
$$\mathcal{R}f = \{v \in \bigcup\bigcup f \mid \bigvee w \in f \bigvee u \in \bigcup w\ w = (u, v)\},$$
$$f^{-1} = \{(v, u) \in \mathcal{R}f \times \mathcal{D}f \mid (u, v) \in f)\},$$
$$f \circ g = \{(u, w) \in \mathcal{D}f \times \mathcal{R}g \mid \bigvee v \in \bigcup\bigcup f\ (u, v) \in f \wedge (v, w) \in g\},$$

etc., de modo que

$$x \in \mathcal{D}f \leftrightarrow \bigvee y\,(x, y) \in f, \qquad y \in \mathcal{D}f \leftrightarrow \bigvee x\,(x, y) \in f,$$
$$(v, u) \in f^{-1} \leftrightarrow (u, v) \in f, \qquad (u, w) \in f \circ g \leftrightarrow \bigvee v((u, v) \in f \wedge (v, w) \in g),$$

etc. También es fácil comprobar que todos estos conceptos son Δ_0 respecto de la jerarquía de Lévy. Véase la sección A.3 del apéndice A. Ahora ya podemos probar:

Teorema 6.11 (Σ_1-reemplazo) *Para toda fórmula ϕ de tipo Σ_1 se cumple:*

$$\bigwedge u \in x \overset{1}{\bigvee} v\,\phi(u, v) \to \bigvee f y(f : x \longrightarrow y \text{ suprayectiva} \wedge \bigwedge u \in x\,\phi(u, f(u))).$$

Demostración: Supongamos que $\bigwedge u \in x \overset{1}{\bigvee} v\,\phi(u, v)$. Por Σ_1-recolección, existe un z tal que $\bigwedge u \in x \bigvee v \in z\,\phi(u, v)$ y, claramente, de hecho,

$$\bigwedge u \in x \overset{1}{\bigvee} v \in z\,\phi(u, v).$$

Ahora observamos que, para todo $w \in x \times z$ es equivalente

$$\bigvee uv(w = (u, v) \wedge \phi(u, v)) \leftrightarrow \neg\bigvee uvv'(w = (u, v) \wedge v \neq v' \wedge \phi(u, v')),$$

y ambas fórmulas son Σ_1 y Π_1 respectivamente, luego por Δ_1-especificación existe el conjunto

$$f = \{(u, v) \in x \times z \mid \phi(u, v)\},$$

y cumple lo pedido con $y = \mathcal{R}x$. ∎

6.3 KP como teoría aritmética

Analicemos la demostración (en Z^*) del teorema 3.26. En primer lugar, por el teorema 6.3, la fórmula $\chi(n, x, a)$ mostrada explícitamente en la nota posterior es Σ_1 si lo son las fórmulas ϕ y ψ, y en segundo lugar observamos que todas las inducciones que se realizan en la prueba son respecto de fórmulas Σ_1. La conclusión es que en KP se demuestra el siguiente teorema de recursión (compárese también con 5.57):

Teorema 6.12 *Sea X una clase, sea $G : \omega \times X \longrightarrow X$ y sea $a \in X$. Si X y G son de tipo Σ_1, existe una función $F : \omega \longrightarrow X$ de tipo Σ_1 tal que*

$$F(0) = a \wedge \bigwedge n \in \omega\ F(n') = G(n, F(n)).$$

Esto basta para definir en KP la suma y el producto de números naturales, pues para la suma podemos aplicar el teorema anterior a la clase $X = \omega$ (que es de tipo Δ_0) y a la función $G(i, n) = m \leftrightarrow m = n'$ (también Δ_0). Esto nos da el término $m + n$ de tipo Σ_1 y, por consiguiente, Δ_1.

A su vez, para definir el producto aplicamos el teorema de nuevo a $X = \omega$ y a la función $G(i, n) = r \leftrightarrow r = n + m$, que es Σ_1, luego el producto $m \cdot n$ es un término Δ_1. En resumen, el teorema siguiente se cumple sin más que tomar $x \in \mathbb{N} \equiv x \in \omega$ (ya hemos visto que los primeros apartados son teoremas de B):

Teorema 6.13 (Axiomas de Peano) *Existe una fórmula $x \in \mathbb{N}$, y términos x', $x + y$, $x \cdot y$, 0 del lenguaje $\mathcal{L}_{\mathrm{tc}}$, todos ellos de tipo Δ_1, tales que en* KP *se demuestra:*

1. $0 \in \mathbb{N}$
2. $\bigwedge n \in \mathbb{N}\ n' \in \mathbb{N}$,
3. $\bigwedge mn \in \mathbb{N}\, m + n \in \mathbb{N}$,
4. $\bigwedge mn \in \mathbb{N}\, m \cdot n \in \mathbb{N}$,
5. $\bigwedge n \in \mathbb{N}\, n' \neq 0$,
6. $\bigwedge mn \in \mathbb{N}(m' = n' \to m = n)$,
7. $\bigwedge m \in \mathbb{N}\ m + 0 = m$,
8. $\bigwedge mn \in \mathbb{N}\ m + n' = (m + n)'$,
9. $\bigwedge m \in \mathbb{N}\ m \cdot 0 = 0$,
10. $\bigwedge mn \in \mathbb{N}\ m \cdot n' = m \cdot n + m$,
11. $\phi(0) \wedge \bigwedge n \in \mathbb{N}(\phi(n) \to \phi(n')) \to \bigwedge n \in \mathbb{N}\, \phi(n)$, *para toda fórmula $\phi(x)$ de tipo Σ_1, tal vez con más variables libres.*

Así pues, KP interpreta a $\mathcal{L}_a$ y cumple la definición de teoría aritmética dada en 3.28 salvo por que el principio de inducción está limitado a fórmulas de tipo Σ_1. También hay que destacar que, en principio, nos referimos a fórmulas de tipo Σ_1 en la jerarquía de Lévy, no a fórmulas de tipo Σ_1 en el sentido aritmético definido en 5.17, pero enseguida vamos a ver que éstas son un caso particular de aquéllas.

En efecto, en primer lugar, llamando θ_{tc} a la traducción de una expresión de $\mathcal{L}_a$ a $\mathcal{L}_{\mathrm{tc}}$, el teorema 3.29 nos da que

$$\bigwedge x_1 \cdots x_n \in \mathbb{N}\, t_{\mathrm{tc}} \in \mathbb{N}.$$

Si t es un término sin descriptores de $\mathcal{L}_a$, una simple inducción sobre la longitud de t prueba que t_{tc} es un término Δ_1 en KP (puesto que todos los conceptos aritméticos están definidos en KP mediante términos y fórmulas Δ_1).

Por lo tanto, si t_1 y t_2 son términos sin descriptores de $\mathcal{L}_a$,

$$(t_1 = t_2)_{\mathrm{tc}} \equiv t_{1\mathrm{tc}} = t_{2\mathrm{tc}}, \qquad (t_1 \leq t_2)_{\mathrm{tc}} \equiv \bigvee z \in \omega\ z + t_{1\mathrm{tc}} = t_{2\mathrm{tc}}$$

son fórmulas Σ_1 en KP.

De aquí se sigue inmediatamente que la traducción de toda fórmula abierta de $\mathcal{L}_a$ es una fórmula Σ_1 en KP, luego la traducción de un caso del principio de inducción correspondiente a una fórmula abierta es un caso de Σ_1-inducción en KP, luego es un teorema de KP. El teorema 3.30 nos da entonces que en KP se demuestran las traducciones de todos los teoremas de IA. En particular, en KP son demostrables las propiedades de la relación de orden aritmética

$$x \leq_a y \leftrightarrow \bigvee z \in \omega\ z + x = y.$$

En particular, en KP se demuestra que

$$\bigwedge m \in \omega\ (m \leq_a 0 \leftrightarrow m = 0), \qquad \bigwedge mn \in \omega(m \leq_a n' \leftrightarrow m \leq_a n \vee m = n'),$$

pues ambas sentencias son traducciones de teoremas de IA. Por otro lado tenemos la relación de orden conjuntista definida sobre los ordinales:

$$x \leq_c y \leftrightarrow x \subset y \leftrightarrow x \in y \vee x = y,$$

y en KP (en B, de hecho) se demuestra que cumple también las dos propiedades anteriores. Así, una simple inducción sobre n prueba que

$$\bigwedge mn \in \omega\ (m \leq_a n \leftrightarrow m \leq_c n).$$

(La inducción es sobre la fórmula $m \leq_a n \leftrightarrow m \leq_c n$, claramente de tipo Σ_1 en el sentido conjuntista.)

Ahora estamos en condiciones de probar lo siguiente:

Teorema 6.14 *Si γ es una fórmula Δ_0 de $\mathcal{L}_a$ con variables libres $x_1, \ldots, x_n$, entonces existe una fórmula γ^* de tipo Δ_1 en* KP *(con las mismas variables libres) tal que en* KP *se demuestra*

$$\bigwedge x_1 \cdots x_n \in \omega\ (\gamma_{\mathrm{tc}} \leftrightarrow \gamma^*).$$

Demostración: Por inducción sobre la longitud de γ. Si $\gamma \equiv t_1 = t_2$, entonces $\gamma_{\mathrm{tc}} \equiv t_{1\mathrm{tc}} = t_{2\mathrm{tc}}$ es Δ_1, luego basta tomar $\gamma^* \equiv \gamma_{\mathrm{tc}}$. Si $\gamma \equiv t_1 \leq t_2$, entonces $\gamma_{\mathrm{tc}} \equiv t_{1\mathrm{tc}} \leq_a t_{2\mathrm{tc}}$. Si $x_1, \ldots, x_n \in \omega$, se cumple que $t_{1\mathrm{tc}}, t_{2\mathrm{tc}} \in \omega$, luego

$$\bigwedge x_1 \cdots x_n \in \omega(\gamma_{\mathrm{tc}} \leftrightarrow t_{1\mathrm{tc}} \subset t_{2\mathrm{tc}}),$$

y la fórmula $t_{1\mathrm{tc}} \subset t_{2\mathrm{tc}}$ es Δ_1.

Si el teorema es cierto para γ_1 y γ_2, es inmediato que también vale para $\neg\gamma_1$ y $\gamma_1 \to \gamma_2$. Supongamos finalmente que $\gamma \equiv \bigwedge x \leq y\, \alpha(x)$, de modo que, por hipótesis de inducción existe α^* de tipo Δ_1 tal que en KP se demuestra

$$\bigwedge x x_1 \cdots x_n \in \omega(\alpha_{\mathrm{tc}}(x) \leftrightarrow \alpha^*(x)).$$

Entonces

$$\gamma_{\mathrm{tc}} \leftrightarrow \bigwedge x(x \in \omega \wedge x \leq_a y \to \alpha_{\mathrm{tc}}(x)).$$

Si suponemos $y, x_1, \ldots, x_n \in \omega$, en KP podemos probar las equivalencias siguientes:

$$\gamma_{\mathrm{tc}} \leftrightarrow \bigwedge x(x \in \omega \wedge x \leq_c y \to \alpha^*(x)) \leftrightarrow \bigwedge x \in y\, \alpha^*_{\mathrm{tc}}(x) \wedge \alpha^*(y),$$

luego basta tomar $\gamma^* \equiv \bigwedge x \in y\, \alpha^*(x) \wedge \alpha^*(u)$, pues esta fórmula es claramente de tipo Δ_1 en KP. ∎

A su vez, el teorema anterior vale también para fórmulas γ de tipo Σ_1 o Π_1 en el sentido aritmético, pues en el primer caso $\gamma \equiv \bigvee x\, \alpha$, con α de tipo Δ_0 y basta tomar $\gamma^* \equiv \bigvee x(x \in \omega \wedge \alpha^*)$, que es una fórmula de tipo Σ_1 en la jerarquía de Lévy y claramente cumple lo pedido. Igualmente sucede si γ es Π_1.

Finalmente, si tenemos un caso de Σ_1-inducción en $\mathcal{L}_a$:

$$\gamma \equiv \phi(0) \wedge \bigwedge x\ (\phi(x) \to \phi(x')) \to \bigwedge x\, \phi(x),$$

donde la fórmula ϕ es Σ_1 y tiene variables libres $x, x_1, \ldots, x_n$, su traducción es

$$\gamma_{\mathrm{tc}} \equiv \phi_{\mathrm{tc}}(0) \wedge \bigwedge x \in \omega\ (\phi_{\mathrm{tc}}(x) \to \phi_{\mathrm{tc}}(x')) \to \bigwedge x \in \omega\, \phi_{\mathrm{tc}}(x).$$

Si suponemos $x_1, \ldots, x_n \in \omega$, tenemos que esto equivale a

$$\gamma^* \equiv \phi^*(0) \wedge \bigwedge x \in \omega\ (\phi^*(x) \to \phi^*(x')) \to \bigwedge x \in \omega\, \phi^*(x),$$

y, como ϕ^* es Σ_1, sabemos que γ^* es un teorema de KP, que a su vez implica trivialmente γ^c_{tc}. Con esto ya podemos afirmar que en KP se demuestran las traducciones de todos los axiomas de $\mathrm{I}\Sigma_1$, por lo que el teorema 3.30 nos da que en KP se demuestran las traducciones de todos los teoremas de $\mathrm{I}\Sigma_1$:

Teorema 6.15 *Si γ es una sentencia de $\mathcal{L}_a$ y $\underset{\mathrm{I}\Sigma_1}{\vdash} \gamma$, entonces $\underset{\mathrm{KP}}{\vdash} \gamma_{\mathrm{tc}}$.*

En particular, como consecuencia de este teorema podemos considerar definidos en KP los conceptos de resta, máximo común divisor, número primo, etc., y podemos contar con las propiedades que hemos demostrado en $\mathrm{I}\Sigma_1$ para estos conceptos. Además hemos visto la equivalencia de las relaciones

$$x \leq_a y \equiv x \in \omega \wedge y \in \omega \wedge \bigvee z \in \omega\ y = z + x,$$

$$x \leq_c y \equiv x \in \omega \wedge y \in \omega \wedge x \subset y.$$

De la prueba se desprende también que si γ es una fórmula aritmética de tipo Σ_1 o Π_1 (en sentido amplio) en $\mathrm{I}\Sigma_1$, entonces existe una fórmula γ' de tipo Σ_1 o Π_1 en KP con las mismas variables libres tal que en KP se demuestra

$$\underset{\mathrm{KP}}{\vdash} \bigwedge x_1 \cdots x_n \in \omega\ (\gamma_{\mathrm{tc}} \leftrightarrow \gamma').$$

En efecto, $\underset{\mathrm{I}\Sigma_1}{\vdash} (\gamma \leftrightarrow \gamma^*)$, para cierta fórmula γ^* de tipo Σ_1 o Π_1 (en sentido estricto) con las mismas variables libres, luego por el teorema anterior $\underset{\mathrm{KP}}{\vdash} \bigwedge x_1 \cdots x_n \in \omega\, (\gamma_{\mathrm{tc}} \leftrightarrow \gamma^*_{\mathrm{tc}})$, y a su vez hemos visto que existe una fórmula γ' del mismo tipo, pero para la jerarquía de Lévy, tal que

$$\underset{\mathrm{KP}}{\vdash} \bigwedge x_1 \cdots x_n \in \omega\, (\gamma^*_{\mathrm{tc}} \leftrightarrow \gamma').$$

∎

La interpretación de KP en $\mathrm{I}\Sigma_1$ A cada expresión θ de $\mathcal{L}_{\mathrm{tc}}$ le podemos asociar una expresión θ_a de $\mathcal{L}_a$ sin más que reemplazar cada fórmula $x \in y$ por la fórmula correspondiente definida en $\mathrm{I}\Sigma_1$. Por ejemplo, la traducción de

$$\bigwedge xy(\bigwedge u(u \in x \leftrightarrow u \in y) \to x = y)$$

es

$$\bigwedge xy(\bigwedge u(u \in x \leftrightarrow u \in y) \to x = y),$$

donde en la primera fórmula $\in$ es el relator de $\mathcal{L}_{\mathrm{tc}}$ y en la segunda $u \in x$ es la fórmula definida en $\mathrm{I}\Sigma_1$. En este caso la traducción θ_a de una expresión θ es casi literalmente "la misma expresión", sólo que interpretada de otro modo. Ahora bien, es fácil ver que si γ es una fórmula Δ_0 en $\mathcal{L}_{\mathrm{tc}}$, entonces γ_a es Δ_1 en $\mathrm{I}\Sigma_1$. En efecto, esto es cierto para $x = y$ y $x \in y$, si vale para α y β es claro que vale para $\neg\alpha$ y $\alpha \to \beta$, mientras que en el caso en que $\gamma \equiv \bigwedge x \in y\, \beta$, tenemos que

$$\gamma_a \equiv \bigwedge x \in y\, \beta_a \leftrightarrow \bigwedge x < y(x \in y \to \beta_a),$$

que es Δ_1 en $\mathrm{I}\Sigma_1$ si β_a lo es. Por consiguiente, si γ es Σ_1 o Π_1 en $\mathcal{L}_{\mathrm{tc}}$ se cumple que γ_a es Σ_1 o Π_1 en $\mathrm{I}\Sigma_1$.

Ya hemos visto que en AP se demuestran los axiomas de Z^* (incluso los de ZF$-$AI, que es una teoría más fuerte). Ahora probamos que en $\mathrm{I}\Sigma_1$ se demuestran los de KP:

Teorema 6.16 *Si γ es un axioma de* KP, *entonces γ_a es un teorema de* $\mathrm{I}\Sigma_1$.

Demostración: El axioma de extensionalidad está probado en 5.53, el del par y el de la unión en 5.56, el de especificación en 5.54, teniendo en cuenta que las fórmulas Δ_0 de $\mathcal{L}_{\mathrm{tc}}$ se traducen a fórmulas Δ_1 en $\mathrm{I}\Sigma_1$, según hemos observado antes del teorema.

Respecto al axioma de recolección, hemos probado que en KP (sin dicho axioma) equivale a

$$\bigwedge x \in u \bigvee y\, \phi(x, y) \to \bigvee v \bigwedge x \in u \bigvee y \in v\, \phi(x, y),$$

y en $\mathrm{I}\Sigma_1$ hemos demostrado (teorema 5.22):

$$\bigwedge x \le u \bigvee y\, \phi(x, y) \to \bigvee v \bigwedge x \le u \bigvee y \le v\, \phi(x, y).$$

Si partimos que una fórmula ϕ de tipo Δ_0 en $\mathcal{L}_{\mathrm{tc}}$, entonces ϕ_a es Δ_1 en $\mathrm{I}\Sigma_1$, al igual que lo es

$$\Phi(x, y) \equiv (x \in u \wedge \phi_a(x, y)) \vee (x \notin u \wedge y = 0).$$

Aplicando a esta fórmula el principio de recolección aritmético obtenemos la traducción del de KP.

Sólo falta probar el axioma de regularidad, pero si $\phi(u)$ es de tipo Π_1 en $\mathcal{L}_{\mathrm{tc}}$, entonces $\phi_a(u)$ lo es en $\mathrm{I}\Sigma_1$, y si suponemos $\bigvee u\, \phi_a(u)$ podemos tomar el mínimo u que cumpla esto (por 5.25) y claramente dicho u cumple lo exigido por el axioma de regularidad. ∎

El teorema siguiente es inmediato:

Teorema 6.17 *Si γ es una fórmula de $\mathcal{L}_{\mathrm{tc}}$ y $\underset{\mathrm{KP}}{\vdash} \gamma$, entonces $\underset{\mathrm{I}\Sigma_1}{\vdash} \gamma_a$.*

En efecto, una demostración de γ en KP puede verse sin cambio alguno como una demostración de γ_a en $\mathrm{I}\Sigma_1$, pues lo que son axiomas en KP son teoremas en $\mathrm{I}\Sigma_1$.

Este teorema y las observaciones previas a 6.16 implican a su vez que la traducción de toda expresión de tipo Σ_1 o Π_1 (en sentido amplio) en KP es Σ_1 o Π_1 en $\mathrm{I}\Sigma_1$.

Cómo trabajar simultáneamente en KP y en $\mathrm{I}\Sigma_1$ Vamos a ver que, tomando unas mínimas precauciones, es posible trabajar simultáneamente en $\mathrm{I}\Sigma_1$ y KP.

Para ello escribiremos $x \in \mathbb{N}$ dejando dos posibles interpretaciones a esta fórmula: si trabajamos en $\mathrm{I}\Sigma_1$ habrá que entenderla como $x = x$, mientras que si trabajamos en KP habrá que entenderla como $x \in \omega$. Similarmente, cuando hablemos de la suma y el producto de números naturales, o del 0, se interpretarán como los funtores y la constante de $\mathcal{L}_a$ o como los términos definidos en KP. Por ejemplo, consideremos esta fórmula:

$$\bigwedge x(x \subset \mathbb{N} \wedge x \neq \varnothing \to \bigvee n \in x \bigwedge m \in x\ n \leq m).$$

Podemos considerar indistintamente que se trata de una fórmula de $\mathcal{L}_a$ o una fórmula de $\mathcal{L}_{\mathrm{tc}}$. En el primer caso $x \subset \mathbb{N} \equiv \bigwedge u(u \in x \to u \in \mathbb{N})$ se cumple trivialmente, pues $u \in \mathbb{N} \equiv u = u$. Además $n \in x$ se interpreta como la fórmula definida en $\mathrm{I}\Sigma_1$, mientras que $n \leq m \equiv \bigvee z\ z + n = m$. Si, por el contrario, la interpretamos como fórmula de $\mathcal{L}_{\mathrm{tc}}$ hemos de entender que $x \subset \mathbb{N}$ no es trivial, pues ahora $u \in \mathbb{N} \equiv u \in \omega$, mientras que en $m \in x$ el signo $\in$ es el relator de pertenencia de $\mathcal{L}_{\mathrm{tc}}$ y

$$m \leq n \equiv m \in \mathbb{N} \wedge n \in \mathbb{N} \wedge m \subset n \leftrightarrow m \in \mathbb{N} \wedge n \in \mathbb{N} \wedge \bigvee z \in \mathbb{N}\ z + m = n.$$

La equivalencia la hemos probado en la demostración del teorema 6.15. A partir de ahora siempre interpretaremos así la fórmula $x \leq y$ cuando la consideremos como fórmula de $\mathcal{L}_{\mathrm{tc}}$, y análogamente

$$x < y \equiv x \in \mathbb{N} \wedge y \in \mathbb{N} \wedge x \in y \leftrightarrow x \in \mathbb{N} \wedge y \in \mathbb{N} \wedge \bigvee z \in \mathbb{N}(z \neq 0 \wedge z + x = y).$$

Observemos que $z = x + y$ es Δ_0 en $\mathrm{I}\Sigma_1$, pero es Δ_1 en KP, mientras que con $x \in y$ sucede lo contrario. Ahora bien, toda fórmula construida a partir de $x \in \mathbb{N}$, 0, x', $x + y$, $x \cdot y$, $x \leq y$, $x < y$, $x \in y$, si es Σ_1 o Π_1 en $\mathrm{I}\Sigma_1$, también lo es en KP, y viceversa. Esto es inmediato a partir de dos observaciones: la primera es que todos estos términos y fórmulas son Δ_1 respecto de ambas interpretaciones, y la segunda es que si una fórmula α es Σ_1 o Π_1 respecto de ambas interpretaciones, lo mismo le sucede a las fórmulas

$$\bigwedge x \in y\, \alpha, \quad \bigvee x \in y\, \alpha, \quad \bigwedge x \leq y\, \alpha, \quad \bigvee x \leq y\, \alpha, \quad \bigwedge x < y\, \alpha, \quad \bigvee x < y\, \alpha.$$

En efecto, las dos primeras son trivialmente Σ_1 o Π_1 en KP, y también lo son en $\mathrm{I}\Sigma_1$, porque sabemos que $x \in y \to x \leq y$, luego los cuantificadores pueden acotarse también. Igualmente, las dos siguientes son trivialmente Σ_1 o Π_1 en $\mathrm{I}\Sigma_1$, mientras que en KP tenemos que

$$\bigwedge x \leq y\, \alpha \leftrightarrow \bigwedge x \in y\, \alpha \wedge \alpha(y),$$

que también es Σ_1 o Π_1, y con las dos últimas es aún más fácil.

Así, la fórmula que hemos puesto como ejemplo más arriba es Π_1 tanto en $\mathrm{I}\Sigma_1$ como en KP, pues la fórmula tras el $\bigwedge x$ es Δ_1 interpretada en ambas teorías.

Señalemos también que si α es una fórmula de $\mathcal{L}_a$ y consideramos su traducción α_{tc} a $\mathcal{L}_{\mathrm{tc}}$, al interpretar esta traducción como fórmula de $\mathcal{L}_a$, resulta ser una fórmula lógicamente equivalente a α, pues lo único que hacemos al traducir es cambiar los funtores de $\mathcal{L}_a$ por los términos correspondientes de $\mathcal{L}_{\mathrm{tc}}$ (y al interpretar α_{tc} de nuevo como fórmula de $\mathcal{L}_a$ volvemos a interpretarlos como los funtores originales) y cambiar cada $\bigwedge x$ por $\bigwedge x \in \mathbb{N}$, lo cual, interpretado de nuevo en $\mathcal{L}_a$ (es decir, con $x \in \mathbb{N} \equiv x = x$), es equivalente a $\bigwedge x$.

A partir de aquí trabajaremos en KP, pero sin apoyarnos en ningún momento en las construcciones concretas que hemos dado de los números naturales, la suma y el producto, es decir, trabajaremos en KP más el supuesto de que tenemos una cierta fórmula $x \in \mathbb{N}$, un cierto designador 0 y unos ciertos términos x', $x + y$, $x \cdot y$ (todos de tipo Δ_1), tales que cumplen el teorema 6.13.

Sabemos que existen tales términos y fórmulas, por lo que todo cuanto demostraremos a partir de aquí serán teoremas de KP sin más que interpretar $x \in \mathbb{N}$ como $x \in \omega$, etc. Pero si no usamos las definiciones concretas de estos términos y fórmulas, sucede que todas las fórmulas que demostremos a partir de los axiomas de KP y del teorema 6.13 podrán interpretarse también como teoremas de $\mathrm{I}\Sigma_1$, sin cambio alguno en las demostraciones, pues hemos visto (teorema 6.16) que todos los axiomas de KP son teoremas de $\mathrm{I}\Sigma_1$ cuando se interpretan como fórmulas de $\mathcal{L}_a$.

En realidad, si queremos que todo lo que demostremos en KP valga también en $\mathrm{I}\Sigma_1$ no hay ningún inconveniente en que en un momento dado demostremos un teorema en KP usando que los números naturales son ordinales a condición de que demos otra prueba válida en $\mathrm{I}\Sigma_1$. Por ejemplo, en KP podemos probar que

$$\bigwedge n \in \mathbb{N} \overset{1}{\bigvee} x \bigwedge u(u \in x \leftrightarrow u < n)$$

diciendo simplemente que basta tomar $x = n$. Esto no prueba el resultado en $\mathrm{I}\Sigma_1$, pero ahí ya lo tenemos probado de otro modo. Concluimos que

$$I_n \equiv x | \bigwedge u(u \in x \leftrightarrow u < n)$$

es una descripción propia en KP siempre que $n \in \mathbb{N}$ y esto también es válido (por otro argumento) en $\mathrm{I}\Sigma_1$, pero debemos tener en cuenta que $\bigwedge n \in \mathbb{N}\ I_n = n$ es un teorema de KP que es falso en $\mathrm{I}\Sigma_1$. Mientras no usemos resultados como éste (o si los usamos, demos un par de pruebas alternativas para cada teoría), los resultados que probemos valdrán en las dos teorías.

Observemos por último que al interpretar un teorema de KP como teorema de $\mathrm{I}\Sigma_1$, podemos, si queremos, reinterpretar los pares ordenados conjuntistas $(u,v) \equiv \{\{u\},\{u,v\}\}$ como pares aritméticos $\langle u,v\rangle$. En efecto, lo único que usaremos de los pares ordenados es que cumplen

$$(u,v) = (u',v') \leftrightarrow u = u' \wedge v = v',$$

no su definición concreta.[7]

La prueba que hemos dado en KP de la existencia del producto cartesiano $x \times y$ de dos conjuntos no vale en $\mathrm{I}\Sigma_1$ si interpretamos los pares ordenados en el sentido aritmético, pero en este caso tenemos una prueba mucho más simple, pues basta tener en cuenta que si $u \in x \wedge v \in y$, entonces $\langle u,v\rangle < \langle x,y\rangle$, luego el teorema anterior se cumple tomando

$$z = \{w \in I_{(x,y)} \mid \bigvee u \in x \bigvee v \in y\ w = (u,v)\}.$$

Así pues, tanto en $\mathrm{I}\Sigma_1$ como en KP tenemos que

$$x \times y \equiv z | \bigwedge w(w \in z \leftrightarrow \bigvee u \in x \bigvee v \in y\ w = (x,y))$$

es una descripción propia (o, equivalentemente, que el producto cartesiano de dos conjuntos es un conjunto). El término $x \times y$ es Δ_0 en KP y Δ_1 en $\mathrm{I}\Sigma_1$.

6.4 Conjuntos finitos, cardinales

Los resultados de esta sección se demuestran en KP y en $\mathrm{I}\Sigma_1$, según acabamos de explicar. Es fácil definir el concepto de conjunto finito, así como el cardinal (el número de elementos) de un conjunto finito:

Definición 6.18 x es finito $\equiv \bigvee f \bigvee n \in \mathbb{N}\ f : I_n \longrightarrow x$ biyectiva.

El teorema siguiente implica que el número natural n está unívocamente determinado:

Teorema 6.19 $\bigwedge mn \in \mathbb{N} \bigwedge f(f : I_m \longrightarrow I_n$ *inyectiva* $\to m \leq n)$.

Demostración: Razonamos por inducción[8] sobre m. Si $m = 0$ es trivial. Supuesto cierto para m, supongamos que $f : I_{m+1} \longrightarrow I_n$ inyectiva. Podemos suponer que $f(m) = n-1$. En efecto, si $n-1 \notin \mathcal{R}f$, basta tomar

$$f^* = (f \setminus \{(m,f(m))\}) \cup \{(m,n-1)\},$$

y es claro que $f^* : I_{m+1} \longrightarrow I_n$ inyectiva y $f^*(m) = n-1$. Si $n-1 \in \mathcal{R}f$, pero no es $f(m)$, sea $u < m$ tal que $f(u) = n-1$. Tomamos

$$f^* = (f \setminus \{(m,f(m)),(u,n-1)\}) \cup \{(u,f(m)),(m,n-1)\},$$

y es claro que esta f^* cumple lo mismo que en el caso anterior.

[7] En realidad sí que usaremos su definición concreta al acotar cuantificadores de la forma $\bigvee uv\ x = (u,v)$ como $\bigvee w \in x \bigvee uv \in w\ x = (u,v)$, para comprobar que determinadas expresiones son Σ_1 o Π_1, pero este tipo de acotaciones se realizan de forma más simple en $\mathrm{I}\Sigma_1$, donde podemos escribir $\bigvee uv < x\ x = (u,v)$.

[8] La fórmula es claramente Π_1. Recordemos que disponemos tanto de la inducción Σ_1 como de la inducción Π_1.

Pero entonces, $f \setminus \{(m, n-1)\} : I_m \longrightarrow I_{n-1}$ inyectiva, luego por hipótesis de inducción $m \leq n-1$, luego $m+1 \leq n$. ∎

En particular,

$$\bigwedge nmf(f : I_m \longrightarrow I_n \text{ biyectiva} \to m = n),$$

luego podemos definir el *cardinal* de un conjunto finito:

Definición 6.20 $|x| \equiv n | \bigvee f\ f : I_n \longrightarrow x$ *biyectiva.*

Teorema 6.21 *Si x e y son conjuntos finitos, se cumple:*

1. $\bigwedge xy(|x| = |y| \leftrightarrow \bigvee f\ f : x \longrightarrow y$ *biyectiva*$)$,
2. $\bigwedge xy(|x| \leq |y| \leftrightarrow \bigvee f\ f : x \longrightarrow y$ *inyectiva*$)$,

Demostración: 1) Si $|x| = |y| = n$, existen $g : I_n \longrightarrow x$, $h : I_n \longrightarrow y$ biyectivas, y basta tomar $f = g^{-1} \circ h$. Si $f : x \longrightarrow y$ biyectiva y $|x| = n$, entonces existe $g : I_n \longrightarrow x$ biyectiva, luego $g \circ f : I_n \longrightarrow y$ biyectiva, luego $|y| = n = |x|$.

2) Sean $|x| = m$, $|y| = n$, $g : I_m \longrightarrow x$ biyectiva y $h : I_n \longrightarrow y$ biyectiva. Si $m \leq n$ entonces $I_m \subset I_n$, luego $g^{-1} \circ h : x \longrightarrow y$ inyectiva. Recíprocamente, si $f : x \longrightarrow y$ inyectiva, $g \circ f \circ h^{-1} : I_m \longrightarrow I_n$ es inyectiva, luego $m \leq n$ por el teorema anterior. ∎

En $\mathrm{I}\Sigma_1$ (pero no en KP) puede probarse que todo conjunto es finito:

Teorema 6.22 ($\mathbf{I\Sigma_1}$) $\bigwedge x \bigvee n f(f : I_n \longrightarrow x$ *biyectiva*$)$.

Demostración: Lo probamos por inducción[9] sobre x. Lo suponemos cierto para todo $y < x$. Si $x = 0$ basta tomar $n = 0$, pues entonces $x = I_0 = \varnothing$. Si $x \neq 0$, entonces existe un $u \in x$. Sea $y = x \setminus \{u\}$. Entonces $y \subset x$, luego $y \leq x$, y de hecho $y < x$, pues no tienen los mismos elementos. Por hipótesis de inducción existe $f : I_n \longrightarrow y$ biyectiva, y es claro que $f \cup \{(n, u)\} : I_{n+1} \longrightarrow x$ es biyectiva. ∎

Teorema 6.23 *Si x, y son conjuntos finitos y $x \cap y = \varnothing$, entonces $x \cup y$ es finito y $|x \cup y| = |x| + |y|$).*

Demostración: Veamos por inducción[9] sobre n que

$$\bigvee f\ f : (I_m \times \{0\}) \cup (I_n \times \{1\}) \longrightarrow I_{m+n} \text{ biyectiva.}$$

Si $n = 0$ es fácil construir una biyección $I_m \times \{0\} \longrightarrow I_m$ biyectiva. Supuesto cierto para n, observamos que

$$(I_m \times \{0\}) \cup (I_{n+1} \times \{1\}) = (I_m \times \{0\}) \cup (I_n \times \{1\}) \cup \{(n, 1)\}.$$

[9] La fórmula es obviamente Σ_1.

Si f es la aplicación dada por hipótesis de inducción, es claro que

$$f^* = f \cup \{((n,1), m+n)\}$$

es una biyección que prueba el resultado para $n+1$.

Ahora, si $|x| = m$ y $|y| = n$, es fácil construir una aplicación biyectiva

$$x \cup y \longrightarrow (I_m \times \{0\}) \cup (I_n \times \{1\}),$$

con lo que llegamos a una biyección $x \cup y \longrightarrow I_{m+n}$. Esto prueba el teorema. ∎

Teorema 6.24 *Si x e y son conjuntos finitos, entonces $\bigwedge xy\ |x \times y| = |x| \cdot |y|$.*

Demostración: La prueba es análoga a la del teorema anterior, probando ahora por inducción sobre n que

$$\bigvee f\ f : I_m \times I_n \longrightarrow I_{mn} \text{ biyectiva.}$$

Notemos que

$$I_m \times I_{n+1} = (I_m \times I_n) \cup (I_m \times \{n\}),$$

y la unión es disjunta, por lo que podemos aplicarle el teorema anterior: si $I_m \times I_n$ es finito por hipótesis de inducción, entonces la unión también lo es, y su cardinal es $mn + m = m(n+1)$. ∎

Teorema 6.25 *Si x es un conjunto finito y $u \subset x$, entonces u es finito y cumple $|u| \leq |x|$. Además, se cumple $|u| = |x|$ si y sólo si $u = x$.*

Demostración: Sea $|x| = n$. Entonces u se puede biyectar con un $v \subset I_n$, y basta ver que v es finito con $|v| \leq n$. Veamos por inducción sobre m que

$$\bigvee f \bigvee r \leq m\ f : I_r \longrightarrow v \cap I_m \text{ biyectiva.}$$

Si $m = 0$ es trivial. Si vale para m, o bien $m \notin v$, en cuyo caso tenemos que $v \cap I_m = v \cap I_{m+1}$ y la conclusión es obvia, o bien $m \in v$, en cuyo caso tenemos que $v \cap I_{m+1} = (v \cap I_m) \cup \{m\}$ y la unión es disjunta. Por hipótesis de inducción el primer conjunto es finito y $|v \cap I_m| \leq m$, luego

$$|v \cap I_{m+1}| = |v \cap I_m| + 1 \leq m + 1.$$

En particular, aplicando esto a $m = n$ tenemos que $|v| \leq m$.

Si $|u| = |x|$ entonces $x = u \cup (x \setminus u)$, luego $|x| = |u| + |x \setminus u|$, luego $|x \setminus u| = 0$, luego $x \setminus u = \varnothing$, luego $u = x$. ∎

Teorema 6.26 *Si x e y son conjuntos finitos, entonces $x \cup y$ es finito y además $|x \cup y| \leq |x| + |y|$.*

Demostración: $x \cup y = x \cup (y \setminus x)$ es una unión de conjuntos finitos disjuntos, luego es finita. Además $|x \cup y| = |x| + |y \setminus x| \leq |x| + |y|$. ∎

Más en general:

Teorema 6.27 *Si a es un conjunto finito y $\bigwedge x \in a\ x$ es finito, entonces $\bigcup a$ es finito.*

Demostración: Sea $f : I_n \longrightarrow a$ biyectiva. Basta probar por inducción[10] sobre m que $m \leq n \to \bigcup f[I_m]$ es finito. Si $m = 0$, entonces $\bigcup f[I_0] = \varnothing$ es finito. Si vale para m y $m + 1 \leq n$, entonces $f[I_{m+1}] = f[I_m] \cup \{f(m)\}$, luego $\bigcup f[I_{m+1}] = \bigcup f[I_m] \cup f(m)$. El primer conjunto es finito por hipótesis de inducción y el segundo por hipótesis, luego la unión es finita. ∎

Veamos ahora algunas propiedades de los conjuntos finitos:

Teorema 6.28 *Sea $x \neq \varnothing$ un conjunto finito.*

1. *x admite un buen orden.*
2. *Toda relación de orden en x tiene un elemento maximal y un minimal.*
3. *Toda relación de orden total en x es un buen orden.*

Demostración: 1) Sea $f : I_n \longrightarrow x$ biyectiva. Definimos

$$R = \{(u, v) \in x \times x \mid f^{-1}(u) \leq f^{-1}(v)\}.$$

Claramente es un buen orden en x.

2) Si R es una relación de orden en x, definimos

$$R' = \{(u, v) \in I_n \times I_n \mid f(u)\, R\, f(v)\}$$

y así R' es una relación de orden en x, y basta ver que I_n tiene maximal respecto a R' (para concluir que tiene minimal aplicamos el resultado a R'^{-1}). Para ello probamos por inducción sobre m que si $1 \leq m \leq n$ entonces I_m tiene R'-maximal.[11] Obviamente 0 es R' maximal para I_1. Si I_m tiene R'-maximal, digamos u, entonces, o bien $\neg u\, R'\, m$, en cuyo caso u es R'-maximal de I_{m+1}, o bien $u\, R'\, m$, en cuyo caso m es un R'-maximal de I_{m+1}.

3) Si $y \subset x$ es un subconjunto no vacío, entonces es finito, luego tiene R-minimal, pero un R-minimal para una relación de orden total es un mínimo. ∎

Nota En particular, como en $\mathrm{I}\Sigma_1$ se demuestra que todo conjunto es finito, tenemos que todo conjunto admite un buen orden, y veremos que uno de los enunciados equivalentes del axioma de elección es precisamente que todo conjunto admite un buen orden, por lo que tenemos probado que $\mathrm{I}\Sigma_1$ satisface todos los axiomas de ZFC−AI. ∎

El teorema siguiente es válido para conjuntos arbitrarios, aunque no sean finitos, pero lo incluimos aquí como resultado previo al teorema posterior, que sí que es exclusivo de conjuntos finitos:

[10] La fórmula es Σ_1.

[11] Explícitamente $\bigvee u < m \bigwedge v < m\, (v \neq u \to (u, v) \notin R')$, que es Δ_1.

Teorema 6.29 *Sean x e y conjuntos no vacíos.*

1. *si $f : x \longrightarrow y$ y $g : y \longrightarrow x$ son dos aplicaciones tales que $f \circ g = i_x$ (donde i_x representa la identidad en x), se cumple que f es inyectiva y g es suprayectiva.*
2. *Si $f : x \longrightarrow y$ es inyectiva, existe $g : y \longrightarrow x$ suprayectiva tal que $f \circ g = i_x$.*
3. *Si $g : y \longrightarrow x$ es suprayectiva e y admite un buen orden (en particular si y es finito), entonces existe $f : x \longrightarrow y$ inyectiva tal que $f \circ g = i_x$.*

DEMOSTRACIÓN: 1) es un hecho general que se demuestra trivialmente en la teoría básica B.

2) Tomemos $a \in x$. Basta definir $g = f^{-1} \cup (y \setminus f[x]) \times \{a\})$. En otras palabras,

$$g(u) = \begin{cases} f^{-1}(u) & \text{si } u \in f[x], \\ a & \text{si } u \notin f[x]. \end{cases}$$

3) Sea $R \subset y \times y$ un buen orden en y. Sea

$$f = \{(u, v) \in x \times y \mid (v, u) \in g \wedge \bigwedge v' \in y(v' \, R \, v \wedge v' \neq v \to (v', u) \notin g)\}.$$

En otras palabras, $f(u)$ es el mínimo de $g^{-1}[u]$ respecto de R. Es claro que f cumple lo pedido. ∎

Teorema 6.30 *Si $f : x \longrightarrow x$ y x es finito, entonces f es inyectiva si y sólo si es suprayectiva.*

DEMOSTRACIÓN: Podemos suponer que $x \neq \varnothing$. Si f es inyectiva, entonces $f[x] \subset x$ cumple $|f[x]| = |x|$, luego $f[x] = x$ y f es suprayectiva.

Si f es suprayectiva, sea $g : x \longrightarrow x$ inyectiva tal que $g \circ f = I_x$. Entonces g es biyectiva, luego $f = g^{-1}$ también lo es. ∎

Las sucesiones de una longitud fija en un rango fijo forman un conjunto:

Teorema 6.31 $\bigwedge n \in \mathbb{N} \bigwedge x \overset{1}{\bigvee} y \bigwedge s(s \in y \leftrightarrow s : I_n \longrightarrow x)$.

DEMOSTRACIÓN: Vamos a usar el teorema 5.57 en $\mathrm{I}\Sigma_1$ o bien 6.12 en KP. Consideramos la fórmula Δ_1

$$\psi(x, n, a, y) \equiv \bigwedge f \in y \, (f : I_{n+1} \longrightarrow x \wedge f|_{I_n} \in a) \wedge$$

$$\bigwedge s \in a \bigwedge u \in x(s : I_n \longrightarrow x \to \bigvee f \in y \, (f|_{I_n} = s \wedge f(n) = u)).$$

Vamos a probar que $\bigwedge x a \bigwedge n \in \mathbb{N} \overset{1}{\bigvee} y \, \psi(x, n, a, y)$. Dado un conjunto a, consideramos el subconjunto

$$b = \{s \in a \mid s : I_n \longrightarrow x\},$$

que existe por Δ_1-especificación. Consideramos también la fórmula Δ_1

$$\phi(f,s,u) \equiv f : I_{n+1} \longrightarrow x \wedge f|_{I_n} = s \wedge f(n) = u$$

y observamos que $\bigwedge s \in b \bigwedge u \in x \overset{1}{\bigvee} f\ \phi(f,s,u)$, luego por Δ_1-reemplazo existe una aplicación $h : b \times x \longrightarrow y$ suprayectiva tal que

$$\bigwedge s \in b \bigwedge u \in x\ \phi(s,x,h(s,x)).$$

Claramente y es el conjunto buscado. La unicidad se sigue sin dificultad del axioma de extensionalidad, pues y contiene exactamente a las sucesiones en x que extienden a sucesiones de a.

Como la fórmula ψ es claramente Δ_1, aplicando el teorema 5.57 o 6.12 a $G(x,a,n) \equiv y|\psi(x,n,a,y)$ obtenemos una fórmula $\chi(x,n,y)$ de tipo Δ_1 de modo que $\bigwedge x \bigwedge n \in \mathbb{N} \overset{1}{\bigvee} y\, \chi(x,n,y)$ y, si llamamos $x^n \equiv y|\chi(x,n,y)$, se cumple

$$x^0 = \{\varnothing\} \wedge \bigwedge n \in \mathbb{N}\ x^{n+1} = G(x,n,x^n).$$

Ahora bien, si $f : I_n \longrightarrow x$ es cualquier aplicación, podemos probar por Σ_1-inducción que

$$\bigwedge i \in \mathbb{N}(i \leq n \to \bigvee y(y = x^i \wedge \bigvee g \in y\ g = f|_{I_i})).$$

En particular, $f \in x^n$, de modo que $y = x^n$ cumple lo pedido. ∎

Definición 6.32 $x^n \equiv y|\bigwedge s(s \in y \leftrightarrow s : I_n \longrightarrow x)$.

Se cumple que el término x^n es de tipo Δ_1, pues la fórmula $y = x^n$ es equivalente a la definida en la demostración del teorema anterior.

En particular, por Σ_1-reemplazo existe el conjunto $\{x^m \mid m < n\}$, y también $\bigcup\{x^m \mid m < n\}$. Esto implica que

$$x^{<n} \equiv y|\bigwedge s(s \in y \leftrightarrow \bigvee m < n\ s : I_m \longrightarrow x)$$

es una descripción propia para todo $n \in \mathbb{N}$ y es un término Σ_1 (luego Δ_1), pues

$$y = x^{<n} \leftrightarrow \bigwedge s \in y \bigvee m < n\ s \in x^m \wedge \bigwedge m < n \bigvee z(z = x^m \wedge z \subset y).$$

Ahora es fácil generalizar:

Teorema 6.33 $\bigwedge xy(y \textit{ es finito} \to \overset{1}{\bigvee} z \bigwedge s(s \in z \leftrightarrow s : y \longrightarrow x))$.

Demostración: Como y es finito existe $n \in \mathbb{N}$ y $f : y \longrightarrow I_n$ biyectiva. Basta aplicar Δ_1-reemplazo a la fórmula $\phi(t,s) \equiv s = f \circ t$ y al conjunto $a = x^n$, con lo que obtenemos un conjunto z tal que

$$\bigwedge s(s \in z \leftrightarrow \bigvee t \in x^n\ s = f \circ t).$$

Es fácil ver que z cumple lo pedido, y es único por extensionalidad. ∎

Definición 6.34 $x^y \equiv z \mid \bigwedge s(s \in z \leftrightarrow s : y \longrightarrow x)$.

Acabamos de probar que x^y es una descripción propia siempre que y es finito.

Teorema 6.35 *Si x e y son finitos, entonces x^y es finito y $|x^y| = |x|^{|y|}$.*

DEMOSTRACIÓN: Sean $|x| = m$, $|y| = n$. Entonces es fácil definir una biyección entre x^y y $(I_m)^n$ (de hecho, la hemos definido en la prueba del teorema anterior). Por lo tanto, basta probar que $(I_m)^n$ es finito y que $|(I_m)^n| = m^n$. Probamos por inducción sobre n que $\bigvee f\ f : I_{m^n} \longrightarrow (I_m)^n$ biyectiva. (La fórmula es Σ_1.)

Para $n = 0$ es claro. Si se cumple para n, usamos que podemos construir una biyección de $(I_m)^{n+1}$ en $(I_m)^n \times I_m$ mediante $s \mapsto (s|_{I_n}, s(n))$, con lo que si $(I_m)^n$ es finito, también lo es $(I_m)^{n+1}$ y además

$$|(I_m)^{n+1}| = |(I_m)^n| \cdot m = m^n \cdot m = m^{n+1}.$$

■

Teorema 6.36 $\bigwedge x(x \text{ es finito} \to \overset{1}{\bigvee} y \bigwedge u(u \in y \leftrightarrow u \subset x))$.

DEMOSTRACIÓN: Basta aplicar Σ_1-reemplazo al conjunto $(I_2)^x$ y a la fórmula $\phi(s, u) \equiv u = s^{-1}[1]$. Así obtenemos un conjunto y tal que

$$\bigwedge u(u \in y \leftrightarrow \bigvee s(s : x \longrightarrow \{0, 1\} \wedge u = s^{-1}[1])).$$

es fácil ver que y cumple lo pedido, y la unicidad se tiene por el axioma de extensionalidad. ■

Definición 6.37 $\mathcal{P}x \equiv y | \bigwedge u(u \in y \leftrightarrow u \subset x)$.

Acabamos de probar que $\mathcal{P}x$ es una descripción propia siempre que x es finito. Notemos que el teorema 5.56 prueba que esto vale en $\mathrm{I}\Sigma_1$ para todo conjunto x, lo cual es coherente, pues en $\mathrm{I}\Sigma_1$ todo conjunto es finito.

Precisando ligeramente la prueba del teorema anterior vemos que la aplicación $(I_2)^x \longrightarrow \mathcal{P}x$ allí construida es biyectiva. Por lo tanto:

Teorema 6.38 *Si x es un conjunto finito, entonces $\mathcal{P}x$ también es finito, y $|\mathcal{P}x| = 2^{|x|}$.*

6.5 Sumas finitas

Cuando los matemáticos tienen definida una operación asociativa (que podemos representar por $+$, aunque en principio es una operación cualquiera) en un conjunto A, automáticamente se permiten escribir expresiones del tipo $x_1 + \cdots + x_n$, para elementos $x_i \in A$. Esto es correcto, pero, desde un punto de vista lógico, requiere una justificación que vamos a presentar aquí.

El contenido de esta sección conviene desarrollarlo en un contexto muy general, así que supondremos dada una fórmula $x \in A$, un término $x + y$ y un designador 0, todos de tipo Δ_1, de modo que se cumpla:

1. $\bigwedge xy \in A\ x + y \in A$
2. $\bigwedge xy \in A\ x + y = y + x$
3. $\bigwedge xyz \in A\ x + (y + z) = (x + y) + z$
4. $\bigwedge x \in A\ x + 0 = x$.

Obviamente todo esto se cumple cuando $A = \mathbb{N}$ y $+$ es la suma usual, pero no nos restringimos a este caso. Sea $n \in \mathbb{N}$ y consideremos una sucesión[12] $s : I_n \longrightarrow A$. Podemos aplicar el teorema 6.12 a la función $G : \mathbb{N} \times A \longrightarrow A$ definida por la fórmula Σ_1

$$G(i, a) = b \leftrightarrow \psi(s, i, a, b) \equiv (i < \ell(s) \wedge b = a + s_i) \vee (i \geq \ell(s) \wedge b = 0).$$

El resultado es una fórmula $\chi(s, n, x)$ de tipo Σ_1 que define una función $F : \mathrm{Suc} \times \mathbb{N} \longrightarrow A$ tal que

$$F(s, 0) = 0 \wedge \bigwedge m < \ell(s)\ F(s, m + 1) = F(s, m) + s_m.$$

En vez de $F(s, n)$ escribiremos $\sum\limits_{i<n} s_i$, que es un término Δ_1. En estos términos la fórmula anterior se escribe así:

$$\sum_{i<0} s_i = 0 \wedge \bigwedge m < \ell(s) \sum_{i<m+1} s_i = \sum_{i<m} s_i + s_m.$$

Una simple inducción demuestra que

$$\bigwedge mn \in \omega \bigwedge s(s : I_n \longrightarrow A \wedge m \leq \ell(s) \to \sum_{i<m} s_i \in A).$$

Teorema 6.39 *Sean $m, n \in \mathbb{N}$, sea $s : I_{m+n} \longrightarrow A$ y sea $t : I_n \longrightarrow A$ tal que $\bigwedge i < n\ t_i = s_{m+i}$. Entonces*

$$\sum_{i<m+n} s_i = \sum_{i<m} s_i + \sum_{i<n} t_i.$$

Demostración: Probaremos que $k \leq n \to \sum\limits_{i<m+k} s_i = \sum\limits_{i<m} s_i + \sum\limits_{i<k} t_i$. Por inducción[13] sobre k.

Para $k = 0$ tenemos que $\sum\limits_{i<m} s_i = \sum\limits_{i<m} s_i + \sum\limits_{i<0} t_i$ porque el último término es 0.

Si vale para k y $k + 1 \leq n$, entonces

$$\sum_{i<m+k+1} s_i = \sum_{i<m+k} s_i + s_{m+k} = \sum_{i<m} s_i + \sum_{i<k} t_i + t_k = \sum_{i<m} s_i + \sum_{i<k+1} t_i.$$

■

[12]Si trabajamos en $\mathrm{I}\Sigma_1$ o si trabajamos en KP, pero se cumple que $A \subset \mathbb{N}$, alternativamente, podemos considerar que s es un número natural, que podemos ver como sucesión finita según lo visto en la sección 5.6. Todo lo que vamos a exponer en esta sección vale en ambos casos, entendiendo que $\ell(s)$ es el n que cumple que $s : I_n \longrightarrow A$ o bien la longitud definida en 5.58, y que s_i es el conjunto que cumple $(i, s_i) \in s$ o bien el definido también en 5.58.

[13]Se trata de una fórmula $\phi(s, t, n, k)$, con ϕ claramente Σ_1.

Teorema 6.40 (Propiedad conmutativa generalizada) *Sea $s : I_n \longrightarrow A$ una sucesión y sea $\sigma : I_n \longrightarrow I_n$ biyectiva. Sea $t = \sigma \circ s$. Entonces*

$$\sum_{i<n} s_i = \sum_{i<n} t_i.$$

DEMOSTRACIÓN: Demostramos por inducción sobre[14] n que

$$\bigwedge st\sigma(s, t : I_n \longrightarrow A \wedge \sigma : I_n \longrightarrow I_n \text{ biyectiva} \wedge t = \sigma \circ s \to \sum_{i<n} s_i = \sum_{i<n} t_i).$$

Si $n = 0$ es trivial. Supongámoslo cierto para n y supongamos que tenemos $s, t : I_{n+1} \longrightarrow A$ y $\sigma : I_{n+1} \longrightarrow I_{n+1}$ con $t = \sigma \circ s$. Sea $k = \sigma(n)$, de modo que $t_n = s_k$. Si $k = n$ entonces $t|_n = \sigma|_n \circ s|_n$ y $t_n = s_n$, luego

$$\sum_{i<n+1} s_i = \sum_{i<n} (s|_n)_i + s_n = \sum_{i<n} (t|_n)_i + t_n = \sum_{i<n+1} t_i.$$

Supongamos ahora que $k < n$.

$$\sum_{i<n+1} s_i = \sum_{i<k+1} s_i + \sum_{i<n-k} u_i = \sum_{i<k} s_i + s_k + \sum_{i<n-k} u_i,$$

donde $u : I_{n-k} \longrightarrow A$ cumple $u_i = s_{k+1+i}$. Definimos ahora $v : I_n \longrightarrow A$ mediante

$$v = \{(i, a) \in I_n \times \mathcal{R}s \mid (i < k \wedge a = s_i \vee k \leq i \wedge a = s_{i+1})\},$$

es decir,

$$v_i = \begin{cases} s_i & \text{si } i < k \\ s_{i+1}, & \text{si } k \leq i. \end{cases}$$

Claramente $\bigwedge i < n - k \; u_i = v_{k+i}$, luego, por el teorema anterior

$$\sum_{i<n+1} s_i = \sum_{i<k} v_i + \sum_{i<n-k} u_i + s_k = \sum_{i<n} v_i + t_n = \sum_{i<n} t_i + t_n = \sum_{i<n+1} t_i,$$

donde hemos aplicado la hipótesis de inducción a $v, t|_n : I_n \longrightarrow A$ y a la biyección $\tau : I_n \longrightarrow I_n$ dada por

$$\tau = \{(i, j) \in I_n \times I_n \mid (\sigma(i) < k \wedge j = \sigma(i)) \vee (k < \sigma(i) \wedge j = \sigma(i) - 1)\}.$$

Es pura rutina comprobar que τ es en efecto una biyección y que $t|_n = \tau \circ v$. ∎

Ahora podemos generalizar la definición de suma finita:

Definición 6.41 $\sum_{i \in x} s_i \equiv a|(s : x \longrightarrow A \wedge \bigvee n \in \mathbb{N} \bigvee tu(t : I_n \longrightarrow x \text{ biyectiva}$
$$\wedge\, u = t \circ s \wedge a = \sum_{i<n} u_i)).$$

[14] La fórmula es Π_1.

Así, si x es un conjunto finito y $s : x \longrightarrow A$, la definición anterior es una descripción propia, pues claramente existe un a que cumple la definición y si a y a^* la cumplen, tenemos $t, t^* : I_n \longrightarrow x$ biyectivas (con el mismo n, pues necesariamente $n = |x|$) y $u = t \circ s$, $u^* = t^* \circ s$, pero entonces tenemos una biyección $\sigma = t^* \circ t^{-1} : I_n \longrightarrow I_n$ tal que $u^* = t^* \circ s = \sigma \circ t \circ s = \sigma \circ u$, luego el teorema anterior nos da que

$$a^* = \sum_{i<n} u_i^* = \sum_{i<n} u_i = a.$$

De aquí se sigue inmediatamente que $\sum_{i\in\varnothing} s_i = 0$ y que $\sum_{i\in\{j\}} s_i = s_j$, así como que si x e y son conjuntos finitos disjuntos, entonces[15]

$$\sum_{i\in x\cup y} s_i = \sum_{i\in x} s_i + \sum_{i\in y} s_i.$$

Para probar esto se usa 6.39 a través de una biyección $I_{m+n} \longrightarrow x \cup y$ definida de modo que se restrinja a una biyección $I_m \longrightarrow x$.

Teorema 6.42 (Propiedad asociativa generalizada) *Sea x un conjunto finito tal que $\bigwedge y \in x\ y$ es finito $\wedge \bigwedge yz \in x(y \neq z \to y \cap z = \varnothing)$, sea $s : \bigcup x \longrightarrow A$. Entonces*

$$\sum_{i\in\cup x} s_i = \sum_{y\in x}\sum_{i\in y} s_i.$$

Demostración: Aquí hay que precisar cómo se define el doble sumatorio. Consideremos la fórmula

$$\phi(y, a) \equiv \bigvee n \in \mathbb{N} \bigvee tu(t : I_n \longrightarrow y \wedge u = t \circ s \wedge a = \sum_{i<n} u_i).$$

Claramente es Σ_1 y $\bigwedge y \in x \overset{1}{\bigvee} a\ \phi(y, a)$, luego por Σ_1-reemplazo (teorema 6.11) existe $t : x \longrightarrow A$ tal que $\bigwedge y \in x\ t(y) = \sum_{i\in y} s_i$. El sumatorio del miembro derecho es, por definición, $\sum_{y\in x} t_y$.

Para probar la igualdad fijamos $f : I_n \longrightarrow x$ biyectiva y probamos por inducción[16] sobre m que

$$m \leq n \to \sum_{i\in\cup f[I_m]} s_i = \sum_{y\in f[I_m]} t_y.$$

Para $m = 0$ ambos miembros son 0. Si vale para m y $m + 1 \leq n$, entonces $f[I_{m+1}] = f[I_m] \cup \{f(m)\}$ y la unión es disjunta, y a su vez

$$\bigcup f[I_{m+1}] = \bigcup f[I_m] \cup f(m),$$

[15] Notemos que escribimos $\sum_{i\in x} s_i \equiv \sum_{i\in x} (s|_x)_i$.

[16] Es fácil ver que la fórmula es Σ_1.

y la unión también es disjunta, por la hipótesis sobre los elementos de x. Por lo tanto, aplicando la descomposición de las sumas considerada justo antes del enunciado de este teorema, tenemos que

$$\sum_{i\in\cup f[I_{m+1}]} s_i = \sum_{i\in\cup f[I_m]} s_i + \sum_{i\in f(m)} s_i = \sum_{y\in f[I_m]} t_y \ + \ t_{f(m)}$$

$$= \sum_{y\in f[I_m]\cup\{f(m)\}} t_y \ = \sum_{y\in f[I_{m+1}]} t_y.$$

Aplicando esto a $m = n$ tenemos la igualdad del enunciado. ∎

A partir de aquí relajaremos la notación para los sumatorios finitos, y escribiremos cosas como

$$\sum_{i=m}^{n} s_i, \qquad \sum_{0\le i<j\le n} s_{ij}, \qquad s_0 + \cdots + s_n, \qquad \text{etc.}$$

Es fácil reducir expresiones como éstas a expresiones de la forma $\sum_{i\in x} s_i$ para x y s adecuados. Además, toda manipulación con sumas finitas puede justificarse sin dificultad a partir de los resultados que hemos demostrado aquí, aunque en algunos casos la prueba pueda ser tediosa.

Cuando la operación de partida la representemos como $\cdot$ en lugar de $+$ escribiremos $\prod$ en lugar de $\sum$.

6.6 La formalización de la aritmética

Llegados a este punto ya es fácil convencerse de que todos los resultados de la aritmética elemental, que sin duda el lector conocerá, son formalizables tanto en KP como en $\mathrm{I}\Sigma_1$ a través de términos y fórmulas de tipo Δ_1. En esta última sección mostraremos algunos ejemplos de cómo hacerlo, los primeros más detallados, los siguientes en forma de meros esbozos, porque si el lector tiene dudas conceptuales sobre ellos haría mejor en estudiar primero algún libro de álgebra en el que se expongan de forma más natural y detallada para después convencerse de que las técnicas que mostramos aquí bastan para formalizar debidamente todos los argumentos.

Más sobre números primos Notemos que el teorema 6.12 aplicado a la función $G_m(i,k) = k\cdot m$ nos da la función exponencial, caracterizada por que

$$\bigwedge m\in\mathbb{N}\ m^0 = 1 \wedge \bigwedge mn\in\mathbb{N}\ m^{n+1} = m^n\cdot m.$$

El término m^n es Σ_1, luego Δ_1. Similarmente, tomando $G(i,k) = k(i+1)$ obtenemos la función factorial:

$$0! = 1 \wedge \bigwedge n\in\mathbb{N}\ (n+1)! = n!(n+1),$$

de modo que la fórmula $y = n!$ también es Δ_1.

Una simple inducción prueba que

$$m^n = \prod_{i=1}^{n} m, \qquad n! = \prod_{i=1}^{n} i.$$

La segunda parte del teorema 5.36 admite la generalización siguiente:

Teorema 6.43 *Si $m : I_n \longrightarrow \mathbb{N}$, p es primo y $p \mid \prod_{i<n} m_i$, entonces existe un $i < n$ tal que $p \mid m_i$.*

Demostración: Supongamos que $\bigwedge i < n\ p \nmid m_i$ y veamos por inducción sobre k que $k \leq n \to p \nmid \prod_{i<k} m_i$.

Para $k = 0$ es trivial y si $p \mid \prod_{i<k+1} m_i = \prod_{i<k} m_i \cdot p_k$, entonces p tiene que dividir al primer factor (lo cual es imposible por hipótesis de inducción) o al segundo (lo cual es imposible por hipótesis). ∎

Teorema 6.44 *Todo número natural $n > 1$ se descompone como $n = \prod_{i=1}^{m} p_i$, donde cada p_i es primo. Además, la descomposición es única salvo el orden, en el sentido de que si $\prod_{i=1}^{m} p_i = \prod_{i=1}^{m'} p'_i$, entonces $m = m'$ y existe una biyección $\sigma : I_m \longrightarrow I_{m'}$ tal que $p_i = p'_{\sigma i}$.*

Demostración: Razonamos por inducción[17] que

$$n > 1 \to \bigvee m \in \mathbb{N} \bigvee p(p : I_m \longrightarrow \mathbb{N} \wedge \bigwedge i < m\ p_i \text{ es primo} \wedge n = \prod_{i<m} p_i).$$

Usamos el teorema 5.24: suponemos el resultado cierto para todo $k < n$. Si $n > 1$, por 5.36 existe un primo q tal que $q \mid n$. Pongamos que $n = qk$. Entonces $k < n$. Si $k = 1$ entonces $n = q$ es primo y el resultado es trivial. En caso contrario, por hipótesis de inducción $k = \prod_{i<m_0} p_i$, con cada p_i primo, luego $n = \prod_{i<m} p_i \cdot q = \prod_{i<m+1} p^*_i$, donde $p^* = p^\frown\langle q\rangle$.

Supongamos ahora que $\prod_{i=1}^{m} p_i = \prod_{i=1}^{m'} p'_i$. No perdemos generalidad si suponemos que $m \leq m'$. Vamos a probar que[17]

$$k \leq m \to \bigvee \sigma\ (\sigma : I_k \longrightarrow I_{m'} \text{ inyectiva} \wedge \bigwedge i < k\ p_i = p'_{\sigma i}).$$

Para $k = 0$ es trivial. Supongamos que $\sigma : I_k \longrightarrow I_{m'}$ cumple lo pedido y que $k + 1 \leq m$. Entonces

$$\prod_{i=0}^{k} p_i \cdot \prod_{i=k}^{m} p_i = \prod_{i\in\sigma[I_k]} p'_i \cdot \prod_{i\in I_{m'}\setminus\sigma[I_k]} p'_i.$$

[17] La fórmula es Σ_1.

Los primeros factores de cada miembro son iguales, por la propiedad conmutativa generalizada, luego

$$p_k \mid \prod_{i=k}^{m} p_i = \prod_{i\in I_{m'}\setminus\sigma[I_k]} p'_i.$$

Por el teorema anterior existe un $i \in I_{m'} \setminus \sigma[I_k]$ tal que $p_k = p'_i$, por lo que $\sigma \cup \{(k,i))\}$ sigue siendo una aplicación inyectiva y cumple lo pedido. En definitiva, llegamos a que existe una aplicación inyectiva $\sigma : I_m \longrightarrow I_{m'}$ tal que $p = \sigma \circ p'$. Por lo tanto,

$$\prod_{i=1}^{m} p_i = \prod_{i=1}^{m'} p'_i = \prod_{i\in\sigma[I_m]} p'_i \cdot \prod_{i\in I_{m'}\setminus\sigma[I_m]} p'_i = \prod_{i=1}^{m} p_i \cdot \prod_{i\in I_{m'}\setminus\sigma[I_m]} p'_i,$$

luego $\displaystyle\prod_{i\in I_{m'}\setminus\sigma[I_m]} p'_i = 1$, y esto sólo puede ser si $I_{m'} \setminus \sigma[I_m] = \varnothing$, es decir, si $\sigma : I_m \longrightarrow I_{m'}$ es biyectiva, lo cual implica que $m = m'$. ∎

Dejamos a cargo del lector continuar, si lo desea, en la formalización de la aritmética de los números primos. Por ejemplo, es fácil definir el exponente $v_p(n)$ con que el primo p divide al número n y demostrar sus propiedades, se puede definir la descomposición de un número en producto de potencias de primos distintos, etc.

Números enteros Ahora mostraremos cómo es posible definir en $\mathrm{I}\Sigma_1$ o en KP los números enteros y su aritmética. Para ello consideramos la fórmula Δ_1 dada por

$$x \sim y \equiv x, y \in \mathbb{N} \wedge x_0 + y_1 = x_1 + y_0,$$

donde x_0, x_1 son las funciones definidas en 5.16, que cumplen que $x = \langle x_0, x_1\rangle$. Claramente es una fórmula Δ_1 y es fácil probar que, para todos los números naturales x, y, z se cumple

$$x \sim x, \qquad x \sim y \to y \sim x, \qquad x \sim y \wedge y \sim z \to x \sim z.$$

Definimos

$$x \in \mathbb{Z} \equiv x \in \mathbb{N} \wedge \bigwedge y < x\ \neg x \sim y.$$

En otras palabras, llamaremos números enteros a los pares (aritméticos)[18] de números naturales que no están relacionados con ningún par menor respecto de la relación $\sim$ que acabamos de definir. Claramente $x \in \mathbb{Z}$ es una fórmula Δ_1. Como la relación $\sim$ también es Δ_1 y se cumple $x \sim x$, sabemos que

$$\bigwedge x \in \mathbb{N} \overset{1}{\bigvee} y \in \mathbb{N}(y \sim x \wedge \bigwedge z < y\ \neg y \sim z),$$

[18]Por conveniencia usamos pares aritméticos tanto en $\mathrm{I}\Sigma_1$ como en KP. Para usar pares conjuntistas en KP tendríamos que definir un buen orden entre ellos, lo cual es posible (por ejemplo, a través de los pares aritméticos), pero es más sencillo usar directamente los pares aritméticos.

lo que equivale a que $\bigwedge x \in \mathbb{N} \overset{1}{\bigvee} y \in \mathbb{Z}\ y \sim x$. Esto nos permite definir[19]

$$[x] \equiv y \in \mathbb{Z} \mid x \sim y,$$

y se cumple que $\bigwedge x \in \mathbb{N}\ ([x] \in \mathbb{Z} \wedge x \sim [x])$. Es fácil ver que

$$\bigwedge xy \in \mathbb{N}\ ([x] = [y] \leftrightarrow x \sim y).$$

En efecto, si $x \sim y$, entonces $[x] \sim x \sim y \sim [y]$, luego $[x] \sim [y]$, pero si fuera $[x] < [y]$ o $[y] < [x]$ no podría ser $[x] \sim [y]$, por definición de número entero, luego tiene que ser $[x] = [y]$.

Abreviaremos $[a, b] \equiv [\langle a, b\rangle]$. En estos términos, si $a, b, c, d \in \mathbb{N}$, se cumple la siguiente relación fundamental:

$$[a, b] = [c, d] \leftrightarrow a + d = b + c.$$

Definimos

$$+n \equiv [n, 0], \qquad -n \equiv [0, n].$$

Así, si $b \leq a$, se cumple que $[a, b] = [a - b, 0]$, mientras que si $a \leq b$ entonces $[a, b] = [0, b - a]$. Esto significa que todo número entero es de la forma $\pm n$, para un cierto $n \in \mathbb{N}$. Más en general, definimos

$$-x \equiv [x_1, x_0],$$

de modo que $\bigwedge ab \in \mathbb{N}\ -[a, b] = [b, a]$. Así, para cada número natural n se cumple que $-(+n) = -n$ y $-(-n) = +n$.

La tabla siguiente muestra los primeros diez números naturales, sus expresiones como pares ordenados y sus expresiones como números enteros cuando lo son. Cuando no lo son se indica el número entero con el que están relacionados:

0	$\langle 0, 0\rangle$	0	5	$\langle 2, 0\rangle$	$+2$
1	$\langle 0, 1\rangle$	-1	6	$\langle 0, 3\rangle$	-3
2	$\langle 1, 0\rangle$	$+1$	7	$\langle 1, 2\rangle$	~ 1
3	$\langle 0, 2\rangle$	-2	8	$\langle 2, 1\rangle$	~ 2
4	$\langle 1, 1\rangle$	~ 0	9	$\langle 3, 0\rangle$	$+3$

Así, el número natural 0 es el número entero $+0 = -0$, pero no hay confusión si lo llamamos simplemente 0. En cambio, no debemos confundir el número entero $+2$, que es el número natural 5, con el número natural 2, que es el número entero $+1$.

[19] En $\mathrm{I}\Sigma_1$ tenemos que $[x]$ es siempre una descripción propia, porque x siempre es un número natural. En cambio, en KP tenemos que

$$y = [x] \leftrightarrow (x \in \mathbb{N} \wedge y \in \mathbb{Z} \wedge x \sim y) \vee (x \notin \mathbb{N} \wedge y = 0).$$

En cualquier caso, la fórmula es Δ_1, luego el término $[x]$ también lo es.

Consideramos ahora el término Δ_1 dado por:[20]

$$x + y \equiv [x_0 + y_0, x_1 + y_1].$$

Es claro que $\bigwedge xy \in \mathbb{Z}\ x+y \in \mathbb{Z}$ y además si $a, b, c, d \in \mathbb{N}$ se cumple la relación:[21]

$$[a, b] + [c, d] = [a + c, b + d].$$

Una comprobación rutinaria muestra las propiedades siguientes:

1. $\bigwedge xyz \in \mathbb{Z}\ ((x + y) + z = x + (y + z)$,
2. $\bigwedge xy \in \mathbb{Z}(x + y = y + x)$,
3. $\bigwedge x \in \mathbb{Z}\ x + 0 = x$,
4. $\bigwedge x \in \mathbb{Z}\ x + (-x) = 0$.

En la práctica escribiremos $n - m \equiv n + (-m)$. Observemos que

$$\bigwedge ab \in \mathbb{N}\ [a, b] = +a - b.$$

Dejamos a cargo del lector los detalles análogos de la definición del producto de números enteros (que es también un término Δ_1) caracterizado por la relación

$$[a, b] \cdot [c, d] = [ac + bd, bc + ad],$$

así como la comprobación de sus propiedades elementales.

Consideramos la fórmula Δ_1

$$x \leq y \equiv x \in \mathbb{Z} \wedge y \in \mathbb{Z} \wedge x_0 + y_1 \leq x_1 + y_0.$$

Se comprueba que $\bigwedge abdc \in \mathbb{N}\ [a, b] \leq [c, d] \leftrightarrow a + d \leq b + c$, así como que se cumplen las propiedades usuales del orden de los números enteros.

Consideramos la fórmula Δ_1 dada por $x \in \mathbb{N}^+ \equiv \bigvee n \leq x\ x = +n$. Es fácil probar lo siguiente:

1. $\bigwedge n \in \mathbb{N}\ + n \in \mathbb{N}^+$,
2. $\bigwedge x \in \mathbb{N}^+ \bigvee n \in \mathbb{N}\ x = +n$,
3. $\bigwedge mn \in \mathbb{N}\ (+m = +n \to m = n)$,

[20] Claramente

$$z = x + y \leftrightarrow (x \in \mathbb{N} \wedge y \in \mathbb{N} \wedge \bigvee u_1 u_2 v_1 v_2 w_1 w_2 (x = \langle u_1, u_2\rangle \wedge y = \langle v_1, v_2\rangle \wedge$$
$$w_1 = u_1 + v_2 \wedge w_2 = u_2 + v_2 \wedge z = [w_1, w_2])) \vee ((x \notin \mathbb{N} \vee y \notin \mathbb{N}) \wedge z = 0) \leftrightarrow$$
$$(x \in \mathbb{N} \wedge y \in \mathbb{N} \wedge \bigwedge u_1 u_2 v_1 v_2 w_1 w_2 (x = \langle u_1, u_2\rangle \wedge y = \langle v_1, v_2\rangle \wedge$$
$$w_1 = u_1 + v_2 \wedge w_2 = u_2 + v_2 \to z = [w_1, w_2])) \vee ((x \notin \mathbb{N} \vee y \notin \mathbb{N}) \wedge z = 0).$$

[21] No es difícil probarlo, pero no es inmediato, pues hay que comprobar que si se cumple $[a, b] = \langle a_0, b_0\rangle$ y $[c, d] = \langle c_0, d_0\rangle$, entonces $[a, b] + [c, d] = [a_0 + c_0, b_0 + d_0] = [a + c, b + d]$.

4. $\bigwedge mn \in \mathbb{N}\ (+(m+n) = +m + (+n))$,
5. $\bigwedge mn \in \mathbb{N} + (mn) = (+m)(+n)$,
6. $\bigwedge mn \in \mathbb{N}\ (m \leq n \leftrightarrow +m \leq +n)$.

Cuando no se preste a confusión, escribiremos $0, 1, 2, \ldots$ para referirnos a los números enteros $0, +1, +2, \ldots$

Números racionales Similarmente, vamos a describir de forma breve cómo se puede formalizar en KP o en $\mathrm{I}\Sigma_1$ la aritmética de los números racionales. Para ello definimos[22]

$$x \sim y \equiv x, y \in \mathbb{N} \wedge (x+1)_0, (x+1)_1, (y+1)_0, (y+1)_1 \in \mathbb{Z} \wedge$$

$$(x+1)_1 \neq 0 \wedge (y+1)_1 \neq 0 \wedge (x+1)_0(y+1)_1 = (y+1)_0(x+1)_1.$$

Tras las comprobaciones análogas a las del apartado precedente, definimos

$$x \in \mathbb{Q} \equiv x \in \mathbb{N} \wedge (x+1)_0, (x+1)_1 \in \mathbb{Z} \wedge (x+1)_1 \neq 0 \wedge \bigwedge y < x\ \neg y \sim x.$$

Ahora se comprueba que $\bigwedge ab \in \mathbb{Z}(b \neq 0 \to \overset{1}{\bigvee} x \in \mathbb{Q}\ x \sim \langle a, b\rangle)$. A su vez, esto nos permite definir

$$\frac{a}{b} \equiv x \in \mathbb{Q} \mid x \sim \langle a, b\rangle - 1,$$

que es una descripción propia siempre que $a, b \in \mathbb{Z} \wedge b \neq 0$. Además, si $a, b, c, d \in \mathbb{Z}$ y $b, d \neq 0$, se cumple la relación fundamental:

$$\frac{a}{b} = \frac{c}{d} \leftrightarrow ad = bc.$$

La tabla siguiente muestra los diez primeros números naturales, x, la expresión de $x+1$ como par de números naturales, su expresión como par de números enteros cuando lo es y su expresión como número racional cuando lo es.

0	$\langle 0,1\rangle$	$\langle 0,-1\rangle$	0	5	$\langle 0,3\rangle$	$\langle 0,-2\rangle$	~ 0
1	$\langle 1,0\rangle$	$\langle -1,0\rangle$	$-$	6	$\langle 1,2\rangle$	$\langle -1,+1\rangle$	-1
2	$\langle 0,2\rangle$	$\langle 0,+1\rangle$	~ 0	7	$\langle 2,1\rangle$	$\langle +1,-1\rangle$	~ 6
3	$\langle 1,1\rangle$	$\langle -1,-1\rangle$	1	8	$\langle 3,0\rangle$	$\langle -2,0\rangle$	$-$
4	$\langle 2,0\rangle$	$\langle +1,0\rangle$	$-$	9	$\langle 0,4\rangle$	$-$	$-$

Vemos que el número natural 0 es el número racional $0/(-1) = 0/(+1)$, al que podemos representar sin ambigüedad como 0. En cambio, el número natural 3 es el número racional $+1/(+1)$, que representaremos simplemente como 1.

[22]Notemos que si en la definición pusiéramos x, y donde hemos puesto $x+1$, $y+1$ no se cumpliría nunca con $x = 0$ o $y = 0$, porque $0 = \langle 0, 0\rangle$ y no se cumpliría la condición $x_1 \neq 0$ o $y_1 \neq 0$. En cambio, así $x = 0$ cumple $x + 1 = \langle 0, 1\rangle$, y verifica la condición $(x+1)_1 \neq 0$, así como que $(x+1)_0 = 0 \in \mathbb{Z}$, $(x+1)_1 = 1 \in \mathbb{Z}$ (pues hemos visto que los números naturales $0, 1$, son, respectivamente, los números enteros 0 y -1).

La suma y el producto de números racionales se definen mediante los términos Δ_1 siguientes:

$$x + y = \frac{x_0y_1 + y_0x_1}{x_1y_1}, \qquad xy = \frac{x_0y_0}{x_1y_1},$$

de modo que si $a, b, c, d \in \mathbb{N}$ se cumple:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}, \qquad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}.$$

Dejamos al lector la comprobación de las propiedades básicas de estas operaciones. En particular, si definimos

$$-x \equiv \frac{-(x+1)_0}{(x+1)_1}, \qquad x^{-1} \equiv \frac{(x+1)_1}{(x+1)_0},$$

se cumple que

$$\bigwedge x \in \mathbb{Q}(x + (-x) = 0 \wedge x \cdot 1 = x), \qquad \bigwedge x \in \mathbb{Q}(x \neq 0 \to xx^{-1} = 1).$$

Teniendo en cuenta que

$$\frac{a}{-b} = \frac{-a}{b} = -\frac{a}{b},$$

resulta que todo número racional es de la forma a/b con $a, b \in \mathbb{Z}$ y $b > 0$.

Consideramos la fórmula Δ_1 dada por

$$x \leq y \equiv \bigvee abcd \in \mathbb{Z}(b > 0 \wedge d > 0 \wedge x = a/b \wedge y = c/d \wedge ad \leq bc)$$

$$\leftrightarrow \bigwedge abcd \in \mathbb{Z}(b > 0 \wedge d > 0 \wedge x = a/b \wedge y = c/d \to ad \leq bc).$$

Se comprueba sin dificultad que esta fórmula satisface las propiedades usuales del orden de los números racionales.

Definimos la clase $\mathbb{Z}_1 = \{n/1 \mid n \in \mathbb{Z}\}$ o, equivalentemente, la fórmula Δ_1 dada por

$$x \in \mathbb{Z}_1 \equiv \bigvee n \leq x(n \in \mathbb{Z} \wedge x = n/1),$$

donde hay que entender que los denominadores 1 son el número entero $+1$.

Hemos visto que el número racional $0/1$ es el número natural 0. Las propiedades siguientes se demuestran sin dificultad:

1. $\bigwedge n \in \mathbb{Z}\ n/1 \in \mathbb{Z}_1$,
2. $\bigwedge x \in \mathbb{Z}_1 \bigvee n \in \mathbb{Z}\ x = n/1$,
3. $\bigwedge mn \in \mathbb{Z}\ (m/1 = n/1 \to m = n)$,
4. $\bigwedge mn \in \mathbb{Z}\ (m + n)/1 = m/1 + n/1$,
5. $\bigwedge mn \in \mathbb{Z}\ (mn)/1 = (m/1)(n/1)$,
6. $\bigwedge mn \in \mathbb{Z}\ (m \leq n \leftrightarrow m/1 \leq n/1)$.

En particular, si consideramos la clase $\mathbb{N}_1 = \{n/1 \mid n \in \mathbb{N}^+\}$, las propiedades anteriores valen igualmente cambiando $\mathbb{Z}$ por $\mathbb{N}$ y entendiendo que $n/1 \equiv (+n)/1$.

A partir de aquí escribiremos $x \in \mathbb{Z}$ para representar la fórmula $x \in \mathbb{Z}_1$ y usaremos la notación $m/n \equiv m \cdot n^{-1}$, donde $m, n \in \mathbb{Q}$, $n \neq 0$. En estos términos se cumple que $n/1 = n$, para todo $n \in \mathbb{Q}$, así como que todo número racional es de la forma m/n, con $m, n \in \mathbb{Z}$, $n \neq 0$. Para cada $n \in \mathbb{N}$ abreviaremos $n_1 \equiv +n/1 \in \mathbb{Q}$, de modo que, si definimos

$$\mathbb{Z}^+ = \{n_1 \mid n \in \mathbb{N} \wedge n > 0\}, \qquad \mathbb{Z}^- \equiv \{-n_1 \mid n \in \mathbb{N} \wedge n > 0\},$$

tenemos que $\mathbb{Z} = \mathbb{Z}^- \cup \{0\} \cup \mathbb{Z}^-$ (unión disjunta) y

$$\mathbb{Q} = \{m/n \mid m, n \in \mathbb{Z} \wedge n \neq 0\} = \{m/n \mid m \in \mathbb{Z} \wedge n \in \mathbb{Z}^+\}.$$

En estos términos se cumple que

1. $\bigwedge n \in \mathbb{N}\ n_1 \in \mathbb{N}_1$,
2. $\bigwedge x \in \mathbb{N}_1 \bigvee n \in \mathbb{N}\ x = n_1$,
3. $\bigwedge mn \in \mathbb{N}\ (m_1 = n_1 \to m = n)$,
4. $\bigwedge mn \in \mathbb{N}\ ((m+n)_1 = m_1 + n_1)$,
5. $\bigwedge mn \in \mathbb{N}\ (mn)_1 = (m_1)(n_1)$,
6. $\bigwedge mn \in \mathbb{N}\ (m \leq n \leftrightarrow m_1 \leq n_1)$.

Estas propiedades nos permiten identificar los objetos que cumplen $x \in \mathbb{N}$ con los que cumplen $x \in \mathbb{N}_1$, pero vamos a detallar lo que esto supone.

Si trabajamos en $\mathrm{I}\Sigma_1$, para cada expresión θ de $\mathcal{L}_a$, llamamos $\bar{\theta} \equiv \theta$, mientras que si trabajamos en KP llamamos $\bar{\theta} \equiv \theta_{\mathrm{tc}}$ a la traducción de θ a KP como teoría aritmética. En ambos casos, llamamos $\tilde{\theta}$ a la traducción en el sentido de la definición 3.28 calculada con la fórmula $x \in \mathbb{N}_1$ y los términos 0, $x+1/1$, $x+y$, $x \cdot y$, donde todas las operaciones son las definidas para números racionales.

Teorema 6.45 *Sea $\theta(x_1, \ldots, x_n)$ una expresión de $\mathcal{L}_a$ cuyas variables libres están entre las indicadas. Si θ es un término, en $\mathrm{I}\Sigma_1$ o* KP *se demuestra*

$$\bigwedge x_1 \cdots x_n \in \mathbb{N}(\bar{\theta}(x_1, \ldots x_n)_1 = \tilde{\theta}((x_1)_1, \cdots (x_n)_1)),$$

y si θ es una fórmula

$$\bigwedge x_1 \cdots x_n \in \mathbb{N}(\bar{\theta}(x_1, \ldots x_n)_1 \leftrightarrow \tilde{\theta}((x_1)_1, \cdots (x_n)_1)).$$

Demostración: Razonamos por inducción sobre la longitud de θ.

1. Si $\theta \equiv x_i$, trivialmente, $\bigwedge x_1 \cdots x_n \in \mathbb{N}((x_i)_1 = (x_i)_1)$.

2. Si $\theta \equiv 0$, trivialmente $\bigwedge x_1 \cdots x_n\, 0 = 0$.

3. Si $\theta \equiv t_1 = t_2$, por hipótesis de inducción, si $x_1, \ldots, x_n \in \mathbb{N}$, tenemos que

$$\bar{t}_i(x_1, \ldots, x_n)_1 = \tilde{t}_i((x_1)_1, \ldots, (x_n)_1),$$

luego

$$\bar{\theta}(x_1, \ldots, x_n)_1 \equiv \bar{t}_1(x_1, \ldots, x_n)_1 = \bar{t}_2(x_1, \ldots, x_n)_1 \leftrightarrow$$
$$\tilde{t}_1((x_1)_1, \ldots, (x_n)_1) = \tilde{t}_2((x_1)_1, \ldots, (x_n)_1) \equiv \tilde{\theta}((x_1)_1, \ldots, (x_n)_1).$$

4. Si $\theta \equiv t'$, por hipótesis de inducción, si $x_1, \ldots, x_n \in \mathbb{N}$, tenemos que

$$\bar{t}(x_1, \ldots, x_n)_1 = \tilde{t}((x_1)_1, \ldots, (x_n)_1),$$

luego, usando la propiedad 4 de la lista previa al enunciado,

$$\bar{\theta}(x_1, \ldots, x_n)_1 \equiv (\bar{t}(x_1, \ldots, x_n)'_1) = (\bar{t}(x_1, \ldots, t_n) + 1)_1 =$$
$$\bar{t}(x_1, \ldots, x_n)_1 + 1_1 = \tilde{t}((x_1)_1, \ldots, (x_n)_1) + 1_1 \equiv \tilde{\theta}((x_1)_1, \ldots, (x_n)_1).$$

5. Si $\theta \equiv t_1 + t_2$, por hipótesis de inducción, si $x_1, \ldots, x_n \in \mathbb{N}$, tenemos que

$$\bar{t}_i(x_1, \ldots, x_n)_1 = \tilde{t}_i((x_1)_1, \ldots, (x_n)_1),$$

luego, de nuevo por la propiedad 4:

$$\bar{\theta}_1(x_1, \ldots, x_n) \equiv (\bar{t}_1(x_1, \ldots, x_n) + \bar{t}_2(x_1, \ldots, x_n))_1 =$$
$$\bar{t}_1(x_1, \ldots, x_n)_1 + \bar{t}_2(x_1, \ldots, x_n)_1 =$$
$$\tilde{t}_1((x_1)_1, \ldots, (x_n)_1) + \tilde{t}_2((x_1)_1, \ldots, (x_n)_1) = \tilde{\theta}((x_1)_1, \ldots, (x_n)_1).$$

6. El caso $\theta \equiv t_1 \cdot t_2$ es análogo usando ahora la propiedad 5.

7. Si $\theta \equiv \neg\alpha$ o $\theta \equiv \alpha \to \beta$, de la hipótesis de inducción se sigue inmediatamente la conclusión.

8. Si $\theta \equiv \bigwedge x\alpha(x, x_1, \ldots, x_n)$, por hipótesis de inducción

$$\bigwedge x x_1 \cdots x_n \in \mathbb{N}(\bar{\alpha}(x, x_1, \ldots, x_n)_1 \leftrightarrow \tilde{\alpha}((x)_1, (x_1)_1, \ldots, (x_n)_1)).$$

Fijados $x_1, \ldots, x_n \in \mathbb{N}$, tenemos que probar que

$$\bigwedge x \in \mathbb{N}\, \bar{\alpha}(x, x_1, \ldots, x_n)_1 \leftrightarrow \bigwedge x \in \mathbb{N}_1\, \tilde{\alpha}(x, (x_1)_1, \ldots, (x_n)_n).$$

Supuesto el miembro izquierdo, dado $x \in \mathbb{N}_1$, tenemos que existe un $x_0 \in \mathbb{N}$ tal que $x = (x_0)_1$, y por la hipótesis se cumple $\bar{\alpha}(x_0, x_1, \ldots, x_n)_1$, que por hipótesis de inducción equivale a $\tilde{\alpha}(x, (x_1)_1, \ldots, (x_n)_1)$, luego se cumple $\bigwedge x \in \mathbb{N}_1\, \tilde{\alpha}(x, (x_1)_1, \ldots, (x_n)_1)$. La implicación opuesta es similar.

9. Si $\theta \equiv x|\alpha(x, x_1, \ldots, x_n)$, por hipótesis de inducción

$$\bigwedge x x_1 \cdots x_n \in \mathbb{N}(\bar{\alpha}(x, x_1, \ldots, x_n)_1 \leftrightarrow \tilde{\alpha}((x)_1, (x_1)_1, \ldots, (x_n)_1)).$$

De aquí se sigue que

$$\overset{1}{\bigvee} x \in \mathbb{N}\, \bar{\alpha}(x, x_1, \ldots, x_n)_1 \leftrightarrow \overset{1}{\bigvee} x \in \mathbb{N}_1\, \tilde{\alpha}(x, (x_1)_1, \ldots, (x_n)_1).$$

Más precisamente, fijados $x_1, \ldots, x_n \in \mathbb{N}$, si x es el único número natural que cumple $\bar{\alpha}(x, x_1, \ldots, x_n)_1$, entonces $(x)_1$ es el único elemento de $\mathbb{N}_1$ que cumple $\tilde{\alpha}((x)_1, (x_1)_1, \ldots, (x_n)_1)$. Por lo tanto, en este caso

$$\bar{\theta}(x_1, \ldots, x_n)_1 \equiv x|(x \in \mathbb{N} \wedge \bar{\alpha}(x, x_1, \ldots, x_n))_1 =$$

$$x|(x \in \mathbb{N}_1 \wedge \tilde{\alpha}(x, (x_1)_1, \ldots, (x_n)_1) = \tilde{\alpha}((x_1)_1, \ldots, (x_n)_1).$$

Si no se dan las unicidades, trivialmente

$$\bar{\theta}(x_1, \ldots, x_n)_1 \equiv x|(x \in \mathbb{N} \wedge \bar{\alpha}(x, x_1, \ldots, x_n))_1 = 0 =$$

$$x|(x \in \mathbb{N}_1 \wedge \tilde{\alpha}(x, (x_1)_1, \ldots, (x_n)_1) = \tilde{\alpha}((x_1)_1, \ldots, (x_n)_1).$$

■

En particular, para toda sentencia α de $\mathcal{L}_a$, tenemos que en $\mathrm{I}\Sigma_1$ o en KP se demuestra $\bar{\alpha} \leftrightarrow \tilde{\alpha}$. Esto significa que $\mathbb{N}$ y $\mathbb{N}_1$ verifican las mismas (traducciones de) sentencias aritméticas. Esto tiene muchas consecuencias:

1. En $\mathrm{I}\Sigma_1$ y en KP se demuestra que $\mathbb{N}_1$ cumple las traducciones de los axiomas de $\mathrm{I}\Sigma_1$, por lo que también son sistemas aritméticos (con la inducción restringida a fórmulas Σ_1 en sentido aritmético) tomando como números naturales los de $\mathbb{N}_1$ y sus operaciones.

2. Más aún AP y Z^* son teorías aritméticas (con inducción para todas las fórmulas aritméticas) con $\mathbb{N}_1$.

3. De hecho, en AP y en Z^* se cumple el principio de inducción para $\mathbb{N}_1$ para fórmulas arbitrarias, pues si probamos $\alpha(0)$ y $\bigwedge n \in \mathbb{N}_1(\alpha(n) \to \alpha(n+1))$, tenemos también $\alpha(0_1)$ y $\bigwedge n \in \mathbb{N}(\alpha(n_1) \to \alpha((n+1)_1)$, luego, por el principio de inducción en $\mathbb{N}$ para la fórmula $\alpha(n_1)$, concluimos que se cumple $\bigwedge n \in \mathbb{N}\, \alpha(n_1)$, que a su vez equivale a $\bigwedge n \in \mathbb{N}_1\, \alpha(n)$.

   Y el hecho que más nos va a interesar:

4. Si en cualquier extensión de $\mathrm{I}\Sigma_1$ o KP probamos una sentencia aritmética para $\mathbb{N}_1$, sabemos que se cumple también para $\mathbb{N}$.

Estos hechos permiten que en la práctica no necesitemos distinguir más entre $\mathbb{N}$ y $\mathbb{N}_1$, de modo que podemos considerar $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q}$.

Notemos que las sumas y productos finitos definidas en la sección anterior son aplicables a la suma y el producto de números racionales.

Nota Hemos visto que en $\mathrm{I}\Sigma_1$ los números naturales pueden verse como conjuntos finitos, de modo que se cumplen los axiomas de KP (o los de ZFC$-$AI si trabajamos en AP). En particular, podemos considerar la fórmula $x \in \omega$ que define a los números naturales en KP como una fórmula de $\mathcal{L}_a$ con la que $\mathrm{I}\Sigma_1$ es también un sistema aritmético con inducción restringida a fórmulas Σ_1 (y AP es un sistema aritmético con inducción para todas las fórmulas aritméticas). Un argumento análogo al que hemos usado para comparar $\mathbb{N}$ y $\mathbb{N}_1$ nos permite llegar a que se da la misma relación entre $\mathbb{N}$ y ω, es decir, que en $\mathrm{I}\Sigma_1$ (resp. AP) se puede probar una sentencia aritmética si y sólo si su traducción es demostrable en KP (resp. ZFC$-$AI). No vamos a necesitar estos hechos, pero los detalles están en [LF 5.48], [LF 5.49], [LF 5.51].

También hemos visto que, además de los axiomas de KP (resp. ZFC$-$AI), en $\mathrm{I}\Sigma_1$ (resp. AP) se puede probar (teorema 6.22) que todo conjunto es finito, lo cual no es un teorema de KP ni de ZFC$-$AI. Si llamamos $\mathrm{KP}_{\mathrm{fin}}$ y $\mathrm{ZFC}_{\mathrm{fin}}$ a KP y ZFC$-$AI más el axioma "todo conjunto es finito", se cumple [LF 5.76], [LF 5.79] que una sentencia de $\mathcal{L}_{\mathrm{tc}}$ es demostrable en $\mathrm{KP}_{\mathrm{fin}}$ (resp. $\mathrm{ZFC}_{\mathrm{fin}}$) si y sólo si es demostrable en $\mathrm{I}\Sigma_1$ (resp. AP).

Estos resultados vienen a decir que AP y $\mathrm{ZFC}_{\mathrm{fin}}$ (o $\mathrm{I}\Sigma_1$ y $\mathrm{KP}_{\mathrm{fin}}$) son esencialmente la misma teoría: los números naturales pueden verse como conjuntos finitos a partir de los cuales pueden definirse números naturales identificables con los números de partida. En otras palabras, que los números naturales pueden envolverse en un marco conjuntista aparentemente más amplio (y más cómodo para trabajar), pero que en realidad no aporta nada esencialmente nuevo. ∎

Cuerpos cuadráticos Al final del capítulo siguiente necesitaremos razonar en $\mathrm{I}\Sigma_1$ con cuerpos cuadráticos de la forma $\mathbb{Q}(\sqrt{d}\,)$, para un número natural d que no sea un cuadrado perfecto. Vamos a mostrar aquí que, en efecto, la aritmética de estos cuerpos es formalizable en $\mathrm{I}\Sigma_1$.

Observemos ante todo que el hecho de que d no sea el cuadrado de un número natural implica que tampoco es el cuadrado de un número racional, pues si $d = p^2/q^2$, entonces el exponente de cada primo en p es diferencia de dos números pares, luego es par, y esto hace que d sea un cuadrado perfecto.

Definimos la fórmula $x \in \mathbb{Q}(\sqrt{d}\,) \equiv x_0 \in \mathbb{Q} \wedge x_1 \in \mathbb{Q}$ y a continuación definimos una suma y un producto mediante

$$x + y \equiv z|(z \in \mathbb{Q}(\sqrt{d}\,) \wedge z_0 = x_0 + y_0 \wedge z_1 = x_1 + y_1),$$

$$x \cdot y \equiv z|(z \in \mathbb{Q}(\sqrt{d}\,) \wedge z_0 = x_0y_0 + x_1y_1d \wedge z_1 = x_0y_1 + x_1y_0).$$

Es fácil ver que estas operaciones cumplen las propiedades de la definición de anillo (asociativa, conmutativa, distributiva, etc.). Más aún, si definimos $\bar{x} \equiv \langle x, 0\rangle$, $\sqrt{d} \equiv \langle 0, 1\rangle$ tenemos que:

1. $\bigwedge x \in \mathbb{Q}(\sqrt{d}\,)\ x = \bar{x}_0 + \bar{x}_1\sqrt{d}$,
2. $\sqrt{d}\sqrt{d} = \bar{d}$,
3. $\bigwedge xy \in \mathbb{Q}\ \overline{x+y} = \bar{x} + \bar{y}$,
4. $\bigwedge xy \in \mathbb{Q}\ \overline{x \cdot y} = \bar{x} \cdot \bar{y}$.

Esto nos permite identificar los números racionales con los elementos de la clase $\overline{\mathbb{Q}} = \{\bar{x} \mid x \in \mathbb{Q}\}$ y afirmar entonces que

$$x \in \mathbb{Q}(\sqrt{d}\,) \leftrightarrow \bigvee ab \in \mathbb{Q}\ x = a + b\sqrt{d}$$

En estos términos

$$(a + b\sqrt{d}) + (a' + b'\sqrt{d}) = a + a' + (b + b')\sqrt{d},$$

$$(a + b\sqrt{d})(a' + b'\sqrt{d}) = aa' + bb'd + (ab' + ba')\sqrt{d}.$$

A partir de aquí se demuestran fácilmente las propiedades algebraicas básicas de los cuerpos cuadráticos. Por ejemplo, el hecho de que sean cuerpos (que todo elemento no nulo tenga inverso) se deduce de que

$$(a + b\sqrt{d})(a - b\sqrt{d}) = a^2 - b^2 d \neq 0$$

si $a + b\sqrt{d} \neq 0$ (porque d no es un cuadrado), por lo que

$$(a + b\sqrt{d})^{-1} = \frac{a}{a^2 - b^2 d} + \frac{-b}{a^2 - b^2 d}\sqrt{d}.$$

Un poco más delicado es comprobar que la relación de orden en $\mathbb{Q}$ puede extenderse a una relación de orden en $\mathbb{Q}(\sqrt{d}\,)$, es decir, que podemos definir una fórmula $x \leq y$ para $x, y \in \mathbb{Q}(\sqrt{d}\,)$ para la que se demuestran las propiedades de una relación de orden total y de modo que sobre $\mathbb{Q}$ coincide con la que ya tenemos definida.

Para probarlo empezamos observando que

$$\bigwedge uv \in \mathbb{Q}(0 \leq u < v \to \bigvee r \in \mathbb{Q}(u < r^2 < v)).$$

En efecto, sea $k \in \mathbb{N}$ tal que $v < k^2$. Fijamos un $n \in \mathbb{N}$ no nulo y, como $v \leq (kn)^2/n^2$, existe un mínimo $m_0 \in \mathbb{N}$ tal que $v \leq m_0^2/n^2$, No puede ser $m_0 = 0$, luego $m_0 = m + 1$ y tenemos que

$$\left(\frac{m}{n}\right)^2 < v \leq \left(\frac{m+1}{n}\right)^2.$$

Por otra parte

$$\left(\frac{m+1}{n}\right)^2 - \left(\frac{m}{n}\right)^2 = \left(\frac{m+1}{n} + \frac{m}{n}\right)\left(\frac{m+1}{n} - \frac{m}{n}\right)$$

$$= \frac{2m+1}{n^2} < \frac{2kn+1}{n^2} = \frac{2k}{n} + \frac{1}{n^2}.$$

Tomando n suficientemente grande podemos hacer que el último término sea menor que $v - u$, con lo que $r = m/n$ cumple que

$$v - r^2 < \left(\frac{m+1}{n}\right)^2 - r^2 < v - u,$$

luego $u < r^2 < v$. ∎

Ahora probamos que si $a, b \in \mathbb{Q}$ no son ambos nulos, existe un $\epsilon > 0$ en $\mathbb{Q}$ tal que, o bien todo $r \in \mathbb{Q}$, $r > 0$, que cumpla $|r^2 - d| < \epsilon$ cumple $a + br > 0$, o bien todo r en dichas condiciones cumple $a + br < 0$. En el primer caso diremos que $a + b\sqrt{d} > 0$ y en el segundo que $a + b\sqrt{d} < 0$.

En efecto, si $b = 0$ es trivial, pues se cumplirá $a + b\sqrt{d} > 0$ o $a + b\sqrt{d} < 0$ según si $a > 0$ o $a < 0$. Supongamos que $b > 0$. Si $a \geq 0$ es trivial, pues todo $r > 0$ cumple $a + br > 0$. Si $a < 0$, como d no es un cuadrado, o bien $d > a^2/b^2$, o bien $d < a^2/b^2$. Llamamos $\epsilon = |d - a^2/b^2| > 0$. Así, si $|d - r^2| < \epsilon$, se cumple $r^2 > a^2/b^2$ o bien $r^2 < a^2/b^2$ según lo que cumpla d, independientemente de r. En el primer caso $r^2b^2 < a^2$, luego $rb < -a$, luego $a + br > 0$. Igualmente en el segundo caso se llega a que $a + br < 0$. Si $b < 0$ se razona análogamente.

Veamos que si $a + b\sqrt{d} > 0$ y $a' + b'\sqrt{d} > 0$, entonces $a + a' + (b + b')\sqrt{d} > 0$. En efecto, sea $\epsilon > 0$ que cumpla la definición para la suma. Podemos tomar $r > 0$ que cumpla $|r^2 - d| < \epsilon/2$ y que además esté lo suficientemente cerca de d para que $a + br > 0$, $a' + b'r > 0$. Entonces $a + a' + (b + b')r > 0$, luego se cumple lo indicado.

Con esto ya podemos definir $a + b\sqrt{d} < a' + b'\sqrt{d}$ si $a' - a + (b' - b)\sqrt{d} \geq 0$ (donde ≥ 0 significa > 0 o $= 0$). Ahora es fácil comprobar que se trata de una relación de orden total que extiende a la de los números racionales.

Observemos ahora que si $a + b\sqrt{d} > 0$, existe un número racional δ tal que $0 < \delta < a + b\sqrt{d}$. En efecto, podemos suponer que $b \neq 0$, pues en caso contrario la conclusión es trivial.

Tomamos r_0 de modo que $|d - r_0^2|$ sea lo suficientemente pequeño como para que $\delta = a + br_0 > 0$. Para que además se cumpla $\delta < a + b\sqrt{d}$ tiene que cumplirse que si $|d - r^2|$ es suficientemente pequeño, entonces $-br_0 + br = b(r - r_0) > 0$, lo cual a su vez equivale a que $r > r_0$ si $b > 0$ o bien $r < r_0$ si $b < 0$.

Para que se cumpla esto basta haber elegido r_0 de modo que $r_0^2 < d$ en el primer caso y $r_0^2 > d$ en el segundo. En efecto, en el primer caso, si exigimos que $|r^2 - d| < \epsilon = d - r_0^2$, entonces $r_0^2 - d < r^2 - d$, luego $r_0^2 < r^2$, luego $r_0 < r$, como se requiere. En el segundo caso exigimos que $|r^2 - d| < \epsilon = r_0^2 - d$, con lo que $r^2 - d < r_0^2 - d$, luego $r < r_0$, como se requiere.

Finalmente, observamos que si $a + b\sqrt{d} > 0$ y $a' + b'\sqrt{d} > 0$, entonces $(a + b\sqrt{d})(a' + b'\sqrt{d}) > 0$.

En efecto, fijamos un número racional $\delta > 0$ tal que

$$0 < \delta < a + b\sqrt{d}, \qquad 0 < \delta < a' + b'\sqrt{d}.$$

Entonces $a - \delta + b\sqrt{d} > 0$, $a' - \delta + b'\sqrt{d} > 0$, luego si $|r^2 - d|$ es suficientemente pequeño, se cumple que $a - \delta + br > 0$ y $a' - \delta + b'r > 0$ o, equivalentemente, $a + br > \delta$, $a' + b'r > \delta$. Por lo tanto,

$$aa' + bb'r^2 + (ab' + a'b)r > \delta^2.$$

Lo que queremos probar es que $aa' + bb'd + (ab' + a'b)r > 0$. Si $bb' = 0$ es trivial, y en caso contrario basta exigir que $|r^2 - d| < \delta^2/|bb'|$, pues entonces

$$aa' + bb'd + (ab' + a'b)r = aa' + bb'r^2 + (ab' + a'b)r + bb'(d - r^2) > \delta^2 - \delta^2 > 0.$$

En resumen, además de las propiedades de relación de orden total, ahora tenemos que

$$\bigwedge xyz \in \mathbb{Q}(\sqrt{d}\,)\ (x \leq y \to x + z \leq y + z),$$

$$\bigwedge xyz \in \mathbb{Q}(\sqrt{d}\,)\ (x \leq y \wedge z \geq 0 \to xz \leq yz).$$

De estas propiedades se deducen algebraicamente (sin necesidad de considerar aproximaciones racionales como hasta ahora) todas las propiedades usuales de la relación de orden.

El lector puede comprobar que en KP o $\mathrm{I}\Sigma_1$ es posible definir los números combinatorios (como términos Δ_1) y demostrar sus propiedades, entre ellas el teorema del binomio, etc. No entramos en detalles porque sólo usaremos esto puntualmente en la última sección del capítulo siguiente y, en cualquier caso, la idea que el lector debería extraer de lo visto aquí es que todos los argumentos aritméticos sobre números racionales (o sistemas numéricos próximos, como los cuerpos cuadráticos) son formalizables de forma natural en KP–$\mathrm{I}\Sigma_1$. La única limitación es que no podemos definir conjuntos infinitos, aunque sí que podemos tratar con fórmulas que caractericen clases infinitas, como las clases de los números enteros o racionales, según hemos visto.

Capítulo VII

La teoría de la recursión

La restricción de la inducción matemática a fórmulas de tipo Σ_1, requerida en la teoría $\mathrm{I}\Sigma_1$, con la consiguiente necesidad de comprobar que los conceptos aritméticos que manejamos son Σ_1 o, a ser posible, Δ_1, puede parecer artificial, pero en este capítulo vamos a probar que las relaciones y funciones Δ_1 tienen una interpretación natural que explica el interés de trabajar en $\mathrm{I}\Sigma_1$ (y no en AP) en la medida de lo posible. Vamos a probar que las relaciones y funciones Δ_1 coinciden con lo que se conoce como relaciones y funciones recursivas, que a su vez pueden caracterizarse como las relaciones y funciones que pueden calcularse en un tiempo finito mediante un algoritmo o, si se prefiere, las relaciones y funciones que puede calcular mecánicamente un ordenador con memoria suficiente para realizar los cálculos, y teniendo en cuenta que la cantidad de memoria requerida dependerá de la magnitud de los datos.

7.1 Funciones y relaciones recursivas

Todas las relaciones y funciones que consideramos aquí son relaciones y funciones n-ádicas (para un $n \geq 1$) sobre el conjunto $\mathbb{N}$ de los números naturales.

Funciones recursivas elementales Llamaremos *funciones recursivas elementales* a las funciones siguientes:

- La función monádica c, dada por $c(n) = 0$ para todo n.
- La función monádica s, dada por $s(n) = n + 1$ para todo n.
- Las funciones k-ádicas p_i^k para $1 \leq i \leq k$, dadas por $p_i^k(n_1, \ldots, n_k) = n_i$.

Observemos que todas las funciones recursivas elementales se pueden calcular explícitamente en cada caso concreto. Las funciones recursivas son las que pueden obtenerse a partir de éstas mediante la aplicación de un número finito de los procesos de definición que indicamos seguidamente.

Definición de funciones a) Una función k-ádica f está definida por *composición* a partir de la función r-ádica g y de las funciones k-ádicas $h_1, \ldots, h_r$ si para todos los naturales $a_1, \ldots, a_k$ se cumple que

$$f(a_1, \ldots, a_k) = g(h_1(a_1, \ldots, a_k), \ldots, h_r(a_1, \ldots, a_k)).$$

Claramente, si tenemos funciones g y $h_1, \ldots, h_r$, la ecuación anterior determina una función f sin ambigüedad alguna. Si disponemos de algoritmos para calcular las funciones g y h_i, es fácil diseñar a partir de ellos un algoritmo que calcule f: basta aplicar los algoritmos de las h_i para calcular las imágenes de los datos y aplicar el algoritmo de g a los resultados que obtengamos.

b) Una función $k+1$-ádica f está definida por *recursión* a partir de la función k-ádica g [o del natural a si $k = 0$] y de la función $k+2$-ádica h si para todos los naturales $a_1, \ldots, a_k$, n se cumple que

$$\begin{aligned} f(a_1, \ldots, a_k, 0) &= g(a_1, \ldots, a_k) \qquad [f(0) = a, \text{ si } k = 0] \\ f(a_1, \ldots, a_k, n+1) &= h(a_1, \ldots, a_k, n, f(a_1, \ldots, a_k, n)). \end{aligned}$$

Si tenemos funciones g, h [o un número a y una función h], las ecuaciones anteriores determinan unívocamente una función f. Si disponemos de algoritmos para calcular g y h también tenemos otro para calcular f: calculamos primero $f(a_1, \ldots, a_k, 0)$ con el algoritmo de g y después vamos calculando $f(a_1, \ldots, a_k, 1)$, $f(a_1, \ldots, a_k, 2)$, etc. mediante el algoritmo de h, hasta llegar a $f(a_1, \ldots, a_k, n)$.

c) Una función k-ádica f está definida por *minimización* a partir de una función $k+1$-ádica g si para todos los naturales $a_1, \ldots, a_k$ se cumple

1. Existe un n tal que $g(a_1, \ldots, a_k, n) = 0$,
2. $f(a_1, \ldots, a_k) = \mu n\ g(a_1, \ldots, a_k, n) = 0$,

donde μn es una abreviatura por "el mínimo n tal que..."

Dada una función g que cumpla 1), la ecuación 2) determina unívocamente una función f que será calculable mediante un algoritmo si lo es g: basta aplicar el algoritmo de g para calcular sucesivamente $g(a_1, \ldots, a_k, 0)$, $g(a_1, \ldots, a_k, 1)$, $g(a_1, \ldots, a_k, 2)$, etc. hasta encontrar el primer n que hace $g(a_1, \ldots, a_k, n) = 0$ (el cual existe por la condición 1.)

Funciones recursivas Una función f es *recursiva primitiva (recursiva)* si existe una sucesión de funciones $f_1, \ldots, f_n$ tales que f_n es f y para todo natural i entre 1 y n, la función f_i es recursiva elemental o bien f_i está definida por composición o recursión (o minimización) a partir de funciones anteriores de la sucesión. Es claro que toda función recursiva primitiva es recursiva.

Así, la única diferencia entre las funciones recursivas y las recursivas primitivas consiste en que en las primeras se admite la minimización como técnica de definición y en las segundas no.

Puesto que las funciones elementales se pueden calcular mediante algoritmos (elementales) y las funciones definidas por composición, recursión o minimización a partir de funciones calculables mediante algoritmos son también calculables mediante algoritmos, es claro que toda función recursiva es calculable mediante un algoritmo. Más concretamente, si f es una función recursiva, una sucesión de funciones $f_1, \ldots, f_n$ según la definición determina un algoritmo para calcular f (en el sentido de que conociendo la sucesión es fácil diseñar el algoritmo correspondiente).

En realidad Gödel llamó funciones recursivas a lo que nosotros hemos llamado funciones recursivas primitivas. Su definición no tenía más pretensión que la de sistematizar algunos resultados previos a sus teoremas de incompletitud, y el nombre de función recursiva aludía simplemente a que el rasgo más característico de estas funciones era que permiten las definiciones recurrentes. Posteriormente Herbrand introdujo lo que llamó funciones *recursivas generales,* cuya definición permitía procedimientos de construcción más generales, tales como recursiones simultáneas en varias variables, definiciones implícitas por sistemas de ecuaciones que cumplieran ciertos requisitos, etc., todo ello sin perder la propiedad de que las funciones así obtenidas eran calculables mediante algoritmos. La definición de Herbrand era tan amplia que resultaba natural conjeturar que cualquier función calculable mediante un algoritmo debía de ser recursiva general. Kleene demostró que las funciones recursivas generales de Herbrand coincidían con las que nosotros hemos definido como funciones recursivas, es decir, que toda la generalidad de la definición de Herbrand se obtenía igualmente sin más que añadir la minimización a la definición de Gödel. Desde entonces, los términos antiguos "función recursiva" y "función recursiva general" han sido sustituidos por "función recursiva primitiva" y "función recursiva", tal y como los hemos introducido nosotros. Finalmente, Turing demostró que las funciones recursivas son exactamente las calculables mediante un algoritmo. Esta afirmación se conoce como *Tesis de Church-Turing* y la probaremos más adelante, pero de momento hemos de empezar por estudiar las funciones recursivas.

Conviene resaltar la similitud formal entre la definición de función recursiva y la definición de teorema en una teoría axiomática: las funciones elementales son el equivalente a los axiomas y los métodos de construcción de funciones son el equivalente a las reglas de inferencia. Como en el caso del cálculo deductivo, la definición que hemos dado de función recursiva es en principio arbitraria, pero ya hemos justificado que no es totalmente arbitraria, ya que las funciones que hemos tomado como recursivas elementales son calculables mediante algoritmos, y los métodos de construcción nos garantizan que producen funciones calculables mediante algoritmos cuando se aplican a funciones calculables mediante algoritmos. Esto es el equivalente al teorema de corrección para el cálculo deductivo. La tesis de Church-Turing es el equivalente al teorema de completitud semántica, pues nos asegurará que son recursivas todas las funciones que "tienen que ser" recursivas, es decir, todas las calculables mediante un algoritmo.

Como en el caso del cálculo deductivo, unas observaciones elementales simplifican notablemente la manipulación de funciones recursivas. En primer lugar,

a la hora de mostrar que una función es recursiva primitiva (o recursiva) podemos admitir que en la sucesión se incorporen funciones que ya hayamos probado que son recursivas primitivas (recursivas) como abreviatura de la sucesión completa, que tendría, en lugar de dicha función, la sucesión que justifica su carácter recursivo.

Claramente, si f es una función n-ádica recursiva (primitiva) e $i_1, \ldots, i_n$ es una reordenación de $1, \ldots, n$, entonces la función g dada por $g(a_1, \ldots, a_n) = f(a_{i_1}, \ldots, a_{i_n})$ es recursiva (primitiva), pues

$$g(a_1, \ldots, a_n) = f(p^n_{i_1}(a_1, \ldots, a_n), \ldots, p^n_{i_n}(a_1, \ldots, a_n)),$$

es decir, g está definida por composición a partir de f y de las proyecciones.

Esto significa que no hemos de preocuparnos por el orden de los argumentos de las funciones. En particular si g y h son recursivas (primitivas) donde g es k-ádica y h es $k+2$-ádica, también será recursiva (primitiva) la función dada por

$$\begin{aligned} f(0, a_1, \ldots, a_k) &= g(a_1, \ldots, a_k), \\ f(n+1, a_1, \ldots, a_k) &= h(n, f(a_1, \ldots, a_k, n), a_1, \ldots, a_k). \end{aligned}$$

De hecho si hubiéramos definido la recursión con estas ecuaciones, la definición de función recursiva (primitiva) correspondiente sería equivalente. Por tanto en adelante usaremos la forma que consideremos más oportuna.

Veamos un primer ejemplo no trivial de función recursiva:

Teorema 7.1 *La suma de números naturales es una función diádica recursiva primitiva.*

Demostración:

$h_1(m) = m$		(p^1_1)
$h_2(m, n, p) = p$		(p^3_3)
$h_3(m) = m + 1$		(s)
$h_4(m, n, p) = h_3(h_2(m, n, p))$	$[= p + 1]$	(composición)
$h_5(m, 0) = h_1(m)$	$[= m]$	(recursión)
$h_5(m, n+1) = h_4(m, n, h_5(m, n))$	$[= h_5(m, n) + 1]$	

Claramente $h_5(m, n) = m + n$. ∎

En la práctica abreviaremos estas demostraciones expresando la función en términos de funciones ya probadas recursivas, sobrentendiendo las proyecciones donde proceda. Por ejemplo la prueba del teorema anterior se puede reducir a

$$m + 0 = m, \qquad m + (n + 1) = (m + n) + 1.$$

Aquí tenemos algunos ejemplos adicionales de funciones recursivas primitivas. Las indicaciones que damos son suficientes para justificar su carácter recursivo.

1) $m \cdot n \qquad m \cdot 0 = 0 \quad m \cdot (n+1) = m \cdot n + m.$

2) $c_a(n) = a \qquad c_a(0) = a \quad c_a(n+1) = c_a(n).$

3) $m^n \qquad m^0 = 1 \quad m^{n+1} = m^n \cdot m.$

4) $n! \qquad 0! = 1 \quad (n+1)! = n! \cdot (n+1).$

5) $\mathrm{sg}(n) = \begin{cases} 0 & \text{si } n = 0 \\ 1 & \text{si } n \neq 0 \end{cases} \qquad \overline{\mathrm{sg}}(n) = \begin{cases} 1 & \text{si } n = 0 \\ 0 & \text{si } n \neq 0 \end{cases}$

$\mathrm{sg}(0) = 0 \quad \mathrm{sg}(n+1) = c_1(n)$

$\overline{\mathrm{sg}}(0) = 1 \quad \overline{\mathrm{sg}}(n+1) = c_0(n)$

6) $\mathrm{pre}(0) = 0 \quad \mathrm{pre}(n+1) = n$

7) $m \dot{-} 0 = m \quad m \dot{-} (n+1) = \mathrm{pre}(m \dot{-} n)$

Relaciones recursivas Cuando hayamos justificado la tesis de Church-Turing será inmediato que una relación es recursiva —en el sentido que introducimos seguidamente— si y sólo si existe un algoritmo para determinar si se cumple o no sobre unos argumentos dados.

Definición 7.2 Si R es una relación n-ádica, llamaremos *función característica* de R a la función n-ádica dada por

$$\chi_R(a_1, \ldots, a_n) = \begin{cases} 1 & \text{si} \quad R(a_1, \ldots, a_n) \\ 0 & \text{si no } R(a_1, \ldots, a_n) \end{cases}$$

Una relación es *recursiva (primitiva)* si lo es su función característica.

Vamos a usar los conectores lógicos con su significado obvio a la hora de nombrar relaciones. Por ejemplo, si R es una relación, llamaremos $\neg R$ a la relación que se cumple sobre unos números si y sólo si no se cumple R, mientras que $R \wedge S$ será la relación que se cumple si y sólo si se cumplen a la vez R y S, etc. Insistimos en que esto debe verse como un mero convenio para nombrar relaciones que, en principio, no tiene ninguna conexión directa con los signos de ningún lenguaje formal.

Teorema 7.3 *Si R, S son relaciones n-ádicas recursivas (primitivas), también lo son*

$$\neg R, \quad R \wedge S, \quad R \vee S, \quad R \to S, \quad R \leftrightarrow S.$$

Demostración: $\chi_{\neg R}(x) = \overline{\mathrm{Sg}}\, \chi_R(x), \quad \chi_{R \wedge S}(x) = \chi_R(x) \cdot \chi_S(x),$

$$R \vee S = \neg(\neg R \wedge \neg S), \quad R \to S = \neg R \vee S, \quad R \leftrightarrow S = (R \to S) \wedge (S \to S).$$

■

Naturalmente, esto sirve de poco si no partimos de ninguna relación que sepamos que es recursiva:

Teorema 7.4 *Las relaciones $m = n$, $m \leq n$, $m < n$ son recursivas primitivas.*

DEMOSTRACIÓN: $\chi_<(m,n) = \mathrm{Sg}(n \dot{-} m)$, $\quad m \leq n \leftrightarrow \neg(n < m)$,

$$m = n \leftrightarrow m \leq n \wedge n \leq m.$$

■

Claramente entonces, si f, g son funciones n-ádicas recursivas (primitivas), las relaciones

$$f(a_1, \ldots, a_n) = g(a_1, \ldots, a_n), \quad f(a_1, \ldots, a_n) \leq g(a_1, \ldots, a_n),$$

$$f(a_1, \ldots, a_n) < g(a_1, \ldots, a_n),$$

también lo son, pues, por ejemplo,

$$\chi_{f \leq g}(a_1, \ldots, a_n) = \chi_\leq(f(a_1, \ldots, a_n), g(a_1, \ldots, a_n)),$$

es decir, la función característica de $f \leq g$ es la composición de la de $\leq$ con f y g. Lo mismo vale para las otras dos.

Teorema 7.5 *Si R es una relación $n+1$-ádica recursiva (primitiva) también lo son las relaciones*

$$S(k, a_1, \ldots, a_n) \leftrightarrow \bigvee m \leq k \; R(m, a_1, \ldots, a_n),$$
$$T(k, a_1, \ldots, a_n) \leftrightarrow \bigwedge m \leq k \; R(m, a_1, \ldots, a_n).$$

y la función $f(k, a_1, \ldots, a_n) = \mu m \leq k \; R(m, a_1, \ldots, a_n)$.

DEMOSTRACIÓN: Definimos la función

$$\begin{cases} h(0, a_1, \ldots, a_n) = \overline{\mathrm{Sg}}\, \chi_R(0, a_1, \ldots, a_n), \\ h(k+1, a_1, \ldots, a_n) = h(k, a_1, \ldots, a_n) \cdot \overline{\mathrm{Sg}}\, \chi_R(k+1, a_1, \ldots, a_n). \end{cases}$$

Así h es recursiva y $h(k, a_1, \ldots, a_n) = 0 \leftrightarrow \bigvee m \leq k \; R(m, a_1, \ldots, a_n)$. Por lo tanto,

$$\chi_S(k, a_1, \ldots, a_n) = \overline{\mathrm{Sg}}\, h(k, a_1, \ldots, a_n),$$

luego S es recursiva (primitiva), al igual que

$$T(k, a_1, \ldots, a_n) \leftrightarrow \neg \bigvee m \leq k \; \neg R(m, a_1, \ldots, a_n).$$

Por último, f puede definirse recursivamente de este modo:

$$\begin{aligned} f(0, a_1, \ldots, a_n) = \;& 0, \\ f(k+1, a_1, \ldots, a_n) = \;& f(k, a_1, \ldots, a_n) + \\ & (h(k, a_1, \ldots, a_n) \dot{-} h(k+1, a_1, \ldots, a_n))(k+1). \end{aligned}$$

■

7.2 Caracterización aritmética

Vamos a demostrar que las relaciones y funciones recursivas son precisamente las relaciones y funciones aritméticas de tipo Δ_1 (definición 5.28), es decir, que una relación n-ádica R es recursiva si y sólo si existen fórmulas $\phi(x_1,\dots,x_n)$ y $\psi(x_1,\dots,x_n)$ de $\mathcal{L}_a$ de tipo Σ_1 y Π_1 respectivamente tales que

$$R(a_1,\dots,a_n) \quad \text{syss} \quad \mathbb{N}\vDash \phi(0^{(a_1)},\dots,0^{(a_n)}) \quad \text{syss} \quad \mathbb{N}\vDash \psi(0^{(a_1)},\dots,0^{(a_n)}),$$

y una función es recursiva si y sólo si lo es la relación $f(x_1,\dots,x_n)=y$.

Definición 7.6 Diremos que una fórmula $\phi(x_1,\dots,x_m,y)$ de $\mathcal{L}_a$ *define* una función m-ádica F si para todos los números naturales $a_1,\dots,a_n$, se cumple

$$\underset{\mathrm{I}\Sigma_1}{\vdash} \overset{1}{\bigvee} y\, \phi(0^{(a_1)},\dots,0^{(a_n)},y)$$

y además $a=f(a_1,\dots,a_n)$ si y sólo si $\mathbb{N}\vDash \phi(0^{(a_1)},\dots,0^{(a_n)},0^{(a)})$.

Las funciones F definidas de esta forma se llaman *funciones aritméticas.* Diremos que F es de tipo Σ_1 si la función ϕ se puede tomar de tipo Σ_1, pero en tal caso, la fórmula Π_1 dada por

$$\psi(x_1,\dots,x_n,y) \equiv \bigwedge x(\phi(x_1,\dots,x_n,x)\to x=y),$$

también cumple que, para todos los números $a_1,\dots,a_n$,

$$\underset{\mathrm{I}\Sigma_1}{\vdash} \overset{1}{\bigvee} y\, \psi(0^{(a_1)},\dots,0^{(a_n)},y),$$

así como que

$$a=f(a_1,\dots,a_n) \quad \text{syss} \quad \mathbb{N}\vDash \psi(0^{(a_1)},\dots,0^{(a_n)},0^{(a)}),$$

por lo que en este caso es más habitual decir que F es de tipo Δ_1.

En efecto, si suponemos que ϕ define a F, en $\mathrm{I}\Sigma_1$ se prueba que el y que cumple $\phi(0^{(a_1)},\dots,0^{(a_n)},y)$ cumple también $\psi(0^{(a_1)},\dots,0^{(a_n)},y)$, y es el único, pues si un y' cumple también $\psi(0^{(a_1)},\dots,0^{(a_n)},y')$, tenemos que

$$\phi(0^{(a_1)},\dots,0^{(a_n)},y)\to y=y',$$

de donde $y=y'$. Además si a cumple $\mathbb{N}\vDash \psi(0^{(a_1)},\dots,0^{(a_n)},0^{(a)})$, y llamamos $a'=f(a_1,\dots,a_n)$, tenemos que

$$\mathbb{N}\vDash \phi(0^{(a_1)},\dots,0^{(a_n)},0^{(a')})\to 0^{(a')}=0^{(a)},$$

de donde $a=a'=f(a_1,\dots,a_n)$. Recíprocamente, si un número natural a' cumple $\mathbb{N}\vDash \phi(0^{(a_1)},\dots,0^{(a_n)},0^{(a')})$, tiene que ser $a'=a$, luego a cumple la relación $\mathbb{N}\vDash \psi(0^{(a_1)},\dots,0^{(a_n)},0^{(a)})$.

Teorema 7.7 *Toda función recursiva es* Δ_1.

DEMOSTRACIÓN: En primer lugar observamos que las funciones recursivas elementales son Δ_1. En efecto

$$c(n) = m \quad \text{syss} \quad \mathbb{N} \vDash 0^{(m)} = 0,$$

$$s(n) = m \quad \text{syss} \quad \mathbb{N} \vDash 0^{(m)} = 0^{(n)} + 1,$$

$$p_i^k(a_1, \dots, a_k) = a \quad \text{syss} \quad \mathbb{N} \vDash 0^{(a_i)} = 0^{(a)},$$

luego basta considerar las fórmulas

$$\phi_c(x, y) \equiv y = 0, \quad \phi_s(x, y) \equiv y = x + 1, \quad \phi_{p_i^k}(x_1, \dots, x_k, y) \equiv y = x_i,$$

pues claramente en $\mathrm{I}\Sigma_1$ se demuestra la condición de unicidad requerida.

Basta probar que toda función definida por composición, recursión o minimización a partir de funciones Δ_1 es Δ_1.

Supongamos en primer lugar que F es la composición de la función m-ádica H y de las funciones n-ádicas $G_1, \dots, G_m$, definidas por fórmulas Σ_1, digamos $\phi, \psi_1, \dots, \psi_m$, y consideramos la fórmula

$$\chi(x_1, \dots, x_n, y) \equiv \bigvee y_1 \dots y_m (\psi_1(x_1, \dots, x_n, y_1) \wedge \dots \wedge \psi_m(x_1, \dots, x_m, y_m)$$

$$\wedge \phi(y_1, \dots, y_m, y)),$$

que claramente es de tipo Σ_1. Fijados $a_1, \dots, a_n$, llamamos $b_i = G_i(a_1, \dots, a_n)$ y $c = H(b_1, \dots, b_m) = F(a_1, \dots, a_n)$. Tenemos que

$$\underset{\mathrm{I}\Sigma_1}{\vdash} \overset{1}{\bigvee} y_i \, \psi_i(0^{(a_1)}, \dots, 0^{(a_n)}, y_i), \qquad \overset{1}{\bigvee} y \, \phi(0^{(b_1)}, \dots, 0^{(b_m)}, y).$$

Puesto que

$$\mathbb{N} \vDash \psi_i(0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(b_i)}), \qquad \mathbb{N} \vDash \phi(0^{(b_1)}, \dots, 0^{(b_m)}, 0^{(c)}),$$

el teorema de Σ_1-completitud 5.29 implica que

$$\underset{\mathrm{I}\Sigma_1}{\vdash} \psi_i(0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(b_i)}), \qquad \underset{\mathrm{I}\Sigma_1}{\vdash} \phi(0^{(b_1)}, \dots, 0^{(b_m)}, 0^{(c)}),$$

de donde $\underset{\mathrm{I}\Sigma_1}{\vdash} \chi(0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(c)})$ y así $\underset{\mathrm{I}\Sigma_1}{\vdash} \bigvee y \, \chi(0^{(a_1)}, \dots, 0^{(a_n)}, y)$.

Si, razonando en $\mathrm{I}\Sigma_1$, suponemos $\chi(0^{(a_1)}, \dots, 0^{(a_n)}, y)$, tenemos que existen $y_1, \dots, y_m$ tales que $\psi_i(0^{(a_1)}, \dots, 0^{(a_n)}, y_i)$, y la unicidad implica que $y = 0^{(b_i)}$, luego $\phi(0^{(b_1)}, \dots, 0^{(b_m)}, y)$ y, de nuevo por la unicidad, $y = 0^{(c)}$, lo que prueba que

$$\underset{\mathrm{I}\Sigma_1}{\vdash} \overset{1}{\bigvee} y \, \chi(0^{(a_1)}, \dots, 0^{(a_n)}, y).$$

Más aún, hemos probado que el único y que cumple esto es $0^{(c)}$, lo que implica que $F(a_1, \dots, a_n) = c$ equivale a $\mathbb{N} \vDash \chi(0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(c)})$.

Supongamos ahora que F está definida por recursión a partir de la función n-ádica G y de la función $n+2$-ádica H, las cuales están definidas por las fórmulas ϕ y χ de tipo Σ_1, y consideramos la fórmula[1]

$$\chi(x_1,\dots,x_n,x,y) \equiv \bigvee s(\ell(s) = x+1 \wedge s_x = y \wedge \phi(x_1,\dots,x_n,s_0)$$
$$\wedge \bigwedge i < x\ \psi(x_1,\dots,x_n,i,s_i,s_{i+1})).$$

Fijamos números naturales $a_1,\dots,a_n,a$ y llamamos s al número natural que se corresponde con la sucesión de longitud $a+1$ dada por $s_i = F(a_1,\dots,a_n,i)$. En particular $s_a = F(a_1,\dots,a_n,a)$. Vamos a abreviar $\bar{s} \equiv 0^{(s)}$. Puesto que $\mathbb{N} \vDash \ell(\bar{s}) = 0^{(a)}+1$ y la fórmula es Σ_1, el teorema de Σ_1-completitud nos da que en $\mathrm{I}\Sigma_1$ se demuestra $\ell(\bar{s}) = 0^{(a)}+1$, e igualmente $\bar{s}_{0^{(a)}} = 0^{(s_a)}$.

Por otro lado, tenemos que $s_0 = F(a_1,\dots,a_n,0) = G(a_1,\dots,a_n)$ y

$$s_{i+1} = F(a_1,\dots,a_n,i,s_i) = H(a_1,\dots,a_n,i,s_i).$$

Como $\mathbb{N} \vDash \bar{s}_{0^{(i)}} = 0^{(s_i)}$ y la fórmula es Δ_1, por Σ_1-completitud tenemos que en $\mathrm{I}\Sigma_1$ se demuestra que $\bar{s}_{0^{(i)}} = 0^{(s_i)}$, luego las igualdades anteriores se traduce en que

$$\mathbb{N} \vDash \phi(0^{(a_1)},\dots,0^{(a_n)},\bar{s}_0), \qquad \mathbb{N} \vDash \psi(0^{(a_1)},\dots,0^{(a_n)},0^{(i)},\bar{s}_{0^{(i)}},\bar{s}_{0^{(i+1)}})$$

y, de nuevo por Σ_1-completitud, en $\mathrm{I}\Sigma_1$ se demuestra

$$\phi(0^{(a_1)},\dots,0^{(a_n)},\bar{s}_0), \qquad \psi(0^{(a_1)},\dots,0^{(a_n)},0^{(i)},\bar{s}_{0^{(i)}},\bar{s}_{0^{(i+1)}})$$

Usando el teorema 5.4, las últimas fórmulas implican

$$\bigwedge i < 0^{(a)}\ \psi(0^{(a_1)},\dots,0^{(a_n)},i,\bar{s}_i,\bar{s}_{i+1}).$$

De aquí se sigue $\chi(0^{(a_1)},\dots,0^{(a_n)},0^{(a)},0^{(s_a)})$, luego en particular

$$\bigvee y\, \chi(0^{(a_1)},\dots,0^{(a_n)},0^{(a)},y).$$

Si suponemos $\chi(0^{(a_1)},\dots,0^{(a_n)},0^{(a)},y)$, tomamos s^* que cumpla la definición, de modo que $\ell(s^*) = 0^{(a)}+1$. Además $\phi(0^{(a_1)},\dots,0^{(a_n)},s^*_0)$, y por la unicidad de ϕ tiene que ser $s^*_0 = \bar{s}_0$. Supuesto que $s^*_{0^{(i)}} = \bar{s}_{0^{(i)}}$, como tenemos

$$\psi(0^{(a_1)},\dots,0^{(a_n)},0^{(i)},s^*_{0^{(i)}},s^*_{0^{(i+1)}}),$$

la unicidad de ψ implica que $s^*_{i+1} = \bar{s}_{i+1}$, y así concluimos que $s^* = \bar{s}$ y, en particular, que $y = s^*_{0^{(a)}} = \bar{s}_{0^{(a)}} = 0^{(s_a)}$.

Así hemos demostrado que el único y que cumple $\chi(0^{(a_1)},\dots,0^{(a_n)},0^{(a)},y)$ es $y = 0^{(s_a)}$, lo que, por una parte, implica $\overset{1}{\bigvee} y\, \chi(0^{(a_1)},\dots,0^{(a_n)},0^{(a)},y)$ y, por otra, que $c = F(a_1,\dots,a_n,a)$ equivale a $\mathbb{N} \vDash \chi(0^{(a_1)},\dots,0^{(a_n)},0^{(a)},0^{(c)})$.

[1]Identificamos los números naturales con sucesiones finitas de números naturales según hemos visto en la sección 5.6.

Dejamos a cargo del lector la adaptación mínima del argumento que se requiere cuando la función F es monádica.

Supongamos, por último, que F se define por minimización a partir de la función G, determinada por la fórmula ϕ, de tipo Σ_1. Sea $\psi(x_1,\dots,x_n,y)$ una fórmula de tipo Σ_1 equivalente en $\mathrm{I}\Sigma_1$ a

$$\phi(x_1,\dots,x_n,y,0) \wedge \bigwedge i < y \bigvee x\ \phi(x_1,\dots,x_n,i,x+1),$$

Fijamos $a_1,\dots,a_n$ y llamamos $a = F(a_1,\dots,a_n)$.

Como $G(a_1,\dots,a_n,a) = 0$, tenemos que $\mathbb{N} \vDash \phi(0^{(a_1)},\dots,0^{(a_n)},0^{(a)},0)$ y, por Σ_1-completitud, en $\mathrm{I}\Sigma_1$ se demuestra $\phi(0^{(a_1)},\dots,0^{(a_n)},0^{(a)},0)$. Usando el teorema 5.4 es fácil probar también que

$$\bigwedge i < 0^{(a)} \bigvee y\ \phi(0^{(a_1)},\dots,0^{(a_n)},i,y+1),$$

con lo que $\psi(0^{(a_1)},\dots,0^{(a_n)},0^{(a)})$, y en particular $\bigvee y, \psi(0^{(a_1)},\dots,0^{(a_n)},y)$.

Por otro lado, si se cumple $\psi(0^{(a_1)},\dots,0^{(a_n)},y)$, no puede ser $0^{(a)} < y$, ya que entonces tendríamos existiría un x tal que $\phi(0^{(a_1)},\dots,0^{(a_n)},0^{(a)},x+1)$, pero, por la unicidad de ϕ, sería $x+1=0$, lo cual es imposible.

Tampoco puede ser $y < 0^{(a)}$, porque y cumple $\phi(0^{(a_1)},\dots,0^{(a_n)},y,0)$, pero el teorema 5.4 nos daría que

$$y = 0^{(0)} \vee y = 0^{(1)} \vee \cdots \vee y = 0^{(a-1)}$$

y, por otra parte, como $F(a_1,\dots,a_n,i) \neq 0$ para todo $i < a$, podemos demostrar que $\neg\phi(0^{(a_1)},\dots,0^{(a_n)},0^{(i)},0)$, luego también $\neg\phi(0^{(a_1)},\dots,0^{(a_n)},y,0)$, con lo que tenemos una contradicción. Esto implica que $y = 0^{(a)}$, luego concluimos que $\overset{1}{\bigvee} y\, \psi(0^{(a_1)},\dots,0^{(a_n)},y)$ y, concretamente, el único y que cumple esto es $0^{(a)}$. Esto implica que $a = F(a_1,\dots,a_n)$ equivale a $\mathbb{N} \vDash \psi(0^{(a_1)},\dots,0^{(a_n)},0^{(a)})$.

■

Vamos a dar un teorema más fuerte de definición por recursión. Para ello, si F es una función $n+1$-ádica de tipo Δ_1, definimos $\bar{F}|_n(a_1,\dots,a_n)$ como el número natural s que cumple $\ell(s) = n$ y, para $i < n$, se cumple $s_i = F(a_1,\dots,a_n,i)$.

Teorema 7.8 (Recursión completa) *Si H es una función $n+2$-ádica de tipo Δ_1, la función $n+1$-ádica F dada por*

$$F(a_1,\dots,a_n,a) = H(a_1,\dots,a_n,a,\bar{F}|_a)$$

es de tipo Δ_1.

Demostración: La prueba es similar al caso de la prueba del teorema anterior correspondiente al caso en que F está definida por recursión. Basta considerar

$$\chi(x_1,\dots,x_n,x,y) \equiv \bigvee s(\ell(s) = x+1 \wedge s_x = y \wedge \bigwedge i \leq x\ s_i = \psi(x_1,\dots,x_n,s|_i)).$$

■

Teorema 7.9 *Toda relación Δ_0 es recursiva primitiva.*

DEMOSTRACIÓN: Observemos en primer lugar que si $t(x_1, \ldots, x_n)$ es un término sin descriptores, entonces la fórmula $y = t(x_1, \ldots, s_n)$ (donde y es una variable que no esté libre en t) define una función recursiva primitiva.

En efecto, si $t \equiv x_i$ entonces $y = x_i$ define la proyección p_i^n, que es recursiva elemental. Si $t \equiv 0$, entonces $y = 0$ define la composición de una proyección cualquiera con la función c, luego es recursiva primitiva.

Si $t \equiv t_0'$ y la fórmula $y = t_0'$ define una función recursiva primitiva, entonces $y = t$ define la composición de dicha función con la función s, luego es también recursiva primitiva.

Si $t \equiv t_1 + t_2$ o $t \equiv t_1 \cdot t_2$ y t_1, t_2 definen funciones recursivas primitivas, entonces $y = t$ define la composición de estas dos funciones con la función suma o producto (y ambas son recursivas primitivas), luego también es recursiva primitiva.

Ahora observamos que si $t_1(x_1, \ldots, x_n)$ y $t_2(x_1, \ldots, x_n)$ son términos sin descriptores, entonces las fórmulas $t_1 = t_2$ y $t_1 \leq t_2$ definen relaciones recursivas. En efecto, basta tener en cuenta la observación tras el teorema 7.4, ya que si f_1 y f_2 son las funciones recursivas primitivas definidas por $y = t_1$ e $y = t_2$, respectivamente, es decir, que

$$a = f_i(a_1, \ldots, a_n) \quad \text{syss} \quad \mathbb{N} \vDash 0^{(a)} = t_i(0^{(a_1)}, \ldots, 0^{(a_n)}),$$

entonces la relación definida por $t_1 = t_2$ es $f_1(a_1, \ldots, a_n) = f_2(a_1, \ldots, a_n)$, que es una relación recursiva primitiva, e igualmente con $\leq$.

Ahora probamos que si ϕ es una fórmula Δ_0, entonces la relación aritmética que define es recursiva primitiva. Ya lo tenemos probado para fórmulas de tipo $t_1 = t_2$ o $t_1 \leq t_2$. Si unas fórmulas ϕ y ψ definen relaciones recursivas primitivas R y S, entonces las fórmulas $\neg\phi$ y $\phi \to \psi$ definen las relaciones $\neg R$ y $R \to S$, que son recursivas primitivas por 7.3. Por último, una fórmula de tipo $\bigwedge x \leq y\, \phi(x, x_1, \ldots, x_n)$ o $\bigwedge x \leq y\, \phi(x, x_1, \ldots, x_n)$ define una relación recursiva primitiva si así lo hace la fórmula ϕ por el teorema 7.5. ∎

Teorema 7.10 *Una función es recursiva si y sólo si es Δ_1.*

DEMOSTRACIÓN: Ya hemos probado una implicación. Si F es una función n-ádica Δ_1, entonces existe una fórmula ϕ de tipo Δ_0 tal que

$$F(a_1, \ldots, a_n) = a \quad \text{syss} \quad \mathbb{N} \vDash \bigvee x\, \phi(x, 0^{(a_1)}, \ldots, 0^{(a_n)}, 0^{(a)})$$

$$\text{syss existe un } b \text{ tal que } \mathbb{N} \vDash \phi(0^{(b)}, 0^{(a_1)}, \ldots, 0^{(a_n)}, 0^{(a)}).$$

Por el teorema anterior, la relación

$$R(a_1, \ldots, a_n, a, b) \quad \text{syss} \quad \mathbb{N} \vDash \phi(0^{(b)}, 0^{(a_1)}, \ldots, 0^{(a_n)}, 0^{(a)})$$

es recursiva primitiva, y tenemos que

$$F(a_1,\dots,a_n)=a \quad \text{syss existe un } b \text{ tal que } R(a_1,\dots,a_n,a,b).$$

Ahora observamos que la relación

$$z=(x,y)=\frac{(x+y)(x+y+1)}{2}+x$$

es recursiva primitiva, pues está determinada por la fórmula $z=\langle x,y\rangle$ definida[2] en 5.14, que es Δ_0. Lo mismo vale para las relaciones

$$x=z_0 \quad \text{syss} \quad \bigvee y\le z\ z=(x,y), \qquad x=z_1 \quad \text{syss} \quad \bigvee x\le z\ z=(x,y).$$

Más aún, la función $H(z)=z_0$ es recursiva primitiva, pues si $R_0(z,x)$ es la relación $x=z_0$, entonces $H(z)=\mu x\le z\ \chi_{\neg R_0}(z,x)=0$.

Por lo tanto, la relación

$$S(a_1,\dots,a_n,c) \quad \text{syss} \quad \bigvee xy\le c(c=(x,y)\wedge R(a_1,\dots,a_n,x,y))$$

es recursiva primitiva, por los teoremas 7.3 y 7.5. Además, dados $a_1,\dots,a_n$, existe siempre un c que cumple $S(a_1,\dots,a_n,c)$, pues basta tomar $c=(a,b)$, donde $a=F(a_1,\dots,a_n)$ y b es el adecuado para que se cumpla ϕ. Recíprocamente, si se cumple $S(a_1,\dots,a_n,c)$, entonces se cumple $R(a_1,\dots,a_n,c_0,c_1)$, luego $c_0=f(a_1,\dots,a_n)$. Podemos definir

$$G(a_1,\dots,a_n)=\mu c\,\chi_{\neg S}(a_1,\dots,a_n,c)=0,$$

que es recursiva, pues siempre existe un c tal que $\chi_{\neg S}(a_1,\dots,a_n,c)=0$ (ya que esto equivale a $S(a_1,\dots,a_n,c)$). Ahora es claro que

$$F(a_1,\dots,a_n)=H(G(a_1,\dots,a_n)),$$

luego F es recursiva. ∎

Teorema 7.11 *Una relación es recursiva si y sólo si es Δ_1.*

Demostración: Si R es una relación n-ádica recursiva, entonces χ_R es una función recursiva, luego es Δ_1. Esto significa que existe una fórmula $\phi(x_1,\dots,x_n,x)$ de tipo Σ_1 tal que

$$\chi_R(a_1,\dots,a_n)=a \quad \text{syss} \quad \mathbb{N}\vDash\phi(0^{(a_1)},\dots,0^{(a_n)},0^{(a)}).$$

Entonces

$$R(a_1,\dots,a_n) \quad \text{syss} \quad \chi_R(a_1,\dots,a_n)=1 \quad \text{syss} \quad \mathbb{N}\vDash\phi(0^{(a_1)},\dots,0^{(a_n)},1)$$

$$\text{syss} \quad \mathbb{N}\vDash\neg\phi(0^{(a_1)},\dots,0^{(a_n)},0),$$

luego R está descrita tanto por una fórmula Σ_1 como por una fórmula Π_1 y es, por consiguiente, Δ_1.

[2]Notemos que estamos usando la misma notación para representar la relación metamatemática que la fórmula que la expresa en $\mathcal{L}_a$.

Supongamos ahora que R es Δ_1. Esto significa que existen fórmulas ϕ y ψ de tipo Σ_1 tales que

$$R(a_1,\dots,a_n) \quad \text{syss} \quad \mathbb{N} \vDash \phi(0^{(a_1)},\dots,0^{(a_n)}) \quad \text{syss} \quad \mathbb{N} \vDash \neg\psi(0^{(a_1)},\dots,0^{(a_n)}).$$

Entonces la fórmula

$$\chi(x_1,\dots,x_n,x) \equiv (\phi(x_1,\dots,x_n) \wedge x = 1) \vee (\psi(x_1,\dots,x_n) \wedge x = 0)$$

es Σ_1 y define a la función característica de R, luego ésta es recursiva y R también. ∎

Ahora es inmediato el carácter recursivo de todos los conceptos que hemos definido aritméticamente (pues todos tienen definiciones Δ_1), por ejemplo, las funciones $\ell(n)$ o n_i que dan la longitud o el elemento i-ésimo de la sucesión codificada por n.

7.3 Funciones recursivas parciales

Según hemos anunciado (y demostraremos después) las funciones recursivas son aquellas que pueden calcularse mediante un algoritmo. Sin embargo, el hecho de que una función $k+1$-ádica g pueda calcularse mediante un algoritmo nos permite ir calculando sucesivamente

$$g(a_1,\dots,a_k,0), \quad g(a_1,\dots,a_k,1), \quad g(a_1,\dots,a_k,2), \quad \dots$$

pero no nos garantiza que vaya a existir un número natural n que cumpla la condición $g(a_1,\dots,a_k,n) = 0$. Por consiguiente, el hecho de que la función g sea recursiva no nos asegura que la función k-ádica f definida por minimización a partir de g sea necesariamente recursiva. La definición exige que sea cierto que para todos los $a_1,\dots,a_k$ exista un n que anule a g, pero una cosa es que esto suceda y otra muy distinta que seamos capaces de saber que sucede. En principio, podría darse el caso de que la función f fuera recursiva y no existiera un argumento que lo justificara.

Este aspecto "no computable" de las funciones recursivas puede eliminarse pasando a una clase mayor de funciones, a las que se permite quedar indefinidas en ocasiones. Vamos a precisar esta idea:

Definición 7.12 Una *función n-ádica parcial* f es un criterio que a ciertos grupos de n números naturales $a_1,\dots,a_n$, repetidos o no y en un cierto orden, les asigna un número natural que representaremos por $f(a_1,\dots,a_n)$. En tal caso se dice que f *está definida* para $a_1,\dots,a_n$ o que $f(a_1,\dots,a_n)$ *está definido*, pero también se admite que f no esté definida para algunos argumentos posibles.

Una función parcial k-ádica está definida por *composición parcial* a partir de las funciones parciales g (r-ádica) y $h_1,\dots,h_r$ (k-ádicas) si f está definida exactamente para aquellos naturales $a_1,\dots,a_k$ tales que están definidas $h_i(a_1,\dots,a_k)$ $(i = 1,\dots,r)$ y $g(h_1(a_1,\dots,a_k),\dots,h_r(a_1,\dots,a_k))$ y se cumple que

$$f(a_1,\dots,a_k) = g(h_1(a_1,\dots,a_k),\dots,h_r(a_1,\dots,a_k)).$$

Una función parcial $k+1$-ádica está definida por *recursión parcial* a partir de la función parcial k-ádica g (o del número natural a si $k = 0$) y la función parcial $k+2$-ádica h si f está definida exactamente para aquellos naturales $a_1, \ldots, a_k$, n tales que

1. $g(a_1, \ldots, a_k)$ está definido [si $k \neq 0$],
2. $f(a_1, \ldots, a_k, u)$ está definido para todo $u < n$,
3. $h(a_1, \ldots, a_k, u, f(u, a_1, \ldots, a_k))$ está definido para todo $u < n$,

y se cumple

$$\begin{aligned} f(a_1, \ldots, a_k, 0) &= g(a_1, \ldots, a_k) \\ f(a_1, \ldots, a_k, u+1) &= h(a_1, \ldots, a_k, u, f(u, a_1, \ldots, a_k)) \quad \text{si } 0 \leq u < n. \end{aligned}$$

Una función parcial k-ádica f está definida por *minimización parcial* a partir de una función parcial $k+1$-ádica g si f está definida exactamente para aquellos naturales $a_1, \ldots, a_k$ tales que existe un natural n que cumple

1. Si $m \leq n$ entonces g está definida para $a_1, \ldots, a_k, m$.
2. $g(a_1, \ldots, a_k, n) = 0$ y se cumple

$$f(a_1, \ldots, a_k) = \mu n\, g(a_1, \ldots, a_k, n) = 0.$$

Una función parcial f es *recursiva parcial* si hay una sucesión de funciones $f_1, \ldots, f_n$ tales que f_n es f y cada f_i es recursiva elemental o está definida por composición, recursión o minimización parcial a partir de funciones anteriores de la sucesión.

Obviamente, toda función recursiva es recursiva parcial. Ahora la condición de minimización parcial no incluye ninguna condición que no sepamos comprobar y por consiguiente nunca hay duda de si la construcción de una función se ajusta o no a la definición de función recursiva parcial. La duda "se traslada" ahora a saber si una función parcial dada está o no definida para unos argumentos dados. En efecto, si una función f está definida por minimización parcial a partir de una función g, podemos encontrarnos con que al intentar calcular

$$g(a_1, \ldots, a_k, 0), \quad g(a_1, \ldots, a_k, 1), \quad g(a_1, \ldots, a_k, 2), \quad \ldots$$

el proceso continúe indefinidamente sin que nunca se obtenga el valor 0, o bien que el cálculo de algún $g(a_1, \ldots, a_k, n)$ no termine nunca (porque a su vez requiera aplicar una definición por minimización parcial). En ambos casos $f(a_1, \ldots, a_k)$ quedará indefinido.

Este concepto de función parcial se ajusta mejor al concepto de "función computable mediante un ordenador", pues un algoritmo programado en un ordenador puede perfectamente caer en un bucle infinito y no proporcionar respuesta alguna a una entrada determinada.

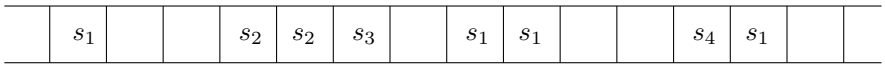
7.4 Máquinas de Turing

El argumento de Turing para probar que las funciones recursivas coinciden con las funciones calculables mediante un algoritmo se basa en el concepto de máquina de Turing. Una máquina de Turing es un modelo teórico de ordenador con infinita memoria disponible. Aunque la descripción que sigue parezca la descripción de una máquina real, debemos tener presente que una máquina de Turing es un concepto abstracto de la misma naturaleza que una teoría axiomática, es decir, un sistema conceptual que fundamenta una serie de afirmaciones objetivas.

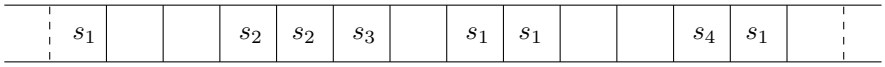
Una *máquina de Turing* consta de una cinta infinita dividida en infinitas casillas contiguas infinitamente prolongable tanto a izquierda como a derecha:

Cada casilla puede estar en blanco o tener impreso un signo de entre los de una lista finita que llamaremos *alfabeto:* $s_1, \ldots, s_j$ $(j \geq 1)$ fija para cada máquina particular. Escribiremos s_0 para nombrar al "blanco" y así la situación posible de una casilla será una de entre $s_0, \ldots, s_j$. En cualquier momento la cinta tendrá un número finito de casillas impresas (con signos distintos de s_0).

Representaremos la cinta con signos así:



Se sobrentiende que el resto de la cinta está en blanco. Si queremos indicar que el resto de la cinta puede no estar en blanco pero que no importa lo que haya impreso, lo representaremos así:



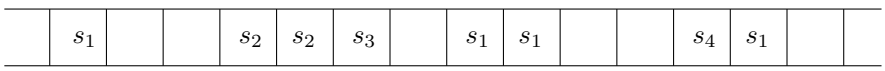
El estado de la cinta en un momento dado lo llamaremos *situación.*

En cada instante la máquina se encontrará en un *estado* de entre un número finito posible de ellos $q_0, \ldots, q_k$ $(k \geq 1)$ fijo para cada máquina particular. Cada estado puede ser *activo* o *pasivo.* El estado q_0 siempre es pasivo, q_1 siempre es activo. A q_1 le llamaremos *estado inicial,* es el estado en que se encuentra la máquina cuando empieza a funcionar.

En cada instante la máquina lee el signo de una casilla de la cinta. Esta casilla se llama *casilla escrutada* y el signo se llama *signo escrutado.*

El estado de la máquina y el signo escrutado determinan la *configuración* de la máquina en un instante dado. La configuración y la situación determinan la *configuración completa.*

Expresaremos las configuraciones completas indicando el estado sobre la casilla escrutada así:



Si no queremos indicar el estado usaremos un guión "—".

Si en un instante dado una máquina de Turing se encuentra en un estado activo, ésta realizará un *acto*. Un acto consiste en:

1. Leer el signo de la casilla escrutada,
2. Imprimir un signo (quizá s_0) en la casilla escrutada,
3. Mover un lugar la cinta de modo que la nueva casilla escrutada pase a ser la contigua izquierda, la misma casilla o la contigua derecha,
4. Cambiar de estado (pasando quizá al mismo),

de tal modo que el signo que se imprime, el movimiento que se hace y el estado al que se pasa, son función exclusivamente de la configuración de la máquina en ese instante.

Si el estado es pasivo no se produce ningún acto: la máquina está *parada.*

Según esto una máquina de Turing viene determinada por:

1. El alfabeto $s_0, \ldots, s_j$, con $j \geq 1$,
2. El conjunto de estados posibles $q_0, \ldots, q_k$, con $k \geq 1$,
3. Una función que a cada configuración activa (s_a, q_b) le asigna una terna (s_c, M, q_d), donde s_c, M, q_d son, respectivamente el signo impreso, el movimiento realizado I, D o C (izquierda, derecha o centro) y el estado al que se pasa cuando la configuración es (s_a, q_b). A esta función se le llama *programa* de la máquina.

En la práctica escribiremos el programa en forma de tabla. Por ejemplo: Sea A la máquina de Turing con alfabeto s_0, s_1, estados q_0, q_1 y programa

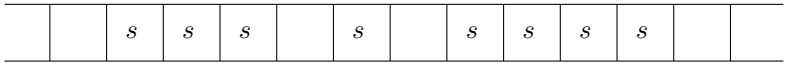
A	s_0	s_1
q_1	$s_1 C q_0$	$s_1 D q_1$

La máquina A se mueve sobre la cinta hacia la derecha hasta encontrar una casilla en blanco, donde imprime s_1 y se para.

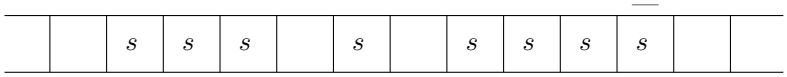
Según advertíamos al principio, las máquinas de Turing no existen (como objetos físicos). No son ordenadores porque ningún ordenador puede trabajar con una "cinta" de memoria infinita. Son un modelo de ordenador ideal exento de limitaciones de memoria. Lo único importante es que podemos hablar consistentemente de ellas y determinar qué hace una máquina dada a partir de unos datos dados, como acabamos de hacer con la máquina A.

Computabilidad Consideremos una máquina de Turing y sea $s = s_1$. Llamaremos *representación* del número natural n a la situación de la cinta que consta de $n + 1$ signos s consecutivos, con el anterior y posterior en blanco.

Llamaremos *representación* de los números $a_1, \ldots, a_n$ a la situación que consta de n secuencias de $a_i + 1$ signos s consecutivos cada una, separadas por un blanco. Por ejemplo, la representación de 2, 0, 3 es



Llamaremos *vacío* en la cinta a dos o más casillas en blanco consecutivas. Llamaremos *representación normal* o *posición normal* de los naturales $a_1, \ldots, a_n$ a la representación de $a_1, \ldots, a_n$ cuando la casilla escrutada es la última casilla impresa de a_n. Por ejemplo 2, 0, 3 en posición normal es

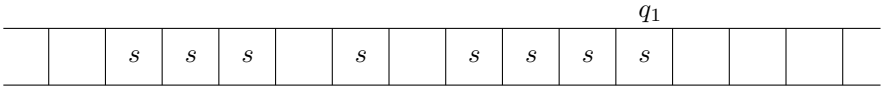


Diremos que una máquina de Turing M *computa* la función parcial n-ádica f si cuando M comienza con los números $a_1, \ldots, a_n$ en posición normal y el resto de la cinta en blanco, termina con la representación normal de

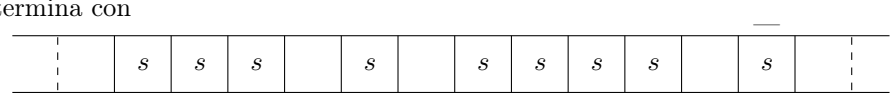
$$a_1, \ldots, a_n, f(a_1, \ldots, a_n)$$

en el caso de que $f(a_1, \ldots, a_n)$ esté definido y no se detiene con $a_1, \ldots, a_n, a$ en posición normal para ningún número a si $f(a_1, \ldots, a_n)$ no está definido.

Por ejemplo si $f(2, 0, 3) = 0$ y M computa f, cuando M comienza con



termina con



No se exige que la posición absoluta de los números en la cinta sea la misma que al comienzo.

Una función parcial es *computable* si hay una máquina de Turing que la computa. Una máquina de Turing M computa $1 \mid 1$ la función parcial n-ádica f si cumple:

1. El alfabeto de M es s_0, s,
2. Si M comienza con $a_1, \ldots, a_n$ en posición normal y el resto de la cinta a la derecha en blanco se cumple:
   (a) Las casillas a la izquierda de la representación de $a_1, \ldots, a_n$ (o sea, a la izquierda del blanco anterior a a_1) no son nunca escrutadas.
   (b) Si $f(a_1, \ldots, a_n)$ está definido, entonces M acaba con
   $$a_1, \ldots, a_n, f(a_1, \ldots, a_n)$$
   en posición normal de modo que la representación comienza en la misma casilla donde comenzaba la de $a_1, \ldots, a_n$ al principio. Además todas las casillas a la derecha quedan en blanco.
   (c) Si $f(a_1, \ldots, a_n)$ no está definido entonces M no se para.

Una función parcial es $1 \mid 1$ *computable* si hay una máquina de Turing que la computa $1 \mid 1$. Vamos a demostrar que una función es computable si y sólo si es $1 \mid 1$ computable si y sólo si es recursiva parcial. El concepto de computabilidad $1 \mid 1$ es un concepto auxiliar técnico para la prueba.

Por el momento trabajaremos con máquinas de un solo signo. Para ellas usaremos la siguiente notación más cómoda:

1. Llamaremos 0 a s_0 y 1 a s_1.
2. Imprimir 1 sobre un 0 lo representaremos E (escribir).
3. Imprimir 0 sobre un 1 lo representaremos B (borrar).
4. Si un signo no se modifica no indicaremos nada.
5. Los estados pasivos serán $0_1, \ldots, 0_n$ (o "0" si sólo hay uno).
6. Los estados activos serán 1, 2, 3, ... (1 es el estado inicial).

Por ejemplo la máquina A de antes se representa ahora así:

A	0	1
1	$E0$	$D1$

Concatenación de máquinas de Turing Si M es una máquina de Turing con estados pasivos $0_1, \ldots, 0_n$ y $N_1, \ldots, N_n$ son otras máquinas de Turing, llamaremos

$$M\left[\begin{array}{c} N_1 \\ \vdots \\ N_n \end{array}\right.$$

a la máquina de Turing definida como sigue:

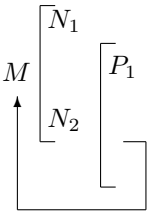
Si $q_1, \ldots, q_j$ son los estados activos de M y $q_1^i, \ldots q_{j_i}^i$ son los estados activos de N_i, los estados activos de la nueva máquina son $q_1, \ldots, q_j, q_1^i, \ldots q_{j_i}^i$, para $i = 1, \ldots, n$.

Los estados pasivos son los de las máquinas $N_1, \ldots, N_n$. El estado inicial es q_1, es decir, el estado inicial de M. El programa es como sigue:

Dada una configuración, se realiza el acto marcado por el programa de la máquina a la que pertenece el estado en curso, a excepción del caso en que M deba pasar al estado 0_i, en cuyo caso se pasa al estado q_1^i.

En otras palabras, se trata de la máquina que empieza actuando como M y, cuando ésta se ha de parar por pasar al estado 0_i, en lugar de ello comienza a actuar la máquina N_i.

La concatenación puede repetirse cuantas veces se quiera, incluso de forma circular. Por ejemplo, la máquina de la figura empieza actuando como M, cuando ésta acaba empieza N_1 o N_2, según el estado pasivo de M al que se llegue; si empieza N_2, cuando ésta acaba empieza P_1 o vuelve a empezar M según el estado pasivo final.



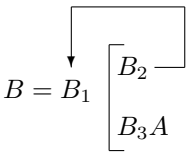
Si M es una máquina de Turing con un único estado pasivo, llamaremos M^n a la máquina que resulta de concatenar M consigo misma n veces.

Construcción de máquinas de Turing Construimos ahora algunas máquinas de Turing concretas e indicamos la actividad que realizan (bajo determinadas condiciones iniciales). Un guión en la tabla del programa indica que no importa la instrucción que pongamos en esa casilla, pues no afecta al comportamiento que se requiere de la máquina.

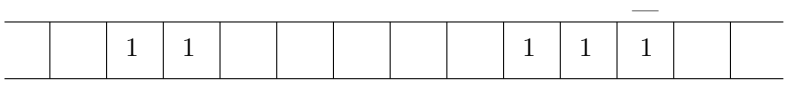
B_1	0	1
1	$-$	$BI2$
2	$EI3$	$I2$
3	$D0_1$	$D0_2$

B_2	0	1
1	$I0$	$D1$

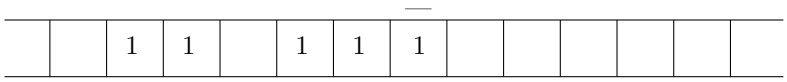
B_3	0	1
1	$-$	$BD0$



Si B comienza con un número en posición normal y otro a su izquierda, mueve el primero hasta eliminar el vacío que los separa (si hay tal vacío) sin escrutar las casillas a la izquierda del segundo número. Por ejemplo, partiendo de



B termina así:



sin escrutar ninguna casilla no representada.

C	0	1
1	$-$	$D2$
2	$D3$	$-$
3	$E0$	$-$

D	0	1
1	$I2$	$I1$
2	$I2$	0

E	0	1
1	$-$	$I2$
2	$D0_1$	$D0_2$

F	0	1
1	$-$	$BI0$

G	0	1
1	$-$	$D2$
2	$D2$	$D3$
3	$I0$	$D3$

H	0	1
1	$-$	$D2$
2	$ED2$	$I3$
3	$-$	$BI0$

El comportamiento de estas máquinas es el siguiente:

C Cuando empieza con un número en posición normal va dos lugares a la derecha e imprime.

D Cuando empieza con un número en posición normal que no sea el extremo izquierdo de la cinta, se sitúa en posición normal respecto al número siguiente por la izquierda.

E Cuando empieza con un número en posición normal toma la salida 0_1 o 0_2 según sea 0 o distinto de 0 y termina en posición normal.

F Cuando comienza en una casilla impresa, borra y va una casilla a la izquierda.

G Va un número a la derecha (al revés que D).

H Cuando comienza con un número en posición normal que no sea el extremo derecho de la cinta, lo completa con unos hasta eliminar el vacío (si existe) que lo separa del siguiente número por la derecha y termina con el número completado en posición normal.

$$I_m = CD^m\ E \begin{cases} HG^m \\ FG^mA \end{cases}$$

La máquina I_m, cuando comienza con $a_1, \ldots, a_m$ en posición normal y con las a_1+2 casillas siguientes por la derecha en blanco, termina con $a_1, \ldots, a_m, a_1$, en posición normal.

$$K_m = AI_m^m FD^m FG^m.$$

La máquina K_m, cuando comienza con $a_1, \ldots, a_m$ en posición normal y las $a_1 + \cdots + a_m + 2m + 1$ casillas siguientes por la derecha en blanco, termina imprimiendo $a_1, \ldots, a_m, , a_1, \ldots, a_m$, donde la doble coma ",," indica un vacío de dos blancos y donde el segundo a_m está en posición normal.

L	0	1
1	$I2$	$I1$
2	$I3$	$BI2$
3	$D4$	$BI2$
4	$D4$	$D5$
5	$I0$	$D5$

Si L comienza con un número en posición normal, borra todos los anteriores a él hasta el primer vacío y vuelve a la posición inicial.

Notemos que ninguna de las máquinas que hemos definido escruta las casillas a la izquierda de los datos.

7.5 La tesis de Church-Turing

Ya sabemos que toda función recursiva es calculable mediante un algoritmo. Una forma más explícita de este hecho es el teorema siguiente:

Teorema 7.13 *Toda función recursiva parcial es* $1 \mid 1$*-computable.*

DEMOSTRACIÓN: Por inducción sobre el número r de funciones de una sucesión que defina a f. Si $r = 1$ se trata de una función recursiva elemental. La función cero c es computada por la máquina C, la función sucesor s es computada por I_1A y la proyección p_i^k es computada por I_{k-i+1}. (En toda esta prueba, "computable" significará "$1 \mid 1$-computable".)

Supongamos ahora que f se define en r pasos y que todas las funciones definibles en menos de r pasos son computables. Distinguimos tres casos, según que f se defina por composición, recursión o minimización a partir de funciones anteriores (que por hipótesis de inducción serán computables).

CASO A) $f(a_1, \ldots, a_n) = g(h_1(a_1, \ldots, a_n), \ldots, h_m(a_1, \ldots, a_n))$, donde las funciones g y h_i son computables por máquinas M_g y M_{h_i} respectivamente.

Veamos que la función f es computada por la máquina

$$\begin{aligned} M_f \quad = \quad & K_n M_{h_1} I_{n+1}^n M_{h_2} I_{n+1}^n \cdots I_{n+1}^n M_{h_m} \\ & I_{(m-1)(n+1)+1} I_{(m-2)(n+1)+2} \cdots I_{0(n+1)+m} M_g L B. \end{aligned}$$

Supongamos definido $f(a_1, \ldots, a_n)$. Si M_f empieza con $a_1, \ldots, \overline{a}_n$ (el guión sobre a_n indica que está en posición normal), en primer lugar K_n copia $a_1, \ldots, a_n$ con un vacío en medio:

$$a_1, \ldots, a_n, , a_1, \ldots, \overline{a}_n.$$

Luego M_{h_1} calcula $h_1(a_1, \ldots a_n)$:

$$a_1, \ldots, a_n, , a_1, \ldots, a_n, \overline{h_1(a_1, \ldots, a_n)}.$$

Ahora I_{n+1}^n copia $a_1, \ldots, a_n$ y M_{h_2} calcula $h_2(a_1, \ldots, a_n)$:

$$a_1, \ldots, a_n, , a_1, \ldots, a_n, h_1(a_1, \ldots, a_n), a_1, \ldots, a_n, \overline{h_2(a_2, \ldots, a_n)}.$$

Tras haber actuado M_{h_m} tenemos

$$a_1, \ldots, a_n, , a_1, \ldots, a_n, h_1(a_1, \ldots, a_n), \ldots, a_1, \ldots, a_n, \overline{h_m(a_2, \ldots, a_n)}.$$

Seguidamente las máquinas $I_{(m-1)(n+1)+1} I_{(m-2)(n+1)+2} \cdots I_{0(n+1)+m}$ copian $h_1(a_1, \ldots, a_n), \ldots, h_m(a_1, \ldots, a_n)$ y entonces M_g calcula la imagen de estos números por g, o sea, calcula $f(a_1, \ldots, a_n)$. La situación de la cinta es entonces:

$$a_1, \ldots, a_n, , x_1, \ldots, x_r, \overline{f(a_1, \ldots, a_n)}.$$

La máquina L borra $x_1, \ldots, x_r$ y B borra el vacío intermedio, hasta quedar

$$a_1, \ldots, a_n, \overline{f(a_1, \ldots, a_n)}.$$

Las casillas a la izquierda del blanco anterior a a_1 nunca han sido escrutadas durante el cálculo.

Si $f(a_1, \ldots, a_n)$ no está definida, entonces no lo está alguna de las funciones $g, h_1, \ldots, h_n$, por lo que la máquina correspondiente no se para y M_f tampoco.

Caso b) La función f está definida por recurrencia a partir de las funciones g y h, es decir:

$$\begin{aligned} f(0, a_2, \ldots, a_n) &= g(a_2, \ldots, a_n), \\ f(a+1, a_2, \ldots, a_n) &= h(a, f(a, a_2, \ldots, a_n), a_2, \ldots, a_n). \end{aligned}$$

Por hipótesis de inducción existen máquinas M_g y M_h que computan a g y h respectivamente. Razonando de forma similar al caso anterior es fácil ver que la función f es computada por la máquina

$$M_f = K_n M_g I_{n+1} E \left[\begin{array}{l} I_2 LB \\ \\ C I_3 I_{n+3}^{n-1} M_h I_{n+3} F E \end{array}\right. \left[\begin{array}{l} I_2 LB \\ \\ I_{n+3} A \end{array}\right.$$

Una ligera modificación da cuenta del caso $n = 1$.

Caso c) $f(a_1, \ldots, a_n) = \mu x\, g(a_1, \ldots, a_n, x) = 0$. Por hipótesis de inducción existe una máquina M_g que computa a g. Entonces la función f es computada por la máquina

$$M_f = K_n C M_g E \left[\begin{array}{l} I_2 LB \\ \\ I_{n+2}^{n+1} A \end{array}\right.$$

De este modo, para cada función recursiva parcial f sabemos construir explícitamente una máquina de Turing que la computa. ∎

Es claro que cualquier función computable por una máquina de Turing es computable mediante un ordenador (salvo limitaciones de memoria). El recíproco no está claro. Las máquinas de Turing tienen, en principio, muy poca capacidad de cálculo. No obstante hemos visto que pueden calcular cualquier función recursiva, lo que, a la larga, se traducirá en que la capacidad de cálculo de una máquina de Turing es idéntica a la de cualquier ordenador (superior —de hecho— por carecer de limitaciones de memoria). El punto más delicado de la demostración de la tesis de Church-Turing es probar el recíproco del teorema anterior. La clave del argumento está en que es general, en el sentido de que no sólo es aplicable a máquinas de Turing, sino que meros cambios técnicos permitirían adaptarlo para justificar que cualquier función calculable por un ordenador cualquiera es recursiva. Luego volveremos sobre este hecho.

Numeración de Gödel para máquinas de Turing Sea M una máquina de Turing con alfabeto $s_0, \ldots, s_j$. Sean $s_{u_0}, \ldots, s_{u_k}$ los signos impresos de izquierda a derecha a la izquierda de una casilla fija de la cinta empezando por el primero que sea distinto de s_0. Llamaremos *número de Gödel* de la cinta a la izquierda de dicha casilla al número $u = \langle u_1, \ldots, u_k \rangle$, es decir el número que codifica la sucesión $u_1, \ldots, u_k$ en el sentido definido en la sección 5.6.

Análogamente, se define el *número de Gödel* de la cinta a la derecha de una casilla dada como $v = \langle v_0, \ldots, v_l \rangle$, donde $s_{v_0}, \ldots, s_{v_l}$ son los signos impresos a la derecha de la casilla de derecha a izquierda, empezando por el primero distinto de s_0.

Si, en un instante dado, el número de Gödel a la izquierda de la casilla escrutada es u, el signo escrutado es s_a, el estado de M es q_c y el número de Gödel a la derecha de la casilla escrutada es v, el *número de Gödel* de la configuración completa de la máquina en ese instante es, por definición, $w = \langle u, a, c, v \rangle$.

Teorema 7.14 *Toda función parcial computable es recursiva parcial.*

DEMOSTRACIÓN: Sea ϕ una función parcial computable por una máquina de Turing M. Sean $q_0, \ldots q_k$ sus estados y $s_0, \ldots, s_j$ su alfabeto. Podemos suponer que q_0 es el único estado pasivo. Definimos en $\mathrm{I}\Sigma_1$ las fórmulas

$$\phi(s,t) \equiv (\ell(s) > 0 \wedge t = s|_{\ell(s)-1}) \vee (s = 0 \wedge t = 0),$$

$$\psi(s,n) \equiv (s > 0 \wedge n = s_{\ell(s)-1}) \vee (s = 0 \wedge n = 0).$$

Claramente, ambas fórmulas son Σ_1 y definen funciones aritméticas, que serán, por consiguiente, recursivas. A la función definida por ϕ la representaremos por s^-, de modo que si el número natural s codifica una sucesión no vacía, entonces s^- codifica la sucesión que resulta de eliminarle su último término, y si $s = 0$ entonces $s^- = 0$. La función recursiva definida por ψ la representaremos por s^+, de modo que si s codifica una sucesión no vacía, entonces s^+ es el último término de dicha sucesión, mientras que si $s = 0$ entonces $s^+ = 0$.

Por último, representaremos por $\langle x, y, z, w \rangle$ a la función recursiva definida por el término Δ_1 que denotamos igual. Así $\langle x, y, z, w \rangle$ es el número natural que codifica la sucesión formada por los cuatro argumentos.

Definamos, para cada configuración activa (s_a, q_c) una función $\rho_{a,c}$ como sigue:

Si el acto tras (s_a, q_c) es $s_b I q_d$, entonces

$$\rho_{a,c}(u,v) = \left\langle u^-, u^+, d, v^\frown \langle b \rangle \right\rangle.$$

Si el acto tras (s_a, q_c) es $s_b C q_d$, entonces

$$\rho_{a,c}(u,v) = \langle u, b, d, v \rangle.$$

Si el acto tras (s_a, q_c) es $s_b D q_d$, entonces

$$\rho_{a,c}(u,v) = \left\langle u^\frown \langle b \rangle, v^+, d, v^- \right\rangle$$

En cualquier caso, si $\langle u, a, c, v \rangle$ es el número de Gödel de la configuración completa de M en un instante, entonces $\rho_{a,c}(u,v)$ es el número de Gödel de la configuración completa siguiente. Es claro que cada función $\rho_{a,c}$ (como función de u, v únicamente) es recursiva, pues es composición de funciones recursivas.

Consideremos la relación $R_{a,c}$ (obviamente recursiva) dada por

$$R_{a,c}(w) \quad \text{syss} \quad \ell(w) = 4 \wedge w_1 = a \wedge w_2 = c.$$

Llamaremos $\chi_{a,c}$ a su función característica (también recursiva). Definimos

$$\rho(w) = \sum_{\substack{a=0,\ldots,j \\ c=1,\ldots,k}} \rho_{a,c}(w_0, w_3) \cdot \chi_{a,c}(w) + w \cdot \overline{\text{sg}}(w_2),$$

que es una función recursiva, por ser suma de funciones recursivas.

Si w es el número de Gödel de una configuración completa, $\rho(w)$ es el número de Gödel de la configuración completa siguiente (el sumando $w \cdot \overline{\text{sg}}(w_2)$ recoge el caso de que el estado sea pasivo, o sea, $w_2 = 0$, con lo que $\overline{\text{sg}}(w_2) = 1$ y así $\rho(w) = w$, es decir, la configuración no cambia).

Definimos ahora la función recursiva

$$\begin{aligned} \theta(w, 0) &= w, \\ \theta(w, z+1) &= \rho(\theta(w, z)). \end{aligned}$$

Si w es el número de Gödel de una configuración completa, $\theta(w, z)$ es el número de Gödel de la configuración completa en que se halla M después de z actos (o la situación final si M se detiene antes).

Para cada número natural n vamos a definir $\tau_n(x_1, \ldots, x_n, c, u, v)$ de modo que si $x_1, \ldots, x_n$ está representado en posición normal, el estado es q_c y los números de Gödel de la cinta a la izquierda de la casilla blanca anterior a x_1 y a la derecha de la casilla blanca posterior a x_n son, respectivamente, u y v, entonces τ_n da el número de Gödel de la configuración completa. Definimos primero

$$1_0 = 0, \qquad 1_{x+1} = 1_x \frown \langle 1 \rangle,$$

de modo que 1_x es el número natural que codifica la sucesión de longitud x cuyos términos son todos iguales a 1. Obviamente es recursiva. Ahora

$$\tau_1(x_1, c, u, v) = \langle u \frown \langle 0 \rangle \frown 1_{x_1}, 1, c, v \frown \langle 0 \rangle \rangle.$$

Supuesta definida τ_n, definimos τ_{n+1} como

$$\tau_{n+1}(x_1, \ldots, x_{n+1}, c, u, v) = \tau_1(x_{n+1}, c, \tau_n(x_1, \ldots, x_{n-1}, x_n + 1, c, u, v)_0, v).$$

Es fácil ver que las funciones τ_n cumplen lo pedido, así como que son recursivas.

Digamos que la función ϕ es n-ádica. Si $x_1, \ldots, x_n$ es escrutado en posición normal con estado q_1 y el resto de la cinta en blanco, la configuración completa es $\tau_n(x_1, \ldots, x_n, 1, 0, 0)$. Así mismo, si $x_1, \ldots, x_n, x$ (para un cierto x) es escrutado en posición normal con estado q_0, la configuración completa es $\tau_{n+1}(x_1, \ldots, x_n, x, 0, u, v)$, para ciertos u, v, y viceversa.

Así pues, $\phi(x_1, \ldots, x_n)$ está definido si y sólo si existen z, x, u y v tales que

$$\theta(\tau_n(x_1, \ldots, x_n, 1, 0, 0), z) = \tau_{n+1}(x_1, \ldots, x_n, x, 0, u, v),$$

y entonces $\phi(x_1, \ldots, x_n) = x$.

Equivalentemente, $\phi(x_1,\dots,x_n)$ está definido si y sólo si existe un número natural t ($t = \langle z, x, u, v\rangle$) que satisface la relación $S(x_1,\dots,x_n,t)$ dada por

$$\ell(t) = 4 \wedge \theta(\tau_n(x_1,\dots,x_n,1,0,0),t_0) = \tau_{n+1}(x_1,\dots,x_n,t_1,0,t_2,t_3),$$

y entonces $\phi(x_1,\dots,x_n) = t_1$. La relación S es claramente recursiva y tenemos que

$$\phi(x_1,\dots,x_n) = \big(\mu t\ \chi_{\neg S}(x_1,\dots,x_n,t) = 0\big)_1,$$

con lo que ϕ es recursiva parcial. ∎

Como consecuencia inmediata tenemos:

Teorema 7.15 *Una función parcial es recursiva parcial si y sólo si es computable, si y sólo si es* $1 \mid 1$ *computable.*

De aquí se deduce una consecuencia no trivial: supongamos que una función n-ádica recursiva parcial f es, de hecho, una función total, es decir, que está definida para todos los argumentos posibles $a_1,\dots,a_n$. En principio, esto no asegura que f sea recursiva, pues en la definición de f se puede haber empleado la minimización parcial, y el hecho de que en todos los casos requeridos para el cómputo de f existan los mínimos requeridos no implica necesariamente que los mínimos existan para todos los posibles argumentos de las funciones intermedias. No obstante:

Teorema 7.16 *Toda función total recursiva parcial es recursiva.*

Demostración: Sea ϕ una función n-ádica recursiva parcial que esté definida para todos los argumentos posibles. Por el teorema 7.13 es computable, luego podemos aplicarle el teorema 7.14, que nos permite expresar

$$\phi(x_1,\dots,x_n) = \big(\mu t\ \chi_{\neg S}(x_1,\dots,x_n,t) = 0\big)_1,$$

donde la función $\chi_{\neg S}$ es recursiva y el hecho de que ϕ esté definida para todos los argumentos posibles equivale a que siempre existe un t que anula a $\chi_{\neg S}$, luego la función $\mu t\,\chi_{\neg S}$ es también recursiva, y por lo tanto ϕ también. ∎

Consecuentemente:

Teorema 7.17 *Una función (total) es recursiva si y sólo si es computable.*

Nota Según anticipábamos, es fácil convencerse de que la prueba del teorema 7.14 anterior puede adaptarse para probar que toda función calculable con un programa de ordenador es recursiva. Para ello sólo hay que complicarla teniendo en cuenta la complejidad adicional de un ordenador frente a una máquina de Turing, pero es claro que no es necesario aportar ninguna idea nueva, sino que el esquema general del argumento sería exactamente el mismo. De hecho es teóricamente más simple, pues la cinta infinita se sustituye por una memoria finita, que sólo puede estar en un número finito de configuraciones. El comportamiento del ordenador está determinado por la configuración de su memoria y

por el estado de su microprocesador (incluyendo aquí cualquier hecho relevante, aunque no corresponda estrictamente al microprocesador). Podemos introducir una numeración de Gödel para la configuración de la memoria y el estado del microprocesador y la función que a partir del número de Gödel de la configuración completa calcula el de la siguiente configuración completa (entendiendo que vale 0 si el número de partida no corresponde a ninguna configuración completa posible) es recursiva (obviamente, pues es una función definida sobre una cantidad finita de números). La función que a partir de la configuración completa inicial calcula la configuración al cabo de n pasos es recursiva, lo cual se prueba exactamente igual que para máquinas de Turing y a su vez nos lleva ya sin ningún cambio a la conclusión del teorema.[3]

Teniendo esto en cuenta, el concepto de función recursiva puede considerarse como una caracterización precisa de la noción de computabilidad, al igual que el concepto de demostración formal es una caracterización precisa del concepto de razonamiento matemático. Algunos autores afirman que la tesis de Church-Turing es indemostrable, porque la noción de computabilidad mediante un algoritmo no admite una definición precisa. Se trata de autores antiguos (los que escribieron cuando todavía no se tenía una idea clara de las posibilidades de los ordenadores) o anticuados (los que se limitan a repetir reverentemente lo que dicen los autores antiguos). La realidad hoy en día es que nadie puede afirmar que una función es computable (en el sentido informal de la palabra) y al mismo tiempo reconocer que no sabría cómo programar a un ordenador para que la computara. El concepto de "ser computable por un ordenador (haciendo abstracción de las limitaciones de memoria)" captura exactamente la noción de "función computable" y ya hemos razonado que el argumento con máquinas de Turing se puede adaptar sin ninguna dificultad conceptual a un argumento análogo sobre cualquier modelo de ordenador que se quiera considerar.

Ejemplo Sea n un número natural. Un problema que se ha convertido en entretenimiento de algunos amantes de los acertijos matemáticos es el siguiente: encontrar una máquina de Turing M con dos signos 0 y 1 y a lo sumo n estados con la condición de que cuando empieza con la cinta en blanco se detiene tras haber escrito el máximo número posible de 1's. En otras palabras, se trata de encontrar el "récord" de unos que puede escribir una máquina de Turing con n estados excluyendo el caso trivial de que no se detenga nunca y escriba infinitos unos.

Más explícitamente, para cada máquina M que acaba deteniéndose cuando empieza con la cinta en blanco, llamamos p_M al número de unos que tiene la cinta cuando esto ocurre. Definimos $\Sigma(n)$ como el máximo de los números p_M, cuando M varía entre las máquinas que acaban deteniéndose al empezar con la cinta en blanco. El problema es, entonces, calcular los valores de Σ.

Teorema 7.18 *La función Σ no es recursiva.*

[3] En este argumento prescindimos de toda incorporación de nuevos datos durante el cálculo, es decir, suponemos que el ordenador no tiene teclado o conexión con otros ordenadores. Esto no es una restricción, pues únicamente nos interesa el intervalo comprendido desde que el ordenador tiene todos los datos introducidos hasta que termina el cálculo.

DEMOSTRACIÓN: Sea f una función recursiva y definamos

$$g(n) = \text{máx}\{f(2n+2), f(2n+3)\}.$$

Es fácil ver que g es recursiva y por lo tanto es computada por una máquina de Turing M. Sea k el número de estados activos de M.

Para cada número natural n sea N_n una máquina que al empezar con la cinta en blanco escriba el número n en la cinta y después actúe como M. Podemos construir N_n con $n+k+2$ estados. Cuando N_n actúa con la cinta en blanco, al acabar está escrito (entre otras cosas) $g(n)$, es decir, hay $g(n)+1$ unos en la cinta como mínimo.

Por lo tanto,

$$\text{máx}\{f(2n+2), f(2n+3)\} + 1 \leq \Sigma(n+k+2) \qquad \text{para todo } n.$$

Si $n \geq k$ tenemos que

$$f(2n+2), f(2n+3) < \Sigma(n+k+2) \leq \Sigma(2n+2) \leq \Sigma(2n+3),$$

ya que Σ es evidentemente creciente. Pero todo número x que cumpla $2k+3 \leq x$ puede expresarse como $x = 2n+2$ o $x = 2n+3$, para un cierto número $n \geq k$, y así $f(x) < \Sigma(x)$.

Hemos probado que Σ supera a cualquier función recursiva a partir de un cierto número natural. En particular Σ no es recursiva. ∎

Estos son algunos datos conocidos[4] sobre Σ:

$$\Sigma(1) = 1, \qquad \Sigma(2) = 4, \qquad \Sigma(3) = 6, \qquad \Sigma(4) = 13,$$

$$\Sigma(5) \geq 4\,098, \qquad \Sigma(6) \geq 3.5 \cdot 10^{18\,267}.$$

El problema de la detención Una pregunta natural que plantea la no recursividad de la función Σ es qué nos impide calcularla. Para calcular $\Sigma(n)$ hay que tomar todas las máquinas de Turing con dos signos y n estados, que son un número finito, seleccionar las que se detienen al empezar con la cinta en blanco y contar el máximo número de unos impreso por cada una de ellas. El único paso que no es evidentemente realizable es determinar cuáles se detienen, por lo que concluimos que no existe un método general para decidir si una máquina de Turing va a detenerse o no cuando comienza con una situación dada, es decir, el problema de la detención de las máquinas de Turing es insoluble.

En algunos casos podremos concluir algo a partir del análisis del programa, por ejemplo es fácil ver que la máquina D no se detiene cuando empieza con la cinta en blanco. En otros casos, en cambio, lo mejor que podremos hacer será ponerla a funcionar y ver si se detiene. Si lo hace sabremos que se para, pero si no se detiene nos quedaremos con la duda de si se va a parar más adelante o si no se va a parar nunca.

[4] Si el lector quiere información actualizada al respecto tiene que buscar "busy beaver" en internet.

7.6 Codificación de las funciones recursivas

Veamos ahora que toda función recursiva puede codificarse mediante un número natural. Para ello empezamos con la definición siguiente:

Definición 7.19 Si $f_0, \ldots, f_n$ es una sucesión de funciones parciales, una *justificación* para $f_0, \ldots, f_n$ es una sucesión de números naturales $\hat{f}_0, \ldots, \hat{f}_n$ tal que para cada $i = 0, \ldots, n$ se cumpla uno de los casos siguientes:

1. $f_i = c$ y $\hat{f}_i = \langle 1, \langle 1\rangle\rangle$.
2. $f_i = s$ y $\hat{f}_i = \langle 2, \langle 1\rangle\rangle$.
3. $f_i = p^k_j$ y $\hat{f}_i = \langle 3, \langle k, j\rangle\rangle$.
4. f_i está definida por composición parcial a partir de f_s, $f_{j_1}, \ldots, f_{j_r}$:

   $$f_i(x_1, \ldots, x_k) = f_s(f_{j_1}(x_1, \ldots, x_k), \ldots, f_{j_r}(x_1, \ldots, x_k)),$$

   para ciertos naturales $s, j_1, \ldots, j_r < i$ y $\hat{f}_i = \Big\langle 4, \Big\langle k, \hat{f}_s, \hat{f}_{j_1}, \ldots, \hat{f}_{j_r}\Big\rangle\Big\rangle$.
5. (a) f_i está definida por recursión parcial a partir de f_j y f_s:
   $f_i(0, x_1, \ldots, x_k) = f_j(x_1, \ldots, x_k)$
   $f_i(n+1, x_1, \ldots, x_k) = f_s(f_i(n, x_1, \ldots, x_k), n, x_1, \ldots, x_k)$,
   para ciertos $j, s < i$, $k \geq 1$ y $\hat{f}_i = \Big\langle 5, \Big\langle k+1, \hat{f}_j, \hat{f}_s\Big\rangle\Big\rangle$.

   (b) f_i está definida por recursión a partir de m y f_s:
   $f_i(0) = m$
   $f_i(n+1) = f_s(f_i(n), n)$
   con $s < i$ y $\hat{f}_i = \Big\langle 5, \Big\langle 1, m, \hat{f}_s\Big\rangle\Big\rangle$.
6. f_i está definida por minimización parcial a partir de f_j:
   $f_i(x_1, \ldots, x_k) = \mu n\ f_j(x_1, \ldots, x_k, n) = 0$,
   para ciertos $j < i$ y $k \geq 1$, y $\hat{f}_i = \Big\langle 6, \Big\langle k, \hat{f}_j\Big\rangle\Big\rangle$.

Aquí $\langle s_0, \ldots, s_n\rangle$ representa el número natural que codifica la sucesión de números naturales indicada. Es evidente que una sucesión de funciones satisface la definición de función recursiva parcial si y sólo si tiene una justificación. Más aún:

Teorema 7.20 *Si dos sucesiones $f_0, \ldots, f_n$ y $g_0, \ldots, g_m$ tienen justificaciones $\hat{f}_0, \ldots, \hat{f}_n$ y $\hat{g}_0, \ldots, \hat{g}_m$ y se cumple que $\hat{f}_i = \hat{g}_j$ para ciertos i, j, entonces $f_i = g_j$.*

DEMOSTRACIÓN: Razonamos por inducción sobre i, es decir, suponemos que el resultado es cierto para todo $i' < i$ y lo probamos para i. Notemos que $\hat{f}_i = \hat{g}_j$ es necesariamente una sucesión de longitud 2, cuya primera componente es $1, 2, 3, 4, 5, 6$. Si dicha primera componente es 1, necesariamente $f_i = g_j = c$, si es 2 entonces $f_i = g_j = s$, y si es 3, entonces la segunda componente es de la forma $\langle k, l\rangle$, y $f_i = g_j = p_l^k$.

Supongamos ahora que la primera componente es 4. Entonces la segunda es de la forma $\left\langle k, \hat{f}_s, \hat{f}_{j_1}, \ldots, \hat{f}_{j_r}\right\rangle = \left\langle k, \hat{g}_{s'}, \hat{g}_{j'_1}, \ldots, \hat{g}_{j'_r}\right\rangle$, y por hipótesis de inducción $f_s = g_{s'}$, $f_{j_l} = g_{j'_l}$ y f_i y g_j coinciden ambas con la función definida por composición parcial a partir de estas funciones. Los casos restantes son análogos a éste. ∎

Definición 7.21 Diremos que un número natural c es un *código de una función recursiva parcial* si existe una sucesión de funciones parciales $f_0, \ldots, f_n$ con una justificación $\hat{f}_0, \ldots, \hat{f}_n$ tal que $c = \hat{f}_n$. Acabamos de probar que la función f_n está unívocamente determinada por c, luego podemos llamarla f_c.

De este modo tenemos que, para cada código c, la función f_c es recursiva parcial, y toda función recursiva parcial es de la forma f_c para algún código c (no necesariamente único).

Teorema 7.22 *El conjunto C de códigos de funciones recursivas parciales es recursivo.*[5]

DEMOSTRACIÓN: Sea f la función dada por $f(n) = \langle \chi_C(0), \ldots, \chi_c(n)\rangle$, es decir, $f(n)$ es el número natural que codifica la sucesión indicada de ceros y unos. Basta probar que la función f es recursiva, pues entonces $\chi_c(n) = f(n)_n$ también lo será (por ser composición de dos funciones recursivas).

Para probarlo llamamos $a = f(0)$ y vamos a probar que f puede definirse en la forma

$$\begin{cases} f(0) = a, \\ f(n+1) = g(n, f(n)), \end{cases}$$

para cierta función recursiva $g(n, k)$. A su vez, basta definir la función g en la forma $g(n, k) = k^\frown \langle h(n, k)\rangle$, para cierta función recursiva h, de modo que lo que necesitamos es que cuando $k = \langle \chi_C(0), \ldots, \chi_C(n)\rangle$ se cumpla $h(n, k) = \chi_C(n+1)$. Definiremos h en la forma[6]

$$h(n, k) = \begin{cases} 1 & \text{si } R(n+1, k) \\ 0 & \text{en caso contrario,} \end{cases}$$

para cierta relación recursiva R, y se trata de definir R de modo que, cuando $k = \langle \chi_C(0), \ldots, \chi_C(n-1)\rangle$, se cumpla $R(n, k)$ si y sólo si $n \in C$. Notemos que,

[5]Notemos que un conjunto es lo mismo que una relación monádica. El teorema afirma que la relación "ser un código" es recursiva.

[6]Esto equivale a $y = h(n, k) \leftrightarrow (R(n+1, k) \wedge y = 1) \vee (\neg(R(n+1, k) \wedge y = 0)$, luego h será recursiva si R lo es.

por la definición de C, si se cumple $n \in C$, entonces el número de argumentos de la función parcial f_n es $(n_1)_0$. Por claridad escribiremos $\mathrm{Nar}(n) = (n_1)_0$, que es una función recursiva.

Definimos R como la conjunción de la relación recursiva

$$n > 0 \wedge \ell(k) = n - 1 \wedge \ell(n) = 2$$

y la disyunción de las relaciones siguientes, también recursivas:

1. $n_0 = 1 \wedge n = \langle 1, \langle 1\rangle\rangle$,
2. $n_0 = 2 \wedge n = \langle 2, \langle 1\rangle\rangle$,
3. $n_0 = 3 \wedge \ell(n_1) = 2 \wedge 1 \le (n_1)_1 \le (n_1)_0$,
4. $n_0 = 4 \wedge \ell(n_1) \ge 3 \wedge \bigwedge i < \ell(n_1)((i > 0 \to k_{(n_1)_i} = 1) \wedge$
   $(i > 1 \to \mathrm{Nar}((n_1)_i) = \mathrm{Nar}(n)) \wedge \ell(n_1) = \mathrm{Nar}((n_1)_1) + 2$.
5. $n_0 = 5 \wedge \ell(n_1) = 3 \wedge ((\mathrm{Nar}(n) > 1 \wedge k_{(n_1)_1} = k_{(n_1)_2} = 1 \wedge \mathrm{Nar}((n_1)_1) =$
   $\mathrm{Nar}(n) - 1 \wedge \mathrm{Nar}((n_1)_2) = \mathrm{Nar}(n) + 1)$
   $\vee\ (\mathrm{Nar}(n) = 1 \wedge k_{(n_1)_2} = 1 \wedge \mathrm{Nar}((n_1)_2 = 2)))$,
6. $n_0 = 6 \wedge \ell(n_1) = 2 \wedge k_{(n_1)_1} = 1 \wedge \mathrm{Nar}((n_1)_1) = \mathrm{Nar}(n) + 1$.

Es fácil ver que R cumple lo pedido.[7] ∎

Ahora consideramos el conjunto $\mathcal{A}$ formado por todos los números de la forma $\langle c, x, m\rangle$, donde $c \in C$ es un código de una función recursiva parcial f con $k = \mathrm{Nar}(c)$ argumentos y x es una sucesión de longitud k tal que está definido $m = f_c(x_0, \ldots, x_{k-1})$. Vamos a probar que $\mathcal{A}$ es Σ_1. Para ello definimos la fórmula de $\mathcal{L}_a$ (claramente Σ_1) dada por:

$$x \in \mathcal{A} \equiv \bigvee u n(\ell(u) = n + 1 \wedge u_n = x \wedge$$

$$\bigwedge i \le n(\ell(u_i) = 3 \wedge \bigvee f z(f = (u_i)_0 \wedge f \in C \wedge z = (u_i)_2 \wedge \cdots))),$$

donde los puntos suspensivos representan la disyunción de las fórmulas siguientes (todas ellas Σ_1):

1. $f = \langle 1, \langle 1\rangle\rangle \wedge \bigvee k\ u_i = \langle f, \langle k\rangle, 0\rangle$,
2. $f = \langle 2, \langle 1\rangle\rangle \wedge \bigvee k\ u_i = \langle f, \langle k\rangle, k + 1\rangle$,
3. $\bigvee x j k(f = \langle 3, \langle k, j\rangle\rangle \wedge \ell(x) = k \wedge u_i = \langle f, x, x_j\rangle)$,

[7] Por ejemplo, la condición 4) afirma que si $n_0 = 4$, entonces $n = \left\langle 4, \left\langle k, \hat{f}_s, \hat{f}_{j_1}, \ldots, \hat{f}_{j_r}\right\rangle\right\rangle$, de modo que $n_1 = \left\langle k, \hat{f}_s, \hat{f}_{j_1}, \ldots, \hat{f}_{j_r}\right\rangle$ cumple que todas sus componentes menos la primera son códigos ($k_{(n_i)_i} = 1$) que codifican funciones con k argumentos y que el número de argumentos de $\hat{f}_s$ es r. Esto equivale a que n sea el código de la composición parcial de las funciones cuyos códigos son $\hat{f}_s, \hat{f}_{j_1}, \ldots, \hat{f}_{j_r}$.

4. $\bigvee khgx(f = \langle 4, \langle k, h, g\rangle\rangle \wedge \ell(x) = k \wedge u_i = \langle f, x, z\rangle \wedge$
$\bigvee rjsw(\ell(g) = r \wedge \ell(j) = r \wedge \ell(w) = r) \wedge$
$s < i \wedge \bigwedge l < r\ (j_l < i \wedge u_{j_l} = \langle g_l, x, w_l\rangle) \wedge u_s = \langle h, w, z\rangle)$,

5. (a) $\bigvee kghx(f = \langle 5, \langle k+1, g, h\rangle\rangle \wedge \ell(x) = k \wedge$
$u_i = \langle f, \langle 0\rangle^\frown x, z\rangle \wedge \bigvee j < i\ u_j = \langle g, x, z\rangle)$,
(b) $\bigvee kghxm(f = \langle 5, \langle k+1, g, h\rangle\rangle \wedge \ell(x) = k \wedge u_i =$
$\langle f, \langle m+1\rangle^\frown x, z\rangle \wedge \bigvee rst(s, t < i \wedge u_t = \langle f, \langle m\rangle^\frown x, r\rangle \wedge u_s =$
$\langle h, \langle r, m\rangle^\frown x, z\rangle))$,
(c) $\bigvee ah(f = \langle 5, \langle 1, a, h\rangle\rangle \wedge u_i = \langle f, \langle 0\rangle, a\rangle)$,
(d) $\bigvee ahm(f = \langle 5, \langle 1, a, h\rangle\rangle \wedge u_i = \langle f, \langle m+1\rangle, z\rangle \wedge \bigvee rst(s, t < i \wedge u_t =$
$\langle f, \langle m\rangle, r\rangle \wedge u_s = \langle h, \langle r, m\rangle, z\rangle))$,

6. $\bigvee kgx(f = \langle 6, \langle k, g\rangle\rangle \wedge \ell(x) = k \wedge u_i = \langle f, x, z\rangle \wedge$
$\bigvee jw(\ell(j) = z+1 \wedge \ell(w) = z+1 \wedge \bigwedge l \le z(j_l < i \wedge u_{j_l} = \langle g, x^\frown\langle l\rangle, w_l\rangle) \wedge$
$w_z = 0 \wedge \bigwedge l < z\ w_l \neq 0))$.

Una simple rutina demuestra que la fórmula $x \in \mathcal{A}$ cumple lo indicado, es decir:

Teorema 7.23 *Si c es un código de una función recursiva parcial y f_c es una función k-ádica, entonces $f_c(a_1, \ldots, a_k)$ está definido si y sólo si existe un m tal que*

$$\mathbb{N} \vDash \left\langle 0^{(c)}, \left\langle 0^{(a_1)}, \ldots, 0^{(a_k)}\right\rangle, 0^{(m)}\right\rangle \in \mathcal{A},$$

y en tal caso $m = f_c(a_1, \ldots, a_k)$.

Ahora tenemos un ejemplo de conjunto no recursivo:

Teorema 7.24 *El conjunto $\mathcal{A}$ es Σ_1, pero no es recursivo.*

Demostración: Ya hemos visto que $\mathcal{A}$ es Σ_1. Si fuera recursivo también sería recursiva la función dada por[8]

$$f(m) = \begin{cases} 1 & \text{si } \langle m, \langle m\rangle, 0\rangle \in \mathcal{A}, \\ 0 & \text{en caso contrario.} \end{cases}$$

Por consiguiente, existiría un código c tal que $f = f_c$, pero entonces

$$f(c) = 1 \leftrightarrow \langle c, \langle c\rangle, 0\rangle \in \mathcal{A} \leftrightarrow f_c(c) = 0 \leftrightarrow f(c) = 0,$$

contradicción. ∎

Para interpretar este resultado conviene introducir una definición general:

[8]Esto equivale a $f(m) = n \leftrightarrow (\langle m, \langle m\rangle, 0\rangle \in \mathcal{A} \wedge y = 1) \vee (\langle m, \langle m\rangle, 0\rangle \notin \mathcal{A} \wedge y = 0)$, que sería una fórmula Σ_1 si $x \in \mathcal{A}$ fuera Δ_1.

Definición 7.25 Una relación en $\mathbb{N}$ es *semirrecursiva* si es Σ_1. Los conjuntos (es decir, las relaciones monádicas) semirrecursivas se llaman también *recursivamente numerables.*

El nombre de "semirrecursivo" se debe a que una relación R es recursiva si y sólo si R y $\neg R$ son semirrecursivas (que es otra forma de decir que una relación es Δ_1 si y sólo si es Σ_1 y Π_1). En términos de la tesis de Church-Turing, esto significa que una relación es semirrecursiva si existe un algoritmo que, cuando se cumple, nos permite comprobar que se cumple, pero si no se cumple, tal vez no lleguemos nunca a saber que no se cumple (el algoritmo puede prolongarse indefinidamente en el tiempo de modo que nunca tengamos la certeza de si terminará en algún momento o si no parará nunca).

El nombre de conjunto "recursivamente numerable" se debe a que un conjunto A de tipo Σ_1 está determinado por una fórmula $\bigvee y\, \phi(x, y)$, donde ϕ es Δ_0 y, por consiguiente, define una relación recursiva. Si suponemos que existe al menos un $a \in A$, la función

$$f(n) = \begin{cases} n_0 & \text{si } \mathbb{N} \vDash \phi(0^{(n_0)}, 0^{(n_1)}), \\ a & \text{en caso contrario,} \end{cases}$$

es claramente recursiva y la sucesión $f(0), f(1), f(2), \ldots$ recorre todos los elementos de A. Esto significa que podemos ir enumerando (mediante un algoritmo, con posibles repeticiones) todos los elementos de A, de modo que si un cierto número está en A, tarde o temprano aparecerá en la lista, mientras que si no está en A siempre nos quedará la duda de si aparecerá más adelante o no aparecerá nunca.

El significado del teorema anterior es, pues, que si la función f_c está definida para unos argumentos dados, podemos comprobar que así es (calculándola), pero si no está definida no tenemos garantía en general de que podamos llegar a confirmarlo (podremos iniciar el cálculo, pero éste no terminará nunca y no tenemos garantías de que podamos saber que así sucederá).

Ahora podemos enumerar los conjuntos recursivamente numerables:

Teorema 7.26 *Existe un conjunto Σ_1 universal U, es decir, un conjunto Σ_1 tal que para todo conjunto A de tipo Σ_1 existe un número n tal que, para todo número m, se cumple $m \in A$ syss $\langle n, m\rangle \in U$.*

Demostración: Consideramos la fórmula Σ_1 dada por

$$x \in U \equiv \bigvee r\ \ \langle x_0, \langle r, x_1\rangle, 1\rangle \in \mathcal{A},$$

y sea U el conjunto definido por ella. Veamos que cumple lo pedido. Si A es un conjunto Σ_1, sabemos que

$$m \in A \quad \text{syss} \quad \mathbb{N} \vDash \bigvee r\, \phi(r, 0^{(m)}),$$

para cierta fórmula ϕ de tipo Δ_0, que define una relación recursiva R. Así

$$m \in A \quad \text{syss} \quad \bigvee r\ \chi_R(r, m) = 1.$$

Como χ_R es una función recursiva, existe un $n \in C$ tal que $\chi_R = f_n$. Entonces

$$m \in A \quad \text{syss} \quad \bigvee r\ f_n(r, m) = 1 \quad \text{syss} \quad \bigvee r\ \langle n, \langle r, m\rangle, 1\rangle \in \mathcal{A} \quad \text{syss} \quad \langle n, m\rangle \in U.$$

■

Notemos que si llamamos U_n al conjunto de todos los m tales que $\langle n, m\rangle \in U$, cada conjunto U_n es Σ_1 (pues está definido por la fórmula $\langle 0^{(n)}, m\rangle \in U$), y acabamos de probar que los conjuntos U_n recorren todos los conjuntos Σ_1.

El conjunto U es otro ejemplo de conjunto semirrecursivo no recursivo. Si fuera recursivo, todos los conjuntos U_n serían recursivos, lo cual es tanto como afirmar que todo conjunto semirrecursivo sería recursivo, y ya sabemos que esto es falso.

7.7 Relaciones diofánticas

Una *ecuación diofántica* es una ecuación polinómica con coeficientes enteros de la que se buscan las posibles soluciones enteras. El *décimo problema de Hilbert* consiste en encontrar un método para encontrar las soluciones de cualquier ecuación diofántica dada o, al menos, determinar si tiene o no solución. Como aplicación de los resultados de este capítulo veremos que este problema no tiene solución, es decir, que no puede existir tal método.

Definición 7.27 Una fórmula de $\mathcal{L}_a$ es *diofántica* si es de la forma

$$\bigvee y_1 \cdots y_m\, P(x_1, \ldots, x_n, y_1, \ldots, y_n) = Q(x_1, \ldots, x_n, y_1, \ldots, y_n),$$

donde P y Q son *polinomios* con coeficientes naturales, es decir, términos que se expresan como suma de *monomios,* que a su vez son términos de la forma $az_1 \cdots z_r$, donde cada z_i es una variable y a es un numeral. Más en general, una fórmula es diofántica en una teoría T sobre $\mathcal{L}_a$ si es equivalente en T a una fórmula diofántica.

Una relación n-ádica R en $\mathbb{N}$ es *diofántica* si

$$R(a_1, \ldots, a_n) \quad \text{syss} \quad \mathbb{N} \vDash \phi(0^{(a_1)}, \ldots, 0^{(a_n)}),$$

donde ϕ es una fórmula diofántica. Una función f es *diofántica* si lo es la relación $f(a_1, \ldots, a_n) = a$.

Hemos definido de este modo las fórmulas diofánticas para que sea evidente que las fórmulas diofánticas son Σ_1 y, de hecho, constituyen una de las clases más simples posibles de fórmulas Σ_1. Sin embargo, vamos a ver ahora que cada fórmula diofántica es equivalente a otra mucho más próxima al concepto de ecuación diofántica.

Ante todo, por el teorema 6.45, es claro que una fórmula de tipo

$$\bigvee y_1 \cdots y_m \, P(x_1, \ldots, x_n, y_1, \ldots, y_m) = Q(x_1, \ldots, x_n, y_1, \ldots, y_m)$$

es equivalente en $\mathrm{I}\Sigma_1$ a

$$\bigvee y_1 \cdots y_m \in \mathbb{N} \, P_1(\bar{x}_1, \ldots, \bar{x}_n, y_1, \ldots, y_m) = Q_1(\bar{x}_1, \ldots, \bar{x}_n, y_1, \ldots, y_m),$$

donde ahora $\mathbb{N} \equiv \mathbb{N}_1$ representa al conjunto de los números naturales definidos en $\mathbb{Q}$, el término $\bar{x} \equiv (+x/1)$ representa al número racional que se identifica con el número natural x (en la sección 6.6 llamábamos x_1 a este mismo término) y P_1, Q_1 son polinomios en $\mathbb{Q}$ con coeficientes naturales, es decir, que son sumas de monomios (respecto de la suma en $\mathbb{Q}$) y los monomios son productos de variables (respecto del producto en $\mathbb{Q}$) cuyos coeficientes son designadores de la forma $+0^{(a)}$. Pero, como ahora "estamos en $\mathbb{Q}$", podemos despejar y reducir la fórmula a

$$\bigvee y_1 \cdots y_m \in \mathbb{N} \, R(\bar{x}_1, \ldots, \bar{x}_n, y_1, \ldots, y_m) = 0,$$

donde $R = \bar{P} - \bar{Q}$ es ahora un polinomio en $\mathbb{Q}$ con coeficientes enteros (es decir, un polinomio cuyos monomios tienen coeficientes de la forma $\pm\overline{0^{(a)}}$). Recíprocamente, todo polinomio con coeficientes enteros puede descomponerse en la forma $R = \bar{P} - \bar{Q}$, donde $\bar{P}$ y $\bar{Q}$ son polinomios en $\mathbb{Q}$ con coeficientes naturales y, por consiguiente, son de la forma $\bar{P} = P_1$, $\bar{Q} = Q_2$, para ciertos polinomios en el sentido de la definición 7.27. Concluimos que una fórmula es diofántica si y sólo si es equivalente en $\mathrm{I}\Sigma_1$ a una de la forma

$$\bigvee y_1 \cdots y_m \in \mathbb{N} \, P(\bar{x}_1, \ldots, \bar{x}_n, y_1, \ldots, y_m) = 0,$$

donde P es un polinomio con coeficientes enteros. Ésta es la forma en la que normalmente manejaremos las relaciones diofánticas, pero vamos a ver que en realidad es equivalente a otra versión en la que aparecen ecuaciones diofánticas propiamente dichas:

Teorema 7.28 *Una fórmula es diofántica si y sólo si es equivalente en* $\mathrm{I}\Sigma_1$ *a otra de la forma*

$$\bigvee y_1 \cdots y_m \in \mathbb{Z} \, P(\bar{x}_1, \ldots, \bar{x}_n, y_1, \ldots, y_m) = 0,$$

donde P es un polinomio con coeficientes enteros.

Demostración: Dado un polinomio P en las condiciones del enunciado, podemos construir otro polinomio Q mediante[9]

$$Q(x_1, \ldots, x_n, y_1, \ldots, y_m) = \prod_{(\epsilon_1, \ldots, \epsilon_m)} P(x_1, \ldots, x_n, \epsilon_1 y_1, \ldots, \epsilon_m y_m),$$

[9]La construcción es metamatemática, es decir, dado el término P, estamos explicando cómo construir otro término Q, a saber, el que se obtiene multiplicando un número finito de variantes de P y operando los productos de sumas mediante la propiedad distributiva hasta obtener una suma de monomios, es decir, un polinomio. Como la propiedad distributiva es demostrable en $\mathrm{I}\Sigma_1$, es claro que en esta misma teoría se prueba la igualdad entre el producto y el polinomio resultante Q.

donde $(\epsilon_1, \ldots, \epsilon_m)$ recorre todos los valores posibles con $\epsilon_i = \pm 1$. Es claro que la fórmula del enunciado equivale en $\mathrm{I}\Sigma_1$ a

$$\bigvee y_1 \cdots y_m \in \mathbb{N}\, Q(\bar{x}_1, \ldots, \bar{x}_n, y_1, \ldots, y_m) = 0,$$

luego es diofántica. Para la otra implicación necesitamos un truco diferente. Lagrange demostró que todo número natural es suma de cuatro cuadrados.[10] Por lo tanto, dada una fórmula

$$\bigvee y_1 \cdots y_m \in \mathbb{N}\, Q(\bar{x}_1, \ldots, \bar{x}_n, y_1, \ldots, y_m) = 0,$$

podemos definir

$$P(x_1, \ldots, x_n, p_1, q_1, r_1, s_1, \ldots, p_m, q_m, r_m, s_m)$$

$$= Q(x_1, \ldots, x_n, p_1^2 + q_1^2 + r_1^2 + s_1^2, \ldots, p_m^2 + q_m^2 + r_m^2 + s_m^2).$$

Este término puede operarse en $\mathrm{I}\Sigma_1$ hasta llegar a un polinomio, y es claro que la fórmula dada equivale a

$$\bigvee p_1 q_1 r_1 s_1 \ldots p_m q_m r_m s_m \in \mathbb{Z}\, P(\bar{x}_1, \ldots, \bar{x}_n, p_1, \ldots, s_m) = 0.$$

■

Así pues, una relación n-ádica R es diofántica si y sólo si existe un polinomio P con coeficientes enteros y n parámetros de modo que $R(a_1, \ldots, a_n)$ es equivalente a que la ecuación diofántica definida por P cuando los parámetros se interpretan como $\bar{0}^{(a_1)}, \ldots, \bar{0}^{(a_n)}$ tiene solución. El teorema central que pretendemos probar es el siguiente:

Teorema 7.29 *Una fórmula de $\mathcal{L}_a$ es diofántica en $\mathrm{I}\Sigma_1$ si y sólo si es Σ_1.*

Esto implica a su vez que una relación es diofántica si y sólo si es Σ_1. Una implicación es trivial y si admitimos la contraria es fácil ver que el décimo problema de Hilbert no tiene solución: sea U cualquier conjunto Σ_1 no recursivo. Entonces

$$n \in U \quad \text{syss} \quad \mathbb{N} \vDash \bigvee y_1 \cdots y_m \in \mathbb{Z}\, P(\overline{0^{(n)}}, y_1, \ldots, y_m) = 0,$$

donde P es un polinomio con coeficientes enteros. Si hubiera una forma de determinar si cada una de las ecuaciones diofánticas $P(\overline{0^{(n)}}, y_1, \ldots, y_m) = 0$ tiene o no solución, tendríamos una forma de determinar si cada número n está o no en U, pero entonces el conjunto U sería recursivo, y no es el caso. Por lo tanto, no puede haber un método para determinar, no ya si cualquier ecuación diofántica dada tiene solución, sino si la tienen las ecuaciones diofánticas de una cierta familia uniparamétrica.

[10] Véase mi libro de Introducción a la teoría algebraica de números (teorema 3.5). La prueba se formaliza sin dificultad en $\mathrm{I}\Sigma_1$.

Algunos hechos elementales Comenzamos probando un par de hechos sencillos que usaremos en todo momento. El primero es que un sistema de ecuaciones diofánticas equivale en realidad a una única ecuación:

Teorema 7.30 *Para cada $i = 1, \dots, r$, sea $P_i(x_1, \dots, x_n, y_1, \dots, y_m)$ un polinomio con coeficientes enteros. Entonces la fórmula*

$$\bigvee y_1 \dots y_m \in \mathbb{N}(P_1 = 0 \wedge \dots \wedge P_r = 0)$$

es diofántica.

DEMOSTRACIÓN: Basta observar que equivale a

$$\bigvee y_1 \dots y_m \in \mathbb{N}\, P_1^2 + \dots + P_r^2 = 0.$$

■

Por otra parte tenemos unos pocos procedimientos generales para construir unas fórmulas diofánticas a partir de otras:

Teorema 7.31 *Si ϕ y ψ son fórmulas diofánticas en $\mathrm{I}\Sigma_1$, también lo son las fórmulas $\phi \wedge \psi$, $\phi \vee \psi$ y $\bigvee x\, \phi$.*

DEMOSTRACIÓN: Supongamos que

$$\begin{aligned} \phi(x_1, \dots, x_n) &\leftrightarrow \bigvee y_1 \cdots y_m \in \mathbb{N}\, P(\bar{x}_1, \dots, \bar{x}_n, y_1, \dots, y_m) = 0, \\ \psi(x_1, \dots, x_n) &\leftrightarrow \bigvee z_1 \cdots z_r \in \mathbb{N}\, Q(\bar{x}_1, \dots, \bar{x}_n, z_1, \dots, z_r) = 0. \end{aligned}$$

Entonces

$$\begin{aligned} (\phi \wedge \psi)(x_1, \dots, x_n) &\leftrightarrow \bigvee y_1 \cdots y_m z_1, \dots, z_r \in \mathbb{N}\, P^2 + Q^2 = 0, \\ (\phi \vee \psi)(x_1, \dots, x_n) &\leftrightarrow \bigvee y_1 \cdots y_m z_1, \dots, z_r \in \mathbb{N}\, PQ = 0. \end{aligned}$$

Si

$$\phi(x_1, \dots, x_n, x) \leftrightarrow \bigvee y_1 \cdots y_m \in \mathbb{N}\, P(\bar{x}_1, \dots, \bar{x}_n, \bar{x}, y_1, \dots, y_m) = 0,$$

entonces

$$\bigvee x\, \phi(x_1, \dots, x_n, x) \leftrightarrow \bigvee x y_1 \cdots y_m \in \mathbb{N}\, P(\bar{x}_1, \dots, \bar{x}_n, x, y_1, \dots, y_m) = 0.$$

■

Algunas expresiones diofánticas Para probar que toda fórmula Σ_1 es diofántica necesitamos justificar el carácter diofántico de varias expresiones. El punto más difícil es el teorema siguiente, cuya prueba dejamos para el final:

Teorema 7.32 (Matiyasevič) *El término x^y es diofántico.*

Este teorema lo probó Matiyasevič en 1970. Los resultados que veremos a continuación los obtuvo Julia Robinson en 1952 tomando como conjetura el teorema anterior.[11]

[11] En realidad Robinson definió las funciones exponencial-diofánticas como las definibles en términos de ecuaciones diofánticas y de la función exponencial y probó que las funciones que vamos a estudiar ahora eran exponencial-diofánticas.

Todos los resultados se demuestran[12] en $\mathrm{I}\Sigma_1$. Supondremos al lector familiarizado con los resultados básicos de la aritmética de los números enteros y racionales, cuyas pruebas se formalizan sin dificultad en $\mathrm{I}\Sigma_1$. Por simplicidad dejamos de distinguir entre un número natural x y su número racional asociado $\bar{x}$. En primer lugar demostramos que los números combinatorios son diofánticos, para lo cual nos apoyamos en el teorema siguiente, en el que E representa la parte entera:

Teorema 7.33 *Si $0 \leq k \leq n$ y $u > 2^n$, entonces*

$$E\left[\frac{(u+1)^n}{u^k}\right] \equiv \binom{n}{k} \text{ (mód } u).$$

Demostración: Por la fórmula del binomio tenemos que

$$\frac{(u+1)^n}{u^k} = \sum_{i=0}^{n}\binom{n}{i}u^{i-k} = \sum_{i=k}^{n}\binom{n}{i}u^{i-k} + \sum_{i=0}^{k-1}\binom{n}{i}u^{i-k}.$$

Ahora bien,

$$\sum_{i=0}^{k-1}\binom{n}{i}u^{i-k} \leq \frac{1}{u}\sum_{i=0}^{k-1}\binom{n}{i} < \frac{1}{u}\sum_{i=0}^{n}\binom{n}{i} = \frac{2^n}{u} < 1.$$

Por consiguiente

$$\sum_{i=k}^{n}\binom{n}{i}u^{i-k} \leq \frac{(u+1)^n}{u^k} < \sum_{i=k}^{n}\binom{n}{i}u^{i-k} + 1,$$

es decir,

$$E\left[\frac{(u+1)^n}{u^k}\right] = \sum_{i=k}^{n}\binom{n}{i}u^{i-k} = \binom{n}{k} + u\sum_{i=k+1}^{n}\binom{n}{i}u^{i-k-1} \equiv \binom{n}{k} \text{ (mód } u).$$

■

Teorema 7.34 *El término $\binom{n}{k}$ es diofántico.*

Demostración: Observemos que $\binom{n}{k} \leq \sum_{i=0}^{n}\binom{n}{i} = 2^n$. Por el teorema anterior, para cualquier $u > 2^n$ tenemos que $\binom{n}{k}$ es el único número natural congruente con $E\left[\frac{(u+1)^n}{u^k}\right]$ módulo u y menor que u. Por lo tanto

$$z = \binom{n}{k} \leftrightarrow \bigvee uv \in \mathbb{N}(v = 2^n \wedge u > v$$

$$\wedge\, w = E\left[\frac{(u+1)^n}{u^k}\right] \wedge z \leq w \wedge z \equiv w \text{ (mód } u) \wedge z < u).$$

[12]No obstante, para probar que las relaciones Σ_1 son diofánticas nos basta con probar que las fórmulas Σ_1 son diofánticas en AP, y para ello podemos trabajar en AP, es decir, despreocupándonos de la complejidad de las expresiones que manejamos.

Por el teorema 7.31, basta probar que las fórmulas que aparecen dentro del paréntesis son diofánticas. Ahora bien:

$v = 2^n \leftrightarrow \bigvee x \in \mathbb{N}(x = 2 \wedge v = x^n)$ es diofántica por 7.32.

$u > v \leftrightarrow \bigvee x \in \mathbb{N}\, u = v + x + 1$, diofántica.

$w = E\left[\frac{(u+1)^n}{u^k}\right]$ equivale a

$$\bigvee xyt \in \mathbb{N}(t = u + 1 \wedge x = t^n \wedge y = u^k \wedge w \leq (x/y) < w + 1)$$

y también a

$$\bigvee xyt \in \mathbb{N}(t = u + 1 \wedge x = t^n \wedge y = u^k \wedge wy \leq x < (w + 1)y),$$

claramente diofántica.

$$z \leq w \wedge z \equiv w \ (\text{mód } u) \wedge z < u \leftrightarrow \bigvee xy \in \mathbb{N}(w = z + xu \wedge u = z + y + 1)$$

diofántica. ∎

Nos ocupamos ahora de la función factorial:

Teorema 7.35 *Si $r > (2x)^{x+1}$, entonces $x! = E[r^x/\binom{r}{x}]$.*

Demostración: Podemos suponer $x > 0$.

$$\frac{r^x}{\binom{r}{x}} = \frac{r^x\, x!}{r(r-1)\cdots(r-x+1)} = \frac{x!}{(1-\frac{1}{r})\cdots(1-\frac{x-1}{r})} < \frac{x!}{(1-\frac{x}{r})^x}.$$

Vamos a usar la desigualdad

$$\frac{1}{1-\frac{x}{r}} < 1 + \frac{2x}{r}$$

(al desarrollarla equivale a $r > 2x$). Se cumple que

$$\left(1 + \frac{2x}{r}\right)^x = \sum_{j=0}^{x} \binom{x}{j} \left(\frac{2x}{r}\right)^j = 1 + \frac{2x}{r} \sum_{j=1}^{x} \binom{x}{j} \left(\frac{2x}{r}\right)^{j-1}$$

$$\leq 1 + \frac{2x}{r} \sum_{j=1}^{x} \binom{x}{j} < 1 + \frac{2x}{r}\, 2^x.$$

Así pues,

$$\frac{r^x}{\binom{r}{x}} < x! \left(1 + \frac{2x}{r}\, 2^x\right) \leq x! + \frac{2^{x+1} x^{x+1}}{r} = x! + \frac{(2x)^{x+1}}{r} < x! + 1,$$

con lo que

$$x! \leq \frac{r^x}{\binom{r}{x}} < x! + 1.$$

La primera desigualdad se sigue, por ejemplo, de la primera línea de ecuaciones. ∎

Teorema 7.36 *El término $n!$ es diofántico.*

DEMOSTRACIÓN: Teniendo en cuenta el teorema anterior vemos que

$$m = n! \leftrightarrow \bigvee rstuv \in \mathbb{N}(s = 2n+1 \wedge t = n+1 \wedge r = s^t \wedge u = r^n$$

$$\wedge\, v = \tbinom{r}{n} \wedge mv \leq u < (m+1)v).$$

■

El último resultado que necesitamos es el siguiente:[13]

Teorema 7.37 *La fórmula $b > 0 \wedge z = \prod_{k=1}^{y} (a+bk)$ es diofántica.*

DEMOSTRACIÓN: Veamos primero que si $bq \equiv a \pmod{M}$ entonces

$$\prod_{k=1}^{y} (a+bk) \equiv b^y y! \binom{q+y}{y} \pmod{M}.$$

Si $y = 0$ es inmediato (entendiendo que el producto vale 1). En otro caso

$$\begin{aligned} b^y y! \binom{q+y}{y} &= b^y (q+y)(q+y-1)\cdots(q+1) \\ &= (bq+yb)(bq+(y-1)b)\cdots(bq+b) \\ &\equiv (a+yb)(a+(y-1)b)\cdots(a+b) \pmod{M}. \end{aligned}$$

Tomemos ahora $M = b(a+by)^y + 1$. Teniendo en cuenta que $b > 0$, es claro que $(M, b) = 1$ y $M > \prod_{k=1}^{y} (a+bk)$. Como b es primo con M, existe un q tal que $bq \equiv a \pmod{M}$. Así, $\prod_{k=1}^{y} (a+bk)$ es el único natural congruente con $b^y y! \binom{q+y}{y}$ menor que M.

$$b > 0 \wedge z = \prod_{k=1}^{y} (a+bk) \leftrightarrow b > 0 \wedge \bigvee Mpqrstuvwx \in \mathbb{N}(r = a+by \wedge s = r^y$$

$$\wedge\, M = bs+1 \wedge bq = a+Mt \wedge u = b^y \wedge v = y! \wedge z < M$$

$$\wedge\, w = q+y \wedge x = \tbinom{w}{y} \wedge z + Mp = uvx),$$

y el miembro derecho es claramente diofántico. ■

[13] Este teorema fue probado por primera vez por Davis y Putnam en 1958, basándose en ideas de Julia Robinson. Aquí damos una prueba posterior debida a Robinson.

Las fórmulas Σ_1 son diofánticas El teorema siguiente se debe a Davis-Putnam-Robinson (1961):

Teorema 7.38 *Si la fórmula $\phi(y, z, x_1, \ldots, x_n)$ es diofántica, también lo es $\bigwedge z \le y\, \phi(y, z, x_1, \ldots, x_n)$.*

DEMOSTRACIÓN: Digamos que

$$\phi(y, z, x_1, \ldots, x_n) \leftrightarrow \bigvee y_1 \ldots y_m \in \mathbb{N}\, P(y, z, x_1, \ldots, x_n, y_1, \ldots, y_m) = 0$$

y llamemos $\psi(y, x_1, \ldots, x_n)$ a la fórmula del enunciado. Por recolección

$$\psi(y, x_1, \ldots, x_n) \leftrightarrow \bigvee u \in \mathbb{N}(u > 0 \wedge \bigwedge z \le y$$

$$\bigvee y_1 \ldots y_m \in \mathbb{N}(y_1, \ldots, y_m \le u \wedge P(y, z, x_1, \ldots, x_n, y_1, \ldots, y_m) = 0)).$$

Expresemos el polinomio P como suma de monomios $P = \sum_{r=1}^{N} t_r$, donde[14]

$$t_r = cy^a z^b x_1^{q_1} \cdots x_n^{q_n} y_1^{s_1} \cdots y_m^{s_m}, \qquad c \in \mathbb{Z}.$$

Sea $u_r = |c| y^{a+b} x_1^{q_1} \cdots x_n^{q_n} u^{s_1+\cdots+s_m}$. Sea

$$Q(y, u, x_1, \ldots, x_n) = u + y + \sum_{r=1}^{N} u_r + 1.$$

Se cumple que $Q(y, u, x_1, \ldots, x_n) > u$, $Q(y, u, x_1, \ldots, x_n) > y$ y si $z \le y$, $y_1, \ldots, y_m \le u$, entonces

$$|P(y, z, x_1, \ldots, x_n, y_1, \ldots, y_m)| \le Q(y, u, x_1, \ldots, x_n).$$

Sea $u > 0$. Veamos que

$$\bigwedge z \le y \bigvee y_1 \ldots y_m \in \mathbb{N}(y_1, \ldots, y_m \le u \wedge P(y, z, x_1, \ldots, x_n, y_1, \ldots, y_m) = 0)).$$

equivale a

$$\bigvee c t a_1 \cdots a_m \in \mathbb{N}(t > 0 \wedge c > 0 \wedge 1 + ct = \prod_{k=1}^{y} (1 + kt) \wedge t = Q(y, u, x_1, \ldots, x_n)!$$

$$\wedge (1 + ct) \mid \prod_{j=0}^{u} (a_1 - j) \wedge \cdots \wedge (1 + ct) \mid \prod_{j=0}^{u} (a_m - j)$$

$$\wedge P(y, c, x_1, \ldots, x_n, a_1, \ldots, a_m) \equiv 0 \ (\text{mód } 1 + ct))$$

$$\wedge \bigvee y_1, \ldots y_m \in \mathbb{N}(y_1, \ldots, y_m \le u \wedge P(y, 0, x_1, \ldots, x_n, y_1, \ldots, y_m) = 0).$$

La condición es suficiente: Tomamos $z \le y$, y hemos de probar que

$$\bigvee y_1 \ldots y_m \in \mathbb{N}(y_1, \ldots, y_m \le u \wedge P(y, z, x_1, \ldots, x_n, y_1, \ldots, y_m) = 0).$$

Para $z = 0$ es exactamente lo que afirma la última parte de la hipótesis. Supongamos que $0 < z \le y$. Sea p_z un divisor primo de $1 + zt$, sea y_i^z el resto de dividir a_i entre p_z, para $i = 1, \ldots, m$. Se cumple que $y_i^z < p_z$. Veamos que

[14] Aquí hay que entender que c y los exponentes son designadores, mientras que y, z, x_i, y_j son variables.

1. $y_i^z \le u$.
2. $P(y, z, x_1, \dots, x_n, y_1^z, \dots, y_m^z) = 0$.

En efecto, $p_z \mid (1+zt) \mid (1+ct) \mid \prod_{j=0}^{u}(a_i - j)$, luego existe un j tal que $0 \le j \le u$ tal que $p_z \mid (a_i - j)$, o sea, $j \equiv a_i \equiv y_i^z \pmod{p_z}$.

Como $p_z|(1+zt)$, tenemos que $p_z \nmid t$ y, como $t = Q(y, u, x_1, \dots, x_n)!$,

$$p_z > Q(y, u, x_1, \dots, x_n) > u.$$

Tenemos que $j \le u < p_z$, $y_i^z < p_z$ y, como son congruentes, ha de ser $y_i^z = j$, luego se cumple 1).

$1 + ct \equiv 0 \equiv 1 + zt \pmod{p_z}$, luego $z + zct \equiv c + zct \pmod{p_z}$ y de aquí que $z \equiv c \pmod{p_z}$. Tenemos también que $y_i^z \equiv a_i \pmod{p_z}$, luego

$$P(y, z, x_1, \dots, x_n, y_1^z, \dots, y_m^z) \equiv P(y, c, x_1, \dots, x_n, a_1, \dots, a_m) \equiv 0 \pmod{p_z}.$$

Pero $|P(y, z, x_1, \dots, x_n, y_1^z, \dots, y_m^z)| \le Q(y, u, x_1, \dots, x_n) < p_z$, lo que implica que $P(y, z, x_1, \dots, x_n, y_1^z, \dots, y_m^z) = 0$. Esto prueba 2) y también la suficiencia.

La condición es necesaria: La última parte es inmediata. Hemos de ver la primera. Para cada $0 < z \le y$, sean $y_1^z, \dots, y_m^z \le u$ tales que

$$P(y, u, x_1, \dots, x_n, y_1^z, \dots, y_m^z) = 0.$$

Sea $t = Q(y, u, x_1, \dots, x_n)! > 0$. Como $\prod_{k=1}^{y}(1 + kt) \equiv 1 \pmod{t}$, existe un $c > 0$ tal que $1 + ct = \prod_{k=1}^{y}(1 + kt)$.

Veamos que si $1 \le k < l \le y$, entonces $(1 + kt, 1 + lt) = 1$. En efecto, si $p \mid (1 + kt)$ y $p \mid (1 + lt)$ es un divisor primo común, entonces $p \mid (l - k)$, luego $p < y$, pero $Q(y, u, x_1, \dots, x_n) > y$, de donde $p \mid t$ y $p \mid (1 + kt)$, y así $p \mid 1$, contradicción.

Por el teorema chino del resto, para cada $1 \le i \le m$ existe un a_i tal que

$$a_i \equiv y_i^z \pmod{1 + zt} \qquad z = 1, \dots, y.$$

Como $1+ct \equiv 0 \pmod{1+zt}$ y $1 \equiv -zt \pmod{1+zt}$, tenemos que $(c-z)t \equiv 0 \pmod{1 + zt}$ y, como $(t, 1+zt) = 1$, resulta que $c - z \equiv 0 \pmod{1 + zt}$, es decir, $c \equiv z \pmod{1 + zt}$. Ahora,

$$P(y, c, x_1, \dots, x_n, a_1, \dots, a_m) \equiv P(y, z, x_1, \dots, x_n, y_1^z, \dots, y_m^z) = 0 \pmod{1+zt},$$

luego $1+zt \mid P(y, c, x_1, \dots, x_n, a_1, \dots, a_m)$ y, como los $1+zt$ son primos entre sí, su producto también divide a P, es decir $1 + ct \mid P(y, c, x_1, \dots, x_n, a_1, \dots, a_m)$ o, equivalentemente,

$$P(y, c, x_1, \dots, x_n, a_1, \dots, a_m) \equiv 0 \pmod{1 + ct}.$$

Como $a_i \equiv y_i^z \pmod{1+zt}$, tenemos que $1+zt \mid a_i - y_i^z$, de donde se sigue que $1+zt \mid \prod_{j=0}^{u}(a_i - j)$. Como los $1+zt$ son primos entre sí, también $1+ct \mid \prod_{j=0}^{u}(a_i - j)$ y tenemos la condición.

Con esto hemos probado que la fórmula $\psi(y, x_1, \ldots, x_n)$ equivale a

$$\bigvee u \in \mathbb{N}(u > 0 \wedge \bigvee ct a_1 \cdots a_m \in \mathbb{N}(t > 0 \wedge c > 0 \wedge 1 + ct = \prod_{k=1}^{y}(1+kt)$$

$$\wedge\, t = Q(y, u, x_1, \ldots, x_n)!$$

$$\wedge\, (1+ct) \mid \prod_{j=0}^{u}(a_1 - j) \wedge \cdots \wedge (1+ct) \mid \prod_{j=0}^{u}(a_m - j)$$

$$\wedge\, P(y, c, x_1, \ldots, x_n, a_1, \ldots, a_m) \equiv 0 \pmod{1+ct})$$

$$\wedge \bigvee y_1, \ldots y_m \in \mathbb{N}(y_1, \ldots, y_m \leq u \wedge P(y, 0, x_1, \ldots, x_n, y_1, \ldots, y_m) = 0)).$$

Claramente esto equivale a su vez a

$$\bigvee uvct a_1 \ldots a_m e f g_1 \ldots g_m h_1 \ldots h_m i y_1 \ldots y_m \in \mathbb{N}(u > 0 \wedge t > 0 \wedge c > 0$$

$$\wedge\, v = u + 1 \wedge e = 1 + ct \wedge e = \prod_{k=1}^{y}(1+kt) \wedge f = Q(y, u, x_1, \ldots, x_n) \wedge t = f!$$

$$\wedge\, g_1 = a_1 - v \wedge \cdots \wedge g_m = a_m - v \wedge h_1 = \prod_{k=1}^{v}(g_1 + k) \wedge \cdots \wedge h_m = \prod_{k=1}^{v}(g_m + k)$$

$$\wedge\, e | h_1 \wedge \cdots \wedge e | h_m \wedge i = P(y, c, x_1, \ldots, x_n, a_1, \ldots, a_m) \wedge e \mid i$$

$$\wedge\, y_1, \ldots, y_m \leq u \wedge P(y, 0, x_1, \ldots, x_n, y_1, \ldots, y_m) = 0).$$

Los resultados del apartado anterior muestran que esta última expresión es diofántica. ∎

Ahora ya podemos probar que toda fórmula Σ_1 es diofántica. De hecho, por el teorema 7.31 basta probar que toda fórmula Δ_0 es diofántica. Si α es una fórmula Δ_0, existe una sucesión de fórmulas $\alpha_0, \ldots, \alpha_n \equiv \alpha$ según la definición 5.5. Vamos a probar inductivamente que tanto α_i como $\neg\alpha_i$ es diofántica.

Si $\alpha_i \equiv t_1 = t_2$, donde los t_i son términos sin descriptores, es fácil ver que existen polinomios P_i con coeficientes naturales de modo que (en $\mathrm{I}\Sigma_1$ se demuestra que) $t_i = P_i$, por lo que α_i es diofántica. En cuanto a $\neg\alpha_i$, es equivalente a

$$\bigvee x(t_1 = t_2 + x + 1 \vee t_2 = t_1 + x + 1),$$

luego también es diofántica, por la parte ya probada.

Si $\alpha_i \equiv t_1 \leq t_2$, entonces es equivalente a $\bigvee x(t_2 = t_1 + x)$, mientras que su negación es equivalente a $\bigvee x(t_1 = t_2 + x + 1)$, luego ambas son diofánticas.

Si $\alpha_i \equiv \neg\alpha_j$ con $j < k$, la conclusión es trivial. Si $\alpha_i \equiv (\alpha_j \to \alpha_k)$, con $j, k < i$, de modo que α_j y α_k son diofánticas por hipótesis de inducción, al igual que sus negaciones, entonces α_i y su negación son equivalentes a $\neg\alpha_j \vee \alpha_k$ y $\alpha_j \wedge \neg\alpha_k$, respectivamente, luego ambas son diofánticas.

Si $\alpha_i \equiv \bigvee x \le y\, \alpha_j$ con $j < i$, entonces α_i es diofántica (porque ya hemos visto que $x \le y$ lo es) y su negación, que es equivalente a $\bigwedge x \le y\, \neg\alpha_j$ es diofántica por el teorema anterior. Por último, si $\alpha_i \equiv \bigwedge x \le y\, \alpha_j$ razonamos análogamente. ∎

La ecuación de Pell Nos falta demostrar que la función x^y es diofántica. La prueba se basa en un estudio minucioso de las soluciones de una ecuación diofántica clásica: la ecuación de Pell. Se trata de la ecuación $x^2 - dy^2 = 1$, donde d es un número natural no cuadrado perfecto.

Las soluciones de esta ecuación están relacionadas con el anillo cuadrático $\mathbb{Z}[\sqrt{d}\,] = \{a + b\sqrt{d} \mid a, b \in \mathbb{Z}\}$. Al final del capítulo anterior hemos visto que es posible trabajar con este anillo en $\mathrm{I}\Sigma_1$.

La norma dada por $\mathrm{N}(a + b\sqrt{d}\,) = a^2 - db^2$ es multiplicativa (es decir, la norma de un producto es el producto de las normas). Llamaremos *unidades* de $\mathbb{Z}[\sqrt{d}\,]$ a los enteros cuadráticos $\alpha \in \mathbb{Z}[\sqrt{d}\,]$ de norma 1. De este modo, un par (a, b) es una solución de la ecuación de Pell si y sólo si $\alpha = a + b\sqrt{d} \in \mathbb{Z}[\sqrt{d}\,]$ es una unidad de $\mathbb{Z}[\sqrt{d}\,]$.

Como la norma es multiplicativa, el producto de unidades es una unidad, y el inverso de una unidad es también una unidad. La solución trivial $(1, 0)$ se corresponde con la unidad $1 \in \mathbb{Z}[\sqrt{d}\,]$. En lo que sigue nos restringiremos al caso particular en que $d = a^2 - 1$, con lo que otra unidad es $\epsilon = a + \sqrt{d}$, correspondiente a la solución $(a, 1)$. Vamos a probar que esta unidad genera a todas las demás y, por consiguiente, nos da todas las soluciones de la ecuación de Pell.

Teorema 7.39 *Con la notación anterior, las unidades de $\mathbb{Z}[\sqrt{d}\,]$ son exactamente las de la forma $\pm\epsilon^n$, donde $n \in \mathbb{Z}$.*

Demostración: Observemos que $1 < \epsilon$. La prueba se basa en que no existe ninguna unidad tal que $1 < \alpha < \epsilon$. Si la hubiera, sea $\alpha = x + y\sqrt{d}$. Tenemos que

$$1 = (x + y\sqrt{d}\,)(x - y\sqrt{d}) = (a + \sqrt{d}\,)(a - \sqrt{d}\,),$$

luego

$$\frac{x - y\sqrt{d}}{a - \sqrt{d}} = \frac{a + \sqrt{d}}{x + y\sqrt{d}} = \frac{\epsilon}{\alpha} > 1,$$

de donde $x - y\sqrt{d} > a - \sqrt{d}$ y $-x + y\sqrt{d} < -a + \sqrt{d}$. Por otra parte

$$x - y\sqrt{d} = \frac{1}{x + y\sqrt{d}} < 1,$$

luego $-1 < -x + y\sqrt{d}$. Tenemos, pues, que

$$-1 < -x + y\sqrt{d} < -a + \sqrt{d},$$
$$1 < x + y\sqrt{d} < a + \sqrt{d}.$$

Sumando miembro a miembro queda $0 < 2y\sqrt{d} < 2\sqrt{d}$, luego $0 < y < 1$, pero y es entero.

Sea ahora η cualquier unidad de $\mathbb{Z}[\sqrt{d}\,]$. Cambiando η por $-\eta$ podemos suponer $\eta > 0$ y cambiando η por η^{-1} podemos suponer que $\eta > 1$. La sucesión ϵ^n es monótona creciente y no está acotada, por lo que existe un n tal que $\epsilon^n \le \eta < \epsilon^{n+1}$. Por consiguiente $1 \le \eta\epsilon^{-n} < \epsilon$. Por lo que acabamos de probar $\eta\epsilon^{-n} = 1$, es decir, $\eta = \epsilon^n$. ∎

Como $\epsilon > 1$, para $n > 0$ se cumple $\epsilon^n > 1$, $0 < \epsilon^{-n} < 1$, $-\epsilon^n < -1$ y $-1 < -\epsilon^{-n} < 0$. Si (x, y) es una solución natural de la ecuación de Pell, entonces $x + y\sqrt{d} \ge 1$, luego ha de ser de la forma ϵ^n.

Definición 7.40 Si n y $a > 1$ son números naturales, definimos $x_n(a)$, $y_n(a)$ como los números naturales que cumplen $\epsilon^n = x_n(a) + y_n(a)\sqrt{d}$, donde $d = a^2 - 1$. Si no hay confusión omitiremos a. Notemos que $x_0 = 1$, $y_0 = 0$.

El hecho de que sean números naturales se prueba fácilmente por inducción a partir de las relaciones (7.1) más abajo. Los pares (x_n, y_n) son todas las soluciones naturales de la ecuación de Pell.

La fórmula $\epsilon^{m\pm n} = \epsilon^m\epsilon^{\pm n}$ se traduce inmediatamente en las relaciones

$$x_{m\pm n} = x_m x_n \pm d y_m y_n, \qquad y_{m\pm n} = x_n y_m \pm x_m y_n.$$

En particular

$$x_{m\pm 1} = a x_m \pm d y_m, \qquad y_{m\pm 1} = a y_m \pm x_m. \tag{7.1}$$

La prueba de que m^n es diofántica se basa en los siguientes resultados sobre las soluciones de la ecuación de Pell:

1. $(x_n, y_n) = 1$

   Demostración: Si $p \mid x_n$ y $p \mid y_n$, entonces $p \mid x_n^2 - d y_n^2 = 1$.

2. $y_m \mid y_n$ syss $m \mid n$.

   Demostración: Veamos por inducción sobre k que $y_m \mid y_{mk}$. Para $k = 0, 1$ es obvio.

   $$y_{m(k+1)} = x_m y_{mk} + x_{mk} y_m,$$

   luego si $y_m \mid y_{mk}$, también $y_m \mid y_{m(k+1)}$.

   Supongamos ahora que $y_m \mid y_n$ pero $m \nmid n$. Sea $n = mq + r$, con $0 < r < m$. Si $q = 0$, entonces $n < m$, luego $y_n < y_m$, pues (7.1) implica que la sucesión y_m es estrictamente creciente, luego $y_m \nmid y_n$. Por lo tanto $q > 0$. Entonces $y_n = x_r y_{mq} + x_{mq} y_r$. Por la parte ya probada $y_m \mid y_{mq}$, luego $y_m \mid x_{mq} y_r$. Ahora bien, $(x_{mq}, y_{mq}) = 1$, luego también $(x_{mq}, y_m) = 1$ y entonces ha de ser $y_m \mid y_r$, pero $y_r < y_m$, contradicción.

3. $y_{nk} \equiv k x_n^{k-1} y_n \pmod{y_n^3}$.

   DEMOSTRACIÓN:

   $$x_{nk} + y_{nk}\sqrt{d} = \epsilon^{nk} = (x_n + y_n\sqrt{d}\,)^k = \sum_{i=0}^{k} \binom{k}{i} x_n^{k-i} y_n^i d^{i/2},$$

   luego

   $$y_{nk} = \sum_{\substack{i=0\\ \text{impar}}}^{k} \binom{k}{i} x_n^{k-i} y_n^i d^{(i-1)/2},$$

   pero los sumandos con $i > 1$ son $\equiv 0 \pmod{y_n^3}$, luego tenemos la congruencia pedida.

4. $y_n^2 \mid y_{ny_n}$.

   DEMOSTRACIÓN: Se sigue inmediatamente de la propiedad anterior para $k = y_n$.

5. Si $y_n^2 \mid y_m$, entonces $y_n \mid m$.

   DEMOSTRACIÓN: Por 2 sabemos que $n \mid m$. Sea $m = nk$. Por 3 tenemos $y_m = k x_n^{k-1} y_n + r y_n^3$, luego $y_n^2 \mid k x_n^{k-1} y_n$. Por 1 ha de ser $y_n \mid k$ y, en consecuencia, $y_n \mid m$.

6. $x_{n+1} = 2a x_n - x_{n-1}, \quad y_{n+1} = 2a y_n - y_{n-1}$.

   Basta sumar las relaciones

   $$\begin{aligned} x_{n+1} &= a x_n + d y_n, & y_{n+1} &= a y_n + x_n,\\ x_{n-1} &= a x_n - d y_n, & y_{n-1} &= a y_n - x_m. \end{aligned}$$

7. $y_n \equiv n \pmod{a-1}$.

   DEMOSTRACIÓN: Se cumple para $y_0 = 0$ e $y_1 = 1$. Por inducción y la propiedad anterior:

   $$y_{n+1} = 2a y_n - y_{n-1} \equiv 2n - (n-1) \equiv n+1 \pmod{a-1}.$$

8. Si $a \equiv b \pmod{c}$, entonces

   $$x_n(a) \equiv x_n(b) \pmod{c}, \qquad y_n(a) \equiv y_n(b) \pmod{c}.$$

   DEMOSTRACIÓN: Para $n = 0$, 1 se da la igualdad. Por inducción:

   $$y_{n+1}(a) = 2a y_n(a) - y_{n-1}(a) \equiv 2b y_n(b) - y_{n-1}(b) = y_{n+1}(b) \pmod{c}.$$

   Con la x se razona análogamente.

9. $n \equiv y_n \pmod 2$.

   DEMOSTRACIÓN: Por inducción: $y_{n+1} = 2a y_n - y_{n-1} \equiv y_{n-1} \pmod 2$, luego si n es par $n \equiv y_0 = 0 \pmod 2$ y si n es impar $n \equiv y_1 = 1 \pmod 2$.

10. $x_n - y_n(a-y) \equiv y^n \pmod{2ay-y^2-1}$. (Ésta es la propiedad que conecta las soluciones de la ecuación de Pell con la exponencial.)

    DEMOSTRACIÓN: $x_0 - y_0(a-y) = 1$, $x_1 - y_1(a-y) = y$, luego se cumple para $n = 0, 1$. Si vale para n,

    $$x_{n+1} - y_{n+1}(a-y) = 2a(x_n - y_n(a-y)) - (x_{n-1} - y_{n-1}(a-y)) \equiv 2ay^n - y^{n-1}$$
    $$= y^{n-1}(2ay-1) \equiv y^{n-1}y^2 = y^{n+1} \pmod{2ay-y^2-1}.$$

11. $n \le y_n < y_{n+1}, \quad a^n \le x_n < x_{n+1}, \quad x_n \le (2a)^n$.

    DEMOSTRACIÓN: Las desigualdades sobre y_n se siguen inmediatamente de (7.1). Para las otras, usando (7.1) y 6:

    $$ax_n \le ax_n + dy_n = 2ax_n - (ax_n - dy_n) = 2ax_n - x_{n-1} = x_{n+1} \le 2ax_n.$$

    O sea, $ax_n \le x_{n+1} \le 2ax_n$. En particular $x_n < x_{n+1}$.

    Por inducción sobre n: $a^0 = 1 = x_0 = (2a)^0$. Si $a^n \le x_n \le (2a)^n$, entonces $a^{n+1} \le ax_n \le x_{n+1} \le 2ax_n \le (2a)^{n+1}$.

12. $x_{2n\pm j} \equiv -x_j \pmod{x_n}$.

    DEMOSTRACIÓN:

    $$x_{2n\pm j} = x_n x_{n\pm j} + dy_n y_{n\pm j} \equiv dy_n(y_n x_j \pm x_n y_j)$$
    $$\equiv dy_n^2 x_j = (x_n^2 - 1)x_j \equiv -x_j \pmod{x_n}.$$

13. $x_{4n\pm j} \equiv x_j \pmod{x_n}$.

    DEMOSTRACIÓN: Por el resultado anterior,

    $$x_{4n\pm j} = x_{2n+(2n\pm j)} \equiv -x_{2n\pm j} \equiv x_j \pmod{x_n}.$$

    Este argumento no vale si el signo es negativo y $2n < j \le 4n$, pero entonces $j = 2n + j'$, con $j' \le 2n$ y $x_{4n-j} = x_{2n-j'} \equiv -x_{j'} \equiv x_{2n+j'} = x_j$.

14. Si $x_i \equiv x_j \pmod{x_n}$ con $i \le j \le 2n$ y $n > 0$, entonces $i = j$ excepto si $a = 2$, $n = 1$, $i = 0$, $j = 2$.

    DEMOSTRACIÓN: Por 11 tenemos que $1 = x_0 < x_1 < \cdots < x_{n-1}$ y por 12 resulta que $x_{n+1}, x_{n+2}, \ldots, x_{2n-1}, x_{2n}$ son congruentes, respectivamente, con $-x_{n-1}, -x_{n-2}, \ldots - x_1, -x_0 = -1$ módulo x_n, luego vemos que $x_0, \ldots, x_{2n}$ son congruentes módulo x_n con

    $$-x_{n-1} < -x_{n-1} < \cdots < -x_1 < -x_0 < x_0 < x_1 < \cdots < x_{n-1} \qquad (7.2)$$

    Sea

    $$q = \begin{cases} (x_n - 1)/2 & \text{si } x_n \text{ es impar,} \\ x_n/2 & \text{si } x_n \text{ es par.} \end{cases}$$

En ambos casos, (7.1) nos da que

$$x_{n-1} \le x_n/a \le x_n/2 \le q,$$

luego los números (7.2) están comprendidos entre $-q$ y q. Si x_n es impar entonces $-q, \dots, q$ forman un sistema de restos módulo x_n, luego los números (7.2) son no congruentes dos a dos y el resultado está probado. Si x_n es par entonces un sistema de restos lo forman los números $-q+1, \dots, q$ y la conclusión es la misma salvo si $x_{n-1} = q$, en cuyo caso $i = n-1$, $j = n+1$ contradicen lo que queremos probar.

Por (7.1), si $ax_{n-1} + dy_{n-1} = x_n = 2q = 2x_{n-1}$, entonces $a = 2$, $y_{n-1} = 0$, $n = 1$, $i = 0$ y $j = 2$.

15. Si $x_i \equiv x_j \pmod{x_n}$ con $0 < i \le n$ y $0 \le j < 4n$, entonces $j = i$ o bien $j = 4n - i$.

    Demostración: Supongamos que $j \le 2n$. Entonces por el resultado anterior tenemos que $i = j$ salvo si $n = 1$, $i = 2$, $j = 0$, pero entonces $i > n$, lo cual es imposible.

    Supongamos que $j > 2n$. Sea $j' = 4n - j$. Así $0 < j' < 2n$ y por 13 se sigue cumpliendo $x_i \equiv x_j \equiv x_{j'} \pmod{x_n}$, y ahora concluimos que $i = j'$.

16. Si $0 < i \le n$ y $x_i \equiv x_j \pmod{x_n}$, entonces $j \equiv \pm i \pmod{4n}$.

    Demostración: Sea $j = 4nq + j'$ con $0 \le j' < 4n$. Por 13 tenemos que $x_j \equiv x_{4n(q-1)+j'} \pmod{x_n}$ y, tras q pasos, llegamos a que $x_j \equiv x_{j'} \pmod{x_n}$. Por el resultado anterior, o bien $i = j'$ o bien $i = 4n - j'$, luego $j \equiv j' \equiv \pm i \pmod{4n}$.

La función exponencial Finalmente estamos en condiciones de probar que la función exponencial $m = n^k$ es diofántica. Éstas son las ecuaciones que la caracterizan:

I $\quad x^2 - (a^2-1)y^2 = 1,$

II $\quad u^2 - (a^2-1)v^2 = 1,$

III $\quad s^2 - (b^2-1)t^2 = 1,$

IV $\quad v = ry^2,$

V $\quad b = 1 + 4(p+1)y = a + (q+1)u,$

VI $\quad s = x + cu,$

VII $\quad t = k + 4dy,$

VIII $\quad y = k + e,$

IX $\quad x = \overline{x} + 1,\ y = \overline{y} + 1,\ u = \overline{u} + 1,\ v = \overline{v} + 1,\ t = \overline{t} + 1,$

X $\quad (x - y(a-n) - m)^2 = f^2(2an - n^2 - 1)^2,$

XI $\quad m + g + 1 = 2an - n^2 - 1,$

XII $\quad w = n + h + 1 = k + l + 1,$

XIII $\quad a^2 - (w^2 - 1)(w - 1)^2(z + 1)^2 = 1.$

Lo probaremos en varios pasos. En primer lugar vemos que las primeras ecuaciones caracterizan a las soluciones de la ecuación de Pell:

Teorema 7.41 *Sean a, x, $k \in \mathbb{N}$, $a > 1$, $k > 0$. El sistema* I – IX *tiene solución natural para las demás variables si y sólo si $x = x_k(a)$.*

Demostración: Supongamos que el sistema tiene solución. Las referencias formadas por un único número arábigo corresponden a los resultados del apartado anterior.

Por V tenemos $b > a > 1$.

Por I, II, III y IX existen $i, j, n > 0$ tales que $x = x_i(a)$, $y = y_i(a)$, $u = x_n(a)$, $v = y_n(a)$, $s = x_j(b)$, $t = y_j(b)$.

Por IV, $y \leq v$, luego $i \leq n$.

Por V y VI, $b \equiv a \pmod{x_n(a)}$, $x_j(b) \equiv x_i(a) \pmod{x_n(a)}$.

Por 8, $x_j(b) \equiv x_j(a) \pmod{x_n(a)}$, luego $x_i(a) \equiv x_j(a) \pmod{x_n(a)}$.

Por 16,

$$j \equiv \pm i \pmod{4n}. \tag{7.3}$$

Por IV, $y_i(a)^2 \mid y_n(a)$, de donde, por 5, $y_i(a) \mid n$.

Por (7.3) tenemos

$$j \equiv \pm i \pmod{4y_i(a)}. \tag{7.4}$$

Por V, $b \equiv 1 \pmod{4y_i(a)}$ y por 7,

$$y_j(b) \equiv j \pmod{4y_i(a)}. \tag{7.5}$$

Por VII

$$y_j(b) \equiv k \pmod{4y_i(a)}. \tag{7.6}$$

Por (7.4), (7.5) y (7.6),

$$k \equiv \pm i \pmod{4y_i(a)}. \tag{7.7}$$

Por VIII, $k \leq y_i(a)$ y por 11, $i \leq y_i(a)$. Como los números $-2y+1, \ldots, 2y$ son un sistema de restos módulo $4y$, necesariamente $k = i$, y así, $x = x_i(a) = x_k(a)$.

Supongamos ahora que $x = x_k(a)$, $y = y_k(a)$. Entonces se cumple I.

Sea $m = 2ky_k(a) > 0$, $u = x_m(a) > 0$, $v = y_m(a) > 0$. Se cumple II.

Por 4 y 2, $y^2 = y_k(a)^2 \mid y_{ky_k(a)}(a) \mid y_m(a) = v$. Se cumple IV.

Por 9, $y_m(a) = v$ es par y por 1, $u = x_m(a)$ cumple $(u, v) = 1$. En particular es impar.

También $(u, 4y) = 1$, pues, como u es impar e $y \mid v$,

$$(u, 4y) = (u, y) \mid (u, v) = 1.$$

Por el teorema chino del resto existe un b_0 tal que

$$b_0 \equiv 1 \text{ (mód } 4y) \qquad \text{y} \qquad b_0 \equiv a \text{ (mód } u),$$

y cualquier $b = b_0 + 4juy$ cumple lo mismo. Se cumple V.

Sean $s = x_k(b)$, $t = y_k(b)$. Se cumplen III y IX.

Como $b > a$, se cumple $s = x_k(b) > x_k(a) = x$ (a partir de (7.1) se razona por inducción sobre $k > 0$ que $x_k(b) > k_k(a)$, $y_k(b) > y_k(a)$).

Por 8, $s \equiv x$ (mód u). Se cumple VI.

Por 11, $k \leq y_k(b) = t$ y por 7, $t \equiv k$ (mód $b - 1$).

Por V, $4y \mid b - 1$, luego $t \equiv k$ (mód $4y$). Se cumple VII.

Por 11, $k \leq y_k(a) = y$. Se cumple VIII. ∎

Para ocuparnos de las cuatro ecuaciones que faltan, necesitamos una observación sencilla:

Teorema 7.42 *Si $a > y^k$, $y > 0$, $k > 0$, entonces $2ay - y^2 - 1 > y^k$.*

Demostración: Basta observar que

$$2ay - y^2 - 1 = a^2 - 1 - (a - y)^2 \geq a^2 - 1 - (a - 1)^2 = 2a - 2 \geq a > y^k.$$

La primera desigualdad es porque $y \leq y^k < a$, y la segunda porque $a \geq 2$). ∎

Teorema 7.43 *Si m, n, k son naturales no nulos, entonces $m = n^k$ si y sólo si el sistema de ecuaciones* I–XIII *tiene solución natural en las demás variables.*

Demostración: Supongamos que el sistema tiene solución. Por XII $w > 1$, luego por XIII, $a > 1$. Por 7.41, las primeras ecuaciones implican que $x = x_k(a)$, $y = y_k(a)$.

Por 10, $x_k(a) - y_k(a)(a - n) \equiv n^k$ (mód $2an - n^2 - 1$), luego por X, tenemos que $m \equiv n^k$ (mód $2an - n^2 - 1$).

Por XII, k, $n < w$.

Por XIII, existe un $j > 0$ tal que $a = x_j(w)$, $(w - 1)(z + 1) = y_j(w)$.

Por 7, $j \equiv (w - 1)(z + 1) \equiv 0$ (mód $w - 1$). Por consiguiente $j \geq w - 1$.

Por 11, $a = x_j(w) \geq w^j \geq w^{w-1} > n^k$.

Por XI, $m < 2an - n^2 - 1$ y, por el teorema anterior, $n^k < 2an - n^2 - 1$.

Como m y n^k son congruentes y menores que el módulo, necesariamente $m = n^k$.

Supongamos ahora que $m = n^k$. Tomemos $w > n$, $w > k$. Se cumple XII.

Sea $a = x_{w-1}(w) > 1$. Por 7, $y_{w-1}(w) \equiv (w-1) \equiv 0$ (mód $w-1$), luego $y_{w-1}(w) = (w-1)(z+1)$, para un z. Se cumple XIII.

Por 11, $a = x_{w-1}(w) \geq w^{w-1} > n^k$, luego podemos aplicar el teorema anterior y $2an - n^2 - 1 > n^k = m$. Se cumple XI

Sean $x = x_k(a)$, $y = y_k(a)$. Por 10, $x - y(a-n) \equiv n^k$ (mód $2an - n^2 - 1$). Se cumple X. Las ecuaciones I–IX se cumplen por 7.41. ∎

Eliminar la restricción sobre que m, n y k sean no nulos es muy fácil:

$$m = n^k \leftrightarrow (m \neq 0 \wedge n \neq 0 \wedge k \neq 0 \wedge m = n^k)$$

$$\vee\, (m = 1 \wedge k = 0) \vee (m = 0 \wedge n = 0 \wedge k > 0)$$

y las tres fórmulas son diofánticas.

Capítulo VIII

La formalización de la lógica

Cuando introdujimos los lenguajes formales explicamos que no es imprescindible considerar que los signos de un lenguaje formal sean signos en el sentido propio de la palabra, es decir, "garabatos" trazables en un papel, sino que perfectamente pueden ser conceptos abstractos. En la práctica, es irrelevante que el implicador de $\mathcal{L}_a$ sea una flechita, un patito o, simplemente, "el implicador de $\mathcal{L}_a$", igual que el número 4 es simplemente "el número 4". En particular explicamos que podríamos considerar que los signos de un lenguaje formal $\mathcal{L}$ son números naturales.

Si lo vemos así, las fórmulas de $\mathcal{L}$ son sucesiones finitas de números naturales, y las demostraciones en una teoría axiomática sobre $\mathcal{L}$ son sucesiones finitas de sucesiones finitas de números naturales (que a su vez se pueden identificar con números naturales según explicamos en la sección 5.6), y esto plantea la posibilidad de formalizar el cálculo deductivo en AP. En este capítulo veremos que es posible hacerlo, de hecho, en $\mathrm{I}\Sigma_1$ y y en el siguiente veremos que esta posibilidad tiene consecuencias impactantes.

Si el lector es más tradicional y prefiere considerar que los signos de un lenguaje formal son "garabatos" trazables en un papel, lo único que tiene que hacer para seguir este capítulo es que asignar a cada uno de ellos (con el criterio que prefiera) un número natural distinto, al que tradicionalmente se le llama su *número de Gödel.* De este modo, cuando hablemos de un signo de $\mathcal{L}$ considerado como número natural, el lector puede entender que hablamos del número de Gödel de dicho signo. Similarmente, cuando consideremos a una cadena de signos de $\mathcal{L}$ como una sucesión finita de números naturales, el lector puede entender que hablamos de la sucesión de los números de Gödel de sus signos, que a su vez puede codificarse con un número natural. Esto nos permite considerar que las cadenas de signos de $\mathcal{L}$ (en particular sus expresiones) *son* números naturales, o bien que cada una de ellas *tiene asignado* un número natural, al que también se llama su número de Gödel. Similarmente, podemos considerar que cada sucesión de cadenas finitas de signos de $\mathcal{L}$ (en particular cada deducción lógica) *es* o *tiene asignado* un número natural, que, desde el segundo punto de vista, se llama también "número de Gödel de la deducción".

Como el lector podrá comprobar, considerar que los signos de un lenguaje formal son números naturales es simplemente una forma práctica e inofensiva de evitar cientos de veces la coletilla "el número de Gödel de" en todo cuanto digamos. Todo lo que sigue son definiciones y teoremas en $\mathrm{I}\Sigma_1$ (cuyas traducciones naturales son, por lo tanto, teoremas de KP y de Z^*).

La primera sección es puramente técnica. El lector puede plantearse la posibilidad de leerla superficialmente, sin preocuparse por los detalles, simplemente para convencerse de que $\mathrm{I}\Sigma_1$ es suficiente para formalizar los hechos que conocemos sobre los lenguajes formales y su gramática si consideramos que los signos, cadenas de signos y sucesiones de cadenas de signos son meros números naturales. Todos los detalles son irrelevantes. Lo mismo se aplica a la tercera sección, donde mostramos en en $\mathrm{I}\Sigma_1$ se puede formalizar la prueba del teorema de Σ_1-completitud de Q. Lo único que vamos a necesitar de dicha sección son las condiciones de Hilbert-Bernays que demostramos al final del capítulo. En cambio, el lector debería prestar especial atención a la sección 2, en la que estudiamos la relación entre una teoría metamatemática y su formalización en $\mathrm{I}\Sigma_1$, pues entender la situación es fundamental para entender los resultados del capítulo siguiente y muchos hechos más de la lógica matemática.

8.1 Lenguajes y teorías formales

Definición 8.1 Un *lenguaje formal* $\mathcal{L}$ está determinado por cinco designadores, que representaremos por $\ulcorner\neg\urcorner, \ulcorner\to\urcorner, \ulcorner\bigwedge\urcorner, \ulcorner|\urcorner, \ulcorner=\urcorner$ (de $\mathcal{L}_a$ o $\mathcal{L}_{\mathrm{tc}}$), cuatro fórmulas de tipo Δ_1, que representaremos[1] por $x \in \mathrm{Var}(\mathcal{L})$, $x \in \mathrm{Rel}(\mathcal{L})$, $x \in \mathrm{Fn}(\mathcal{L})$ y $\mathrm{Nar}(x,n)$, con las variables libres indicadas, de modo que en $\mathrm{I}\Sigma_1$ se demuestra[2] lo siguiente:

1. Cada signo de $\mathcal{L}$ pertenece a una única categoría, es decir:
$$\begin{aligned}&\ulcorner\neg\urcorner \neq \ulcorner\to\urcorner \wedge \ulcorner\neg\urcorner \neq \ulcorner\textstyle\bigwedge\urcorner \wedge \ulcorner\neg\urcorner \neq \ulcorner|\urcorner \wedge \ulcorner\to\urcorner \neq \ulcorner\textstyle\bigwedge\urcorner \wedge \ulcorner\to\urcorner \neq \ulcorner|\urcorner \wedge \ulcorner\textstyle\bigwedge\urcorner \neq \ulcorner|\urcorner\\ &\wedge \ulcorner\neg\urcorner \notin \mathrm{Var}(\mathcal{L}) \wedge \ulcorner\neg\urcorner \notin \mathrm{Rel}(\mathcal{L}) \wedge \ulcorner\neg\urcorner \notin \mathrm{Fn}(\mathcal{L})\\ &\wedge \ulcorner\to\urcorner \notin \mathrm{Var}(\mathcal{L}) \wedge \ulcorner\to\urcorner \notin \mathrm{Rel}(\mathcal{L}) \wedge \ulcorner\to\urcorner \notin \mathrm{Fn}(\mathcal{L})\\ &\wedge \ulcorner\textstyle\bigwedge\urcorner \notin \mathrm{Var}(\mathcal{L}) \wedge \ulcorner\textstyle\bigwedge\urcorner \notin \mathrm{Rel}(\mathcal{L}) \wedge \ulcorner\textstyle\bigwedge\urcorner \notin \mathrm{Fn}(\mathcal{L})\\ &\wedge \ulcorner|\urcorner \notin \mathrm{Var}(\mathcal{L}) \wedge \ulcorner|\urcorner \notin \mathrm{Rel}(\mathcal{L}) \wedge \ulcorner|\urcorner \notin \mathrm{Fn}(\mathcal{L})\\ &\textstyle\bigwedge x \in \mathrm{Var}(\mathcal{L})\ (x \notin \mathrm{Rel}(\mathcal{L}) \wedge x \notin \mathrm{Fn}(\mathcal{L})) \wedge \bigwedge x \in \mathrm{Rel}(\mathcal{L})\ x \notin \mathrm{Fn}(\mathcal{L}).\end{aligned}$$
2. $\textstyle\bigwedge x(x \in \mathrm{Rel}(\mathcal{L}) \vee x \in \mathrm{Fn}(\mathcal{L}) \to \bigvee^1 n\, \mathrm{Nar}(x,n))$.
3. $\ulcorner=\urcorner \in \mathrm{Rel}(\mathcal{L}) \wedge \mathrm{Nar}(\ulcorner=\urcorner, 2)$.
4. $\textstyle\bigwedge a(a \text{ es finito} \to \bigvee x \in \mathrm{Var}(\mathcal{L})\ x \notin a)$.

[1] Léanse: "x es una variable", "x es un relator", "x es un funtor" y "n" es el número de argumentos de x", respectivamente.

[2] Si quisiéramos trabajar en KP en lugar de en $\mathrm{I}\Sigma_1$, necesitaríamos exigir que los signos sean números naturales, es decir, $\ulcorner\neg\urcorner, \ulcorner\to\urcorner, \ulcorner\bigwedge\urcorner, \ulcorner|\urcorner, \ulcorner=\urcorner \in \mathbb{N}$, $\bigwedge x(x \in \mathrm{Var}(\mathcal{L}) \vee x \in \mathrm{Rel}(\mathcal{L}) \vee x \in \mathrm{Fn}(\mathcal{L}) \to x \in \mathbb{N})$, así como que, en 2. el n que cumple $\mathrm{Nar}(x,n)$ sea un número natural.

Usaremos los *ángulos de Quine* siempre que puedan darse confusiones entre los signos lógicos metamatemáticos y los signos lógicos en el sentido de la definición anterior. Por ejemplo, en lugar de escribir $=\neq\neg$ escribimos $\ulcorner=\urcorner \neq \ulcorner\neg\urcorner$ y así queda claro que se trata de la negación de la fórmula que resulta de igualar los designadores $\ulcorner=\urcorner$ y $\ulcorner\neg\urcorner$.

En estas condiciones, escribiremos $\mathrm{Nar}(x) \equiv n|\mathrm{Nar}(x,n)$, que es un término Δ_1, y escribiremos $x \in \mathrm{Const}(\mathcal{L}) \equiv x \in \mathrm{Fn}(\mathcal{L}) \wedge \mathrm{Nar}(x) = 0$, que es una fórmula Δ_1, y

$$x \in \mathrm{Rng}(\mathcal{L}) \equiv x = \ulcorner\neg\urcorner \vee x = \ulcorner\to\urcorner \vee x = \ulcorner\bigwedge\urcorner \vee x = \ulcorner|\urcorner$$

$$\vee\, x \in \mathrm{Var}(\mathcal{L}) \vee x \in \mathrm{Rel}(\mathcal{L}) \vee x \in \mathrm{Fn}(\mathcal{L}),$$

que también es una fórmula Δ_1.

Ejemplos (Comparar con la definición de la página 14) El lenguaje $\ulcorner\mathcal{L}_a\urcorner$ es el determinado por los designadores

$$\ulcorner\neg\urcorner \equiv 10, \quad \ulcorner\to\urcorner \equiv 11, \quad \ulcorner\bigwedge\urcorner \equiv 12, \quad \ulcorner|\urcorner \equiv 13, \quad \ulcorner=\urcorner \equiv 5,$$

y por las fórmulas Δ_1

$$x \in \mathrm{Var}(\mathcal{L}_a) \equiv \bigvee n \le x\ x = 2^n, \quad x \in \mathrm{Rel}(\mathcal{L}_a) \equiv x = 5,$$

$$x \in \mathrm{Fn}(\mathcal{L}_a) \equiv x = 3 \vee x = 6 \vee x = 7 \vee x = 9,$$

$$\mathrm{Nar}(x,n) \equiv (x = 3 \wedge n = 0) \vee (x = 6 \wedge n = 1)$$

$$\vee\, ((x = 5 \vee x = 7 \vee x = 9) \wedge n = 2).$$

Cuando trabajemos con este lenguaje formal escribiremos $\ulcorner 0\urcorner \equiv 3$, $\ulcorner S\urcorner \equiv 6$, $\ulcorner+\urcorner \equiv 7$ y $\ulcorner\cdot\urcorner \equiv 9$.

El lenguaje $\ulcorner\mathcal{L}_{\mathrm{tc}}\urcorner$ es el determinado por los designadores

$$\ulcorner\neg\urcorner \equiv 10, \quad \ulcorner\to\urcorner \equiv 11, \quad \ulcorner\bigwedge\urcorner \equiv 12, \quad \ulcorner|\urcorner \equiv 13, \quad \ulcorner=\urcorner \equiv 5,$$

y por las fórmulas Δ_1

$$x \in \mathrm{Var}(\ulcorner\mathcal{L}_{\mathrm{tc}}\urcorner) \equiv \bigvee n \le x\ x = 2^n,$$

$$x \in \mathrm{Rel}(\ulcorner\mathcal{L}_{\mathrm{tc}}\urcorner) \equiv x = 5 \vee x = 6, \qquad x \in \mathrm{Fn}_n(\ulcorner\mathcal{L}_{\mathrm{tc}}\urcorner) \equiv x \neq x,$$

$$\mathrm{Nar}(x,n) \equiv n = 2.$$

Cuando trabajemos con este lenguaje representaremos $\ulcorner\in\urcorner \equiv 6$. ∎

En lo que sigue suponemos fijado un lenguaje formal $\mathcal{L}$.

Definición 8.2 Llamaremos *cadenas de signos* de $\mathcal{L}$ a las sucesiones finitas[3] de signos de $\mathcal{L}$, es decir,[4]

$$s \in \mathrm{Cad}(\mathcal{L}) \equiv \bigwedge n(n < \ell(s) \to s_n \in \mathrm{Rng}(\mathcal{L})).$$

Similarmente,

$$s \in \mathrm{SucCad}(\mathcal{L}) \equiv \bigwedge n(n < \ell(s) \to s_n \in \mathrm{Cad}(\mathcal{L})).$$

Usando el teorema 5.57, podemos definir un término Δ_1 que cumple:

$$\bigwedge s \in \mathrm{SucCad}(\mathcal{L})(\mathrm{Cnct}(s,0) = \varnothing \wedge \bigwedge i < \ell(s)\ \mathrm{Cnct}(s, i+1) = \mathrm{Cnct}(s,i)^\frown s_i).$$

Definimos $\mathrm{Concat}(s) \equiv \mathrm{Cnct}(s, \ell(s))$, que también es un término Δ_1. Cuando $s \in \mathrm{SucCad}(\mathcal{L})$ y $\ell(s) = n+1$, escribiremos $s_0 \cdots s_n \equiv \mathrm{Concat}(s)$, que es la sucesión que resulta de concatenar los términos de s vistos como sucesiones.

Abreviaremos $Ts \equiv s_0 \in \mathrm{Var}(\mathcal{L}) \vee s_0 \in \mathrm{Fn}(\mathcal{L}) \vee s_0 = \ulcorner\urcorner$ y

$$Fs \equiv s_0 \in \mathrm{Rel}(\mathcal{L}) \vee s_0 = \ulcorner\neg\urcorner \vee s_0 = \ulcorner\to\urcorner \vee s_0 = \ulcorner\bigwedge\urcorner,$$

que son dos fórmulas Δ_1.

Diremos que s es una *sucesión de expresiones* de $\mathcal{L}$ si cumple

$$\mathrm{sucexp}(s) \equiv s \in \mathrm{SucCad}(\mathcal{L}) \wedge \bigwedge i < \ell(s)(\cdots),$$

donde los puntos suspensivos representan la disyunción de las fórmulas siguientes:

1. $\bigvee x(x \in \mathrm{Var}(\mathcal{L}) \wedge s_i = \langle x \rangle)$,
2. $\bigvee c(c \in \mathrm{Const}(\mathcal{L}) \wedge s_i = \langle c \rangle)$,
3. $\bigvee tf(t \in \mathrm{SucCad}(\mathcal{L}) \wedge \bigwedge j < \ell(t) \bigvee k < i\ (Ts_k \wedge t_i = s_k)$
   $\wedge\ f \in \mathrm{Fn}(\mathcal{L}) \wedge \mathrm{Nar}(f) = \ell(t) \wedge s_i = \langle f \rangle^\frown \mathrm{Concat}(t))$,
4. $\bigvee tR(t \in \mathrm{SucCad}(\mathcal{L}) \wedge \bigwedge j < \ell(t) \bigvee k < i\ (Ts_k \wedge t_i = s_k)$
   $\wedge\ R \in \mathrm{Rel}(\mathcal{L}) \wedge \mathrm{Nar}(f) = \ell(t) \wedge s_i = \langle R \rangle^\frown \mathrm{Concat}(t))$,
5. $\bigvee k < i(Fs_k \wedge s_i = \ulcorner\neg\urcorner^\frown s_k)$,
6. $\bigvee kl < i(Fs_k \wedge Fs_l \wedge s_i = \ulcorner\to\urcorner^\frown s_k{}^\frown s_l)$,
7. $\bigvee x \bigvee k < i(x \in \mathrm{Var}(\mathcal{L}) \wedge Fs_k \wedge s_i = \ulcorner\bigwedge\urcorner^\frown \langle x \rangle^\frown s_k)$,
8. $\bigvee x \bigvee k < i(x \in \mathrm{Var}(\mathcal{L}) \wedge Fs_k \wedge s_i = \ulcorner\urcorner^\frown \langle x \rangle^\frown s_k)$.

[3] Véase la sección 5.6.

[4] Se trata de una fórmula Δ_1, pues podemos acotar el cuantificador $\bigwedge n < s$.

Diremos que θ es una *expresión* de $\mathcal{L}$ si cumple:

$$\theta \in \mathrm{Exp}(\mathcal{L}) \equiv \bigvee s \bigvee m(\mathrm{sucexp}(s) \wedge \ell(s) = m+1 \wedge s_m = \theta).$$

Es inmediato que si s es una sucesión de expresiones y $m \leq \ell(\theta)$, entonces la restricción $s|_{I_m}$ también es una sucesión de expresiones, lo cual significa a su vez que que todas las cadenas s_i son expresiones de $\mathcal{L}$.

En principio, la fórmula $\mathrm{sucexp}(s)$ es Σ_1, pero es fácil ver que de hecho es Δ_1. Sólo hemos de comprobar que las variables que hemos dejado sin acotar en la definición pueden acotarse. Claramente, todas se acotan por s salvo a lo sumo la t de 3. y 4. Para acotar esta t definiremos algunos conceptos que usaremos a menudo en acotaciones:

Definición 8.3 El teorema 5.57 nos permite definir

$$\mathrm{Rng}(s,0) = \varnothing \wedge \bigwedge n\, \mathrm{Rng}(s,n+1) = \mathrm{Rng}(s,n) \cup \{s_n\},$$

donde usamos que la fórmula $\psi(s,n,x,y) \equiv y = x \cup \{s_n\}$ es Σ_1. A su vez, si definimos

$$\mathrm{Rng}(s) = \mathrm{Rng}(s, \ell(s)+1),$$

tenemos que $\mathrm{Rng}(s)$ es Δ_1 y es el conjunto de todos los términos de la sucesión s. Similarmente, usando

$$\psi(a,n,x,y) \equiv \bigwedge u \in y(\ell(u) = n+1 \wedge u|_n \in x \wedge u_n \in a) \wedge$$
$$\bigwedge u \in x \bigwedge v \in a(u^\frown \langle v\rangle \in y),$$

podemos definir un término a^n de tipo Δ_1 que cumple

$$a^0 = \langle 0\rangle \wedge \bigwedge n(\bigwedge u \in a^{n+1}(\ell(u) = n+1 \wedge u|_n \in a^n \wedge u_n \in a) \wedge$$
$$\bigwedge u \in a^n \wedge v \in a(u^\frown \langle v\rangle \in a^{n+1})),$$

con lo que a^n es el conjunto de todas las sucesiones de longitud n cuyos términos están en a. A su vez, definimos

$$a^{<0} = 0 \wedge \bigwedge n\, a^{<n+1} = a^{<n} \cup a^n, \qquad a^{\leq n} \equiv a^{<n+1}.$$

De este modo $a^{\leq n}$ es el conjunto de todas las sucesiones de longitud $\leq n$ cuyos términos están en a. Finalmente, definimos el término Δ_1

$$\mathrm{Sub}(s) = \mathrm{Rng}(s)^{\leq \ell(s)},$$

que es el conjunto de todas las sucesiones de longitud $\leq \ell(s)$ de términos de la sucesión s.

Volviendo a la definición de $\mathrm{sucexp}(s)$, ahora ya podemos acotar $t \in \mathrm{Sub}(s)$. Más precisamente, tenemos que

$$\mathrm{sucexp}(s) \leftrightarrow \bigvee z(z = \mathrm{Sub}(s) \wedge \mathrm{sucexp}(s)^*) \leftrightarrow \bigwedge z(z = \mathrm{Sub}(s) \to \mathrm{sucexp}(s)^*),$$

donde $\mathrm{subexp}(s)^*$ es la definición que hemos dado de $\mathrm{subexp}(s)$, pero acotando por s las variables que hemos dejado sin acotar, excepto la t de 3. y 4, que las acotamos por z. Esto prueba que la fórmula es Δ_1.

Veamos ahora que la fórmula $\theta \in \mathrm{Exp}(\mathcal{L})$ también es Δ_1. Para ello observamos que si una sucesión de expresiones s define a θ (es decir, tiene a θ como última expresión), entonces la sucesión que resulta de eliminar todos los términos de s que no son subsucesiones de θ, así como las expresiones repetidas (dejando sólo la que aparezca antes en la sucesión) es también una sucesión de expresiones que define a θ. En otras palabras, que en la definición de expresión no perdemos generalidad si suponemos que la sucesión s no tiene repeticiones y que todos sus elementos son subsucesiones de θ. Observamos entonces que el número de subsucesiones de θ de longitud $i \leq \ell(\theta)$ será a lo sumo $\ell(\theta)$, luego el número total de subsucesiones de $\ell(\theta)$ no será superior a $\ell(\theta)^2$. Por consiguiente

$$s \in (\mathrm{Rng}(\theta)^{\leq \ell(\theta)})^{\leq \ell(\theta)^2}.$$

Por consiguiente,

$$\theta \in \mathrm{Exp}(\mathcal{L}) \leftrightarrow \bigwedge kz(k = \ell(\theta)^2 \wedge z = (\mathrm{Rng}(\theta)^{\leq \ell(\theta)})^{\leq \ell(\theta)^2} \rightarrow$$
$$\bigvee s \in z \bigvee m < k(\mathrm{sucexp}(s) \wedge \ell(s) = m + 1 \wedge s_m = \theta)),$$

luego la fórmula es Π_1 y, por lo tanto, Δ_1.

Definición 8.4 Definimos los términos y las fórmulas de un lenguaje formal mediante las fórmulas siguientes (de tipo Δ_1):

$\theta \in \mathrm{Term}(\mathcal{L}) \equiv \theta \in \mathrm{Exp}(\mathcal{L}) \wedge T\theta$,
$\theta \in \mathrm{Form}(\mathcal{L}) \equiv \theta \in \mathrm{Exp}(\mathcal{L}) \wedge F\theta$.

A partir de aquí utilizaremos las versiones formales de los convenios de notación sobre expresiones que venimos usando a nivel metamatemático. Por ejemplo, en lugar de escribir $\langle \rightarrow \rangle^\frown \alpha^\frown \beta$ escribiremos $\alpha \rightarrow \beta$, en lugar de $\langle \rightarrow, \neg \rangle^\frown \alpha^\frown \beta$ escribiremos $\alpha \vee \beta$, etc. También evitaremos la distinción entre un signo x y la cadena de signos $\langle x \rangle$ de longitud 1 cuando el contexto deje claro a cuál nos estamos refiriendo.

Ahora es inmediato que las condiciones de la definición de sucesión de expresiones equivalen a las siguientes:

1. $\bigvee x(x \in \mathrm{Var}(\mathcal{L}) \wedge s_i = x)$,
2. $\bigvee c(c \in \mathrm{Const}(\mathcal{L}) \wedge s_i = c)$,
3. $\bigvee tfn(t \in \mathrm{SucCad}(\mathcal{L}) \wedge \ell(t) = n + 1 \wedge f \in \mathrm{Fn}(\mathcal{L}) \wedge \mathrm{Nar}(f) = n + 1$
   $\wedge \bigwedge j \leq n \bigvee k < i\ (s_k \in \mathrm{Term}(\mathcal{L}) \wedge t_j = s_k) \wedge s_i = f t_0 \cdots t_n)$,
4. $\bigvee tRn(t \in \mathrm{SucCad}(\mathcal{L}) \wedge \ell(t) = n + 1 \wedge R \in \mathrm{Rel}(\mathcal{L}) \wedge \mathrm{Nar}(f) = n + 1$
   $\wedge \bigwedge j \leq n \bigvee k < i\ (s_k \in \mathrm{Term}(\mathcal{L}) \wedge t_j = s_k) \wedge s_i = R t_0 \cdots t_n)$,
5. $\bigvee k < i(s_k \in \mathrm{Form}(\mathcal{L}) \wedge s_i = \neg s_k)$,
6. $\bigvee kl < i(s_k, s_l \in \mathrm{Form}(\mathcal{L}) \wedge s_i = s_k \rightarrow s_l)$,
7. $\bigvee x \bigvee k < i(x \in \mathrm{Var}(\mathcal{L}) \wedge s_k \in \mathrm{Form}(\mathcal{L}) \wedge s_i = \bigwedge x\, s_k)$,
8. $\bigvee x \bigvee k < i(x \in \mathrm{Var}(\mathcal{L}) \wedge s_k \in \mathrm{Form}(\mathcal{L}) \wedge s_i = x | s_k)$.

Es fácil ver entonces que todo término de $\mathcal{L}$ es de la forma x (una variable), c (una constante), una sucesión de la forma $ft_0\cdots t_n$, donde f es un funtor $n+1$-ádico y cada t_i es un término o bien $x|\alpha$, donde x es una variable y α una fórmula. Similarmente, toda fórmula de $\mathcal{L}$ es de la forma $Rt_0\cdots t_n$, donde R es un relator $n+1$-ádico y cada t_i es un término, o bien $\neg\alpha$, $\alpha\to\beta$ o $\bigwedge x\alpha$, donde α y β son fórmulas. Además, toda sucesión construida de esta forma es un término o una fórmula.

El principio de Σ_1-inducción se particulariza a un principio de inducción sobre expresiones:

Teorema 8.5 *Sea $\phi(x)$ una fórmula Σ_1 tal vez con más variables libres. Si suponemos lo siguiente:*

1. $\bigwedge x \in \mathrm{Var}(\mathcal{L})\ \phi(x)$,
2. $\bigwedge c \in \mathrm{Const}(\mathcal{L})\ \phi(c)$,
3. $\bigwedge fnt(f \in \mathrm{Fn}(\mathcal{L}) \wedge \mathrm{Nar}(f) = n+1 \wedge \ell(t) = n+1 \wedge$
   $\bigwedge i \le n\ (t_i \in \mathrm{Term}(\mathcal{L}) \wedge \phi(t_i)) \to \phi(ft_0\cdots t_n))$,
4. $\bigwedge Rnt(f \in \mathrm{Rel}(\mathcal{L}) \wedge \mathrm{Nar}(R) = n+1 \wedge \ell(t) = n+1 \wedge$
   $\bigwedge i \le n\ (t_i \in \mathrm{Term}(\mathcal{L}) \wedge \phi(t_i)) \to \phi(Rt_0\cdots t_n))$,
5. $\bigwedge \alpha(\alpha \in \mathrm{Form}(\mathcal{L}) \wedge \phi(\alpha) \to \phi(\neg\alpha))$,
6. $\bigwedge \alpha\beta(\alpha,\beta \in \mathrm{Form}(\mathcal{L}) \wedge \phi(\alpha) \wedge \phi(\beta) \to \phi(\alpha\to\beta))$,
7. $\bigwedge x\alpha(x \in \mathrm{Var}(\mathcal{L}) \wedge \alpha \in \mathrm{Form}(\mathcal{L}) \wedge \phi(\alpha) \to \phi(\bigwedge x\alpha))$,
8. $\bigwedge x\alpha(x \in \mathrm{Var}(\mathcal{L}) \wedge \alpha \in \mathrm{Form}(\mathcal{L}) \wedge \phi(\alpha) \to \phi(x|\alpha))$,

entonces $\bigwedge \theta \in \mathrm{Exp}(\mathcal{L})\, \phi(\theta)$.

Demostración: Dada una expresión cualquiera θ, basta tomar una sucesión de expresiones s que defina a θ y probar $\bigwedge i \le \ell(\theta)\, \phi(s_i)$ por inducción sobre i. ∎

También podemos definir funciones recurrentemente sobre expresiones:

Teorema 8.6 *Dadas funciones $F_v(x)$, $F_c(x)$, $F_{\mathrm{fn}}(f,t,t')$, $F_{\mathrm{rl}}(R,t,t')$, $F_{\mathrm{ng}}(\alpha,y)$, $F_{\mathrm{imp}}(\alpha,\beta,y,z)$, $F_{\mathrm{gen}}(x,\alpha,y)$, $F_{\mathrm{desc}}(x,\alpha,y)$ de tipo Σ_1, existe una función (también Σ_1) $F : \mathrm{Exp}(\mathcal{L}) \longrightarrow V$ tal que*

1. $\bigwedge x \in \mathrm{Var}(\mathcal{L})\, F(x) = F_v(x)$,
2. $\bigwedge c \in \mathrm{Const}(\mathcal{L})\, F(c) = F_c(c)$,
3. $\bigwedge f \in \mathrm{Fn}(\mathcal{L}) \bigwedge n \bigwedge tt'(\mathrm{Nar}(f) = n+1 \wedge \ell(t) = n+1 \wedge \ell(t') = n+1 \wedge$
   $\bigwedge i < n\ (t_i \in \mathrm{Term}(\mathcal{L}) \wedge t'_i = F(t_i)) \wedge F(ft_0\ldots t_n) = F_{\mathrm{fn}}(f,t,t'))$,

4. $\bigwedge R \in \mathrm{Rel}(\mathcal{L})\bigwedge n\bigwedge tt'(\mathrm{Nar}(f) = n+1 \wedge \ell(t) = n+1 \wedge \ell(t') = n+1 \wedge$
   $\bigwedge i < n\ (t_i \in \mathrm{Term}(\mathcal{L}) \wedge t'_i = F(t_i)) \wedge F(Rt_0 \dots t_n) = F_{\mathrm{rl}}(R,t,t'))$,
5. $\bigwedge \alpha \in \mathrm{Form}(\mathcal{L})\ F(\neg\alpha) = F_{\mathrm{ng}}(\alpha, F(\alpha))$,
6. $\bigwedge \alpha\beta \in \mathrm{Form}(\mathcal{L})\ F(\alpha \to \beta) = F_{\mathrm{imp}}(\alpha,\beta,F(\alpha),F(\beta))$,
7. $\bigwedge \alpha \in \mathrm{Form}(\mathcal{L})\bigwedge x \in \mathrm{Var}(\mathcal{L})\ F(\bigwedge x\alpha) = F_{\mathrm{gen}}(x,\alpha,F(\alpha))$,
8. $\bigwedge \alpha \in \mathrm{Form}(\mathcal{L})\bigwedge x \in \mathrm{Var}(\mathcal{L})\ F(x|\alpha) = F_{\mathrm{desc}}(x,\alpha,F(\alpha))$.

DEMOSTRACIÓN: Basta tomar como F la función definida por la fórmula

$$\phi(\theta,x) \equiv \bigvee ss'\bigvee m(\mathrm{sucexp}(s) \wedge \ell(s) = m+1 \wedge \ell(s') = m+1 \wedge$$
$$s_m = \theta \wedge y = s'_m \wedge \bigwedge i \le m \cdots),$$

donde los puntos suspensivos son la disyunción de las fórmulas siguientes:

1. $\bigvee x(x \in \mathrm{Var}(\mathcal{L}) \wedge s_i = x \wedge s'_i = F_v(x))$,
2. $\bigvee c(c \in \mathrm{Const}(\mathcal{L}) \wedge s_i = c \wedge s'_i = F_c(c))$,
3. $\bigvee tt'fn(t \in \mathrm{SucCad}(\mathcal{L}) \wedge \ell(t) = n+1 \wedge \ell(t') = n+1$
   $\wedge\ f \in \mathrm{Fn}(\mathcal{L}) \wedge \mathrm{Nar}(f) = n+1 \wedge$
   $\bigwedge j \le n \bigvee k < i\ (s_k \in \mathrm{Term}(\mathcal{L}) \wedge t_j = s_k \wedge t'_j = s'_k) \wedge s_i = ft_0 \cdots t_n$
   $\wedge\ s'_i = F_{\mathrm{fn}}(f,t,t'))$,
4. $\bigvee tt'Rn(t \in \mathrm{SucCad}(\mathcal{L}) \wedge \ell(t) = n+1 \wedge \ell(t') = n+1$
   $\wedge\ f \in \mathrm{Rel}(\mathcal{L}) \wedge \mathrm{Nar}(f) = n+1 \wedge$
   $\bigwedge j \le n \bigvee k < i\ (s_k \in \mathrm{Term}(\mathcal{L}) \wedge t_j = s_k \wedge t'_j = s'_k) \wedge s_i = Rt_0 \cdots t_n$
   $\wedge\ s'_i = F_{\mathrm{rl}}(R,t,t'))$,
5. $\bigvee k < i(s_k \in \mathrm{Form}(\mathcal{L}) \wedge s_i = \neg s_k \wedge s'_i = F_{\mathrm{ng}}(s_k, s'_k))$,
6. $\bigvee kl < i(s_k, s_l \in \mathrm{Form}(\mathcal{L}) \wedge s_i = s_k \to s_l \wedge s'_k = F_{\mathrm{imp}}(s_k, s_l, s'_k, s'_l))$,
7. $\bigvee x\bigvee k < i(x \in \mathrm{Var}(\mathcal{L}) \wedge s_k \in \mathrm{Form}(\mathcal{L}) \wedge s_i = \bigwedge x\, s_k$
   $\wedge\ s'_i = F_{\mathrm{gen}}(x, s_k, s'_k))$,
8. $\bigvee x\bigvee k < i(x \in \mathrm{Var}(\mathcal{L}) \wedge s_k \in \mathrm{Form}(\mathcal{L}) \wedge s_i = x|s_k$
   $\wedge\ s'_i = F_{\mathrm{desc}}(x, s_k, s'_k))$.

Claramente la fórmula es Σ_1. Para probar que $\bigwedge \theta \in \mathrm{Exp}(\mathcal{L})\bigvee y\ \phi(\theta,y)$ basta ver que si s es una sucesión de expresiones que define a θ y $\ell(s) = m+1$, entonces

$$\bigwedge j \le m+1 \bigvee s'(\ell(s') = j \wedge \bigwedge i < j \cdots),$$

donde los puntos suspensivos son la disyunción de las fórmulas 1–8 de la definición de ϕ. Como la fórmula tras el $\bigwedge j$ es Σ_1, podemos razonar por inducción sobre j. Para $j = 0$ es trivial. Supuesto cierto para j, sea s' la sucesión correspondiente y distinguimos casos según la condición de la definición de sucesión de expresiones que cumpla s_j:

1. Si s_j es una variable, tomamos $s'' = s'^\frown \langle F_v(s_j)\rangle$,
2. Si s_j es una constante, tomamos $s'' = s'^\frown \langle F_c(s_j))\rangle$,
3. Si $s_j = ft_0 \cdots t_n$, donde t cumple lo requerido por la definición de sucesión de expresiones, definimos por recurrencia otra sucesión t' tal que

$$\bigwedge l \leq n \bigwedge k < j (t_i = s_k \leftrightarrow t'_i = s'_k),$$

   y entonces definimos $s'' = s'^\frown \langle F_{\text{fn}}(f, t, t')\rangle$.
4. Si $s_j = Rt_0 \cdots t_n$ definimos s'' de forma análoga,
5. Si $s_j = \neg s_k$, para cierto $k < j$, definimos $s'' = s'^\frown \langle F_{\text{neg}}(s_k, s'_k)\rangle$, y los casos restantes son similares a éste.

Una comprobación rutinaria muestra que s'' cumple lo requerido, y aplicando esto a $j = m + 1$ obtenemos un s' tal que $y = s'_m$ cumple $\phi(\theta, y)$.

Para probar que $\bigwedge \theta \in \text{Exp}(\mathcal{L}) \overset{1}{\bigvee} y\ \phi(\theta, y)$ basta probar que si s y $\bar{s}$ son sucesiones de expresiones de longitudes $m + 1$ y $\bar{m} + 1$ respectivamente, que definen a θ y s' y $\bar{s}'$ cumplen la definición de ϕ, entonces $s'_m = \bar{s}'_{\bar{m}}$. A su vez basta probar que

$$\bigwedge i \leq m \bigwedge i' \leq \bar{m} (s_i = \bar{s}_{i'} \to s'_i = \bar{s}'_{i'}).$$

Como la fórmula tras $\bigwedge i$ es Δ_1, podemos razonar por inducción sobre i. Suponemos que el resultado es cierto para todo $j < i$ y que $s_i = \bar{s}_{i'}$. Nuevamente hemos de distinguir casos según qué apartado de la definición de sucesión de expresiones cumple s_i. Consideraremos únicamente el caso c), pues los demás son más sencillos. Tenemos entonces que $s_i = \bar{s}_{i'} = ft_0 \cdots t_n$.

Para cada $j \leq n$, tenemos que $t_j = s_k$, para cierto $k < i$ y también $t_j = \bar{s}_{k'}$, para cierto $k' < i'$. Por hipótesis de inducción $s'_k = \bar{s}'_{k'}$, luego la sucesión t' que cumple la definición de ϕ es la misma en los dos casos. Por lo tanto, $s'_i = F_{\text{fn}}(f, t, t') = \bar{s}'_i$.

Por último, es inmediato que si s y s' cumplen la definición de ϕ, entonces $\bigwedge i \leq m\ s'_i = F(s_i)$, de donde a su vez se sigue que F cumple lo pedido. ∎

Variables libres Como primera aplicación, el teorema anterior nos permite definir el conjunto de las variables libres en una expresión:

Definición 8.7 Llamaremos $\text{Vlib}(\theta)$ (el conjunto de las *variables libres* en θ) al término Δ_1 dado por el teorema anterior de modo que cumpla:

1. $\text{Vlib}(x) = \{x\}$,
2. $\text{Vlib}(c) = \varnothing$,
3. $\text{Vlib}(ft_0 \cdots t_n) = \text{Vlib}(t_0) \cup \cdots \cup \text{Vlib}(t_n)$,

4. $\mathrm{Vlib}(Rt_0\cdots t_n) = \mathrm{Vlib}(t_0)\cup\cdots\cup\mathrm{Vlib}(t_n)$,
5. $\mathrm{Vlib}(\neg\alpha) = \mathrm{Vlib}(\alpha)$,
6. $\mathrm{Vlib}(\alpha\to\beta) = \mathrm{Vlib}(\alpha)\cup\mathrm{Vlib}(\beta)$,
7. $\mathrm{Vlib}(\bigwedge x\alpha) = \mathrm{Vlib}(\alpha)\setminus\{x\}$,
8. $\mathrm{Vlib}(x|\alpha) = \mathrm{Vlib}(\alpha)\setminus\{x\}$.

Por ejemplo, el caso 3) está definido mediante $F_{\mathrm{fn}}(f,t,t') = \bigcup\mathrm{Rng}(t')$, que es claramente Δ_1, e igualmente se comprueban los otros casos.

El conjunto $\mathrm{Vlig}(\theta)$ de *variables ligadas* en θ se define igualmente, cambiando tan sólo $\mathrm{Vlig}(x) = \varnothing$, $\mathrm{Vlig}(\bigwedge x\alpha) = \mathrm{Vlig}(\alpha)\cup\{x\}$, $\mathrm{Vlig}(x|\alpha) = \mathrm{Vlig}(\alpha)\cup\{x\}$.

También podemos definir el conjunto $V(\theta)$ de variables de θ, aunque en este caso no necesitamos una definición recurrente:

$$V(\theta) = \{x\in\mathrm{Rng}(\theta)\mid x\in\mathrm{Var}(\mathcal{L})\}.$$

Una simple inducción (aplicando 8.5) prueba que $V(\theta) = \mathrm{Vlib}(\theta)\cup\mathrm{Vlib}(\theta)$. Como $V(\theta)$ es obviamente un conjunto finito, lo mismo vale para los otros dos.

Sustitución En la definición de lenguaje formal hemos exigido que las variables de $\mathcal{L}$ no sean un conjunto finito, por lo que podemos definir

$$\mathrm{MV}(\theta) \equiv x|(x\in\mathrm{Var}(\mathcal{L})\wedge x\notin V(\theta)\wedge\bigwedge y(y\in\mathrm{Var}(\mathcal{L})\wedge y<x)\to y\in V(\theta))),$$

es decir, $\mathrm{MV}(\theta)$ es la menor variable que no está en θ.

Teorema 8.8 *Existe un término $\mathtt{S}^t_x\theta$ de tipo Δ_1 para el que se demuestran las propiedades siguientes:*

1. $\mathtt{S}^t_x y = \begin{cases} t & si\ y = x,\\ y & si\ y\neq x,\end{cases}$
2. $\mathtt{S}^t_x c = c$,
3. $\mathtt{S}^t_x f t_0\cdots t_n = f\mathtt{S}^t_x t_0\cdots\mathtt{S}^t_x t_n$,
4. $\mathtt{S}^t_x Rt_0\cdots t_n = R\mathtt{S}^t_x t_0\cdots\mathtt{S}^t_x t_n$,
5. $\mathtt{S}^t_x\neg\alpha = \neg\mathtt{S}^t_x\alpha$,
6. $\mathtt{S}^t_x(\alpha\to\beta) = \mathtt{S}^t_x\alpha\to\mathtt{S}^t_x\beta$,
7. $\mathtt{S}^t_x\bigwedge y\alpha = \begin{cases} \bigwedge y\alpha & si\ x\notin\mathrm{Vlib}(\bigwedge y\alpha),\\ \bigwedge y\,\mathtt{S}^t_x\alpha & si\ x\in\mathrm{Vlib}(\bigwedge y\alpha)\wedge y\notin\mathrm{Vlib}(t),\\ \bigwedge z\,\mathtt{S}^t_x\mathtt{S}^z_y\alpha & si\ x\in\mathrm{Vlib}(\bigwedge y\alpha)\wedge y\in\mathrm{Vlib}(t)\\ & \wedge\, z = \mathrm{MV}(\bigwedge y\alpha\wedge y = t),\end{cases}$

$$8.\ \mathtt{S}^t_x y|\alpha = \begin{cases} y|\alpha & si\ x \notin \mathrm{Vlib}(y|\alpha), \\ y|\mathtt{S}^t_x\alpha & si\ x \in \mathrm{Vlib}(y|\alpha) \wedge y \notin \mathrm{Vlib}(t), \\ z|\mathtt{S}^t_x\mathtt{S}^z_y\alpha & si\ x \in \mathrm{Vlib}(y|\alpha) \wedge y \in \mathrm{Vlib}(t) \\ & \wedge\ z = \mathrm{MV}(\bigwedge y\alpha \wedge y = t). \end{cases}$$

Demostración: Observemos que no podemos aplicar el teorema 8.6 porque el apartado 7) no define $\mathtt{S}^t_x\bigwedge y\alpha$ en términos de $x, \alpha, \mathtt{S}^t_x\alpha$, y lo mismo vale para el apartado 8).

Es fácil definir el conjunto v_n formado por las n primeras variables de $\mathcal{L}$, así como probar que el término v_n es Δ_1. Consideremos la fórmula Σ_1

$$\phi(x,t,\theta,\theta') \equiv \bigvee pvgd(p = \ell(\theta) + \ell(t) + 1 \wedge v = v_p$$

$$\wedge\ d = \{\theta_0 \in (\mathrm{Rng}(\theta) \cup v)^{<\ell(\theta)+1} \mid \theta_0 \in \mathrm{Exp}(\mathcal{L})\}$$

$$\wedge\ g : (v \cup \{x\}) \times (v \cup \{t\}) \times d \longrightarrow \mathrm{Exp}(\mathcal{L}) \wedge \theta' = g(x,t,\theta)$$

$$\wedge \bigwedge \xi\tau\theta_0(\xi \in v \cup \{x\} \wedge \tau \in v \cup \{t\} \wedge \theta_0 \in d \to \cdots)),$$

donde los puntos suspensivos son la disyunción de las fórmulas siguientes:

1. $\bigvee y(y \in \mathrm{Var}(\mathcal{L}) \wedge \theta_0 = y$
   $\wedge\ ((y = \xi \wedge g(\xi,\tau,\theta_0) = \tau) \vee (y \neq \xi \wedge g(\xi,\tau,\theta_0) = \theta_0)),$
2. $\bigvee c(c \in \mathrm{Const}(\mathcal{L}) \wedge \theta_0 = c \wedge g(\xi,\tau,\theta_0) = \theta_0),$
3. $\bigvee ftt'n(\ell(t) = \ell(t') = n+1 \wedge f \in \mathrm{Fn}(\mathcal{L})$
   $\wedge\ \theta_0 = ft_0\cdots t_n \wedge \bigwedge i \leq n\ t'_i = g(\xi,\tau,t_i) \wedge g(\xi,\tau,\theta_0) = ft'_0\cdots t'_n),$
4. $\bigvee Rtt'n(\ell(t) = \ell(t') = n+1 \wedge f \in \mathrm{Rel}(\mathcal{L})$
   $\wedge\ \theta_0 = Rt_0\cdots t_n \wedge \bigwedge i \leq n\ t'_i = g(\xi,\tau,t_i) \wedge g(\xi,\tau,\theta_0) = fRt'_0\cdots t'_n),$
5. $\bigvee \alpha(\theta_0 = \neg\alpha \wedge g(\xi,\tau,\theta_0) = \neg g(\xi,\tau,\alpha)),$
6. $\bigvee \alpha\beta(\theta_0 = \alpha \to \beta \wedge g(\xi,\tau,\theta_0) = (g(\xi,\tau,\alpha) \to g(\xi,\tau,\beta))),$
7. $\bigvee yz\alpha(\theta_0 = \bigwedge y\alpha \wedge z = \mathrm{MV}(\bigwedge y\alpha \wedge y = \tau) \wedge ((\xi \notin \mathrm{Vlib}(\theta_0) \wedge g(\xi,\tau,\theta_0) = \theta_0) \vee (\xi \in \mathrm{Vlib}(\theta_0) \wedge y \notin \mathrm{Vlib}(\tau) \wedge g(\xi,\tau,\theta_0) = \bigwedge y\, g(\xi,\tau,\alpha)) \vee (\xi \in \mathrm{Vlib}(\theta_0) \wedge y \in \mathrm{Vlib}(\tau) \wedge g(\xi,\tau,\theta_0) = \bigwedge z\, g(\xi,\tau,g(y,z,\alpha)))),$
8. $\bigvee yz\alpha(\theta_0 = y|\alpha \wedge z = \mathrm{MV}(\bigwedge y\alpha \wedge y = \tau) \wedge ((\xi \notin \mathrm{Vlib}(\theta_0) \wedge g(\xi,\tau,\theta_0) = \theta_0) \vee (\xi \in \mathrm{Vlib}(\theta_0) \wedge y \notin \mathrm{Vlib}(\tau) \wedge g(\xi,\tau,\theta_0) = y|g(\xi,\tau,\alpha)) \vee (\xi \in \mathrm{Vlib}(\theta_0) \wedge y \in \mathrm{Vlib}(\tau) \wedge g(\xi,\tau,\theta_0) = z|g(\xi,\tau,g(y,z,\alpha)))).$

Lo que afirma ϕ es que g es una función definida sobre todas las ternas (ξ,τ,θ_0) tales que ξ es una variable de entre las que aparecen en θ o de entre las del conjunto v formado por las $p = \ell(\theta)+\ell(t)+1$ primeras variables de $\mathcal{L}$, o bien $\xi = x$, a su vez τ es una variable de v o bien el término t y θ_0 es una expresión formada por signos que aparezcan en θ o en v de longitud $\leq \ell(\theta)$. Además, g cumple exactamente las propiedades requeridas para que $g(\xi,\tau,\theta_0)$ sea $\mathtt{S}^\tau_\xi\theta_0$.

La elección del conjunto v se ha hecho para que en los apartados 7) y 8) podamos afirmar que $z \in v$, pues las variables que están en θ_0 o en τ son a lo sumo $\ell(\theta_0) + \ell(\tau) \leq \ell(\theta) + \ell(t) < p$, luego alguna variable de v no está en θ_0 ni en τ, luego $z \in v$. Entonces $g(y, z, \alpha)$ está bien definido y es una expresión de la misma longitud que α, formada por signos de $\mathrm{Rng}(\theta) \cup v$, por lo que $g(\xi, \tau, g(y, z, \alpha))$ también está definido. Lo mismo se aplica al apartado 8).

Hay que probar que existe una g que cumple la definición de ϕ, para lo cual fijamos x, t, θ y probamos por inducción sobre l que para $l \leq \ell(\theta) + 1$ existe una

$$g : (v \cup \{x\}) \times (v \cup \{t\}) \times \{\theta_0 \in (\mathrm{Rng}(\theta) \cup v)^{<l} \mid \theta_0 \in \mathrm{Exp}(\mathcal{L})\} \longrightarrow \mathrm{Exp}(\mathcal{L})$$

que cumple las propiedades indicadas. La inducción es Σ_1, se cumple trivialmente para $l = 0$ y, dada una g definida para fórmulas de longitud $< l$, las propiedades determinan cómo extenderla a fórmulas de longitud l.

También es fácil ver que las propiedades implican que g es única, por lo que

$$\bigwedge x \in \mathrm{Var}(\mathcal{L}) \bigwedge t \in \mathrm{Term}(\mathcal{L}) \bigwedge \theta \in \mathrm{Exp}(\mathcal{L}) \overset{1}{\bigvee} \theta' \, \phi(x, t, \theta, \theta'),$$

y entonces es fácil ver que $\mathtt{S}^t_x\theta \equiv \theta' | \phi(x, t, \theta, \theta')$ cumple lo pedido. ∎

Sustitución simultánea Vamos a necesitar un concepto de sustitución simultánea de varias variables por varios términos en una expresión. Por simplicidad supondremos que los términos son designadores, es decir, que no tienen variables libres, que es el único caso que vamos a necesitar.

Definición 8.9 Sea d un conjunto finito de variables de $\mathcal{L}$, sea e un conjunto finito de designadores de $\mathcal{L}$, sea $v : d \longrightarrow e$ una aplicación arbitraria y sea $y \subset \mathcal{P}d$ tal que $\bigwedge ab(a \subset b \wedge b \in y \to a \in y)$. El teorema 8.6 nos da una función $F_{v,y} : \mathrm{Exp}(\mathcal{L}) \longrightarrow V$ de tipo Σ_1 tal que, para toda $\theta \in \mathrm{Exp}(\mathcal{L})$, se cumple que $F_{v,y}(\theta) : y \longrightarrow \mathrm{Exp}(\mathcal{L})$ y, si representamos $\mathtt{S}_a\theta \equiv F_{v,y}(\theta)(a)$, se cumple

1. Si x es una variable de $\mathcal{L}$, entonces
$$\mathtt{S}_a(x) = \begin{cases} v(x) & \text{si } x \in a, \\ x & \text{si } x \notin a. \end{cases}$$
2. $\mathtt{S}_a c = c$.
3. $\mathtt{S}_a f t_0 \cdots t_n = f \mathtt{S}_a t_0 \cdots \mathtt{S}_a t_n$.
4. $\mathtt{S}_a R t_0 \cdots t_n = R \mathtt{S}_a t_0 \cdots \mathtt{S}_a t_n$.
5. $\mathtt{S}_a(\neg\alpha) = \neg \mathtt{S}_a \alpha$.
6. $\mathtt{S}_a(\alpha \to \beta) = \mathtt{S}_a\alpha \to \mathtt{S}_a\beta$.
7. $\mathtt{S}_a(\bigwedge x \alpha) = \bigwedge x \, \mathtt{S}_{a \setminus \{x\}} \alpha$.
8. $\mathtt{S}_a(x|\alpha) = x | \mathtt{S}_{a \setminus \{x\}} \alpha$.

Es fácil ver que esta definición puede expresarse ciertamente en los términos requeridos por el teorema 8.6, es decir, en términos de funciones Σ_1. El hecho de que d e y no intervengan explícitamente en las condiciones anteriores tiene una consecuencia clara: si $d \subset d'$, $y \subset y'$, $v' : d' \longrightarrow e'$ y $v = v'|_d$, entonces, para todo $a \in y$ y toda expresión θ de $\mathcal{L}$, se cumple que $\mathtt{S}_a\theta$ es el mismo calculado con v, d, y o con v', d', y' (se razona por inducción sobre θ sin más que constatar que la definición es la misma en todos los casos).

Tenemos así que $\mathtt{S}_a\theta$ es un término Δ_1 (con variables libres θ, a, v, y) que representa la expresión que resulta de sustituir en θ cada variable $x \in a$ por el designador $v(x)$. La relación con la sustitución ordinaria de una única variable es la siguiente:

$$\bigwedge a \in y \bigwedge x \in a\ \mathtt{S}_a\theta = \mathtt{S}_x^{v(x)}\mathtt{S}_{a\setminus\{x\}}\theta.$$

Esto se demuestra fácilmente por inducción sobre θ a partir de las definiciones. Como ilustración veamos el caso en que $\theta = \bigwedge y\alpha$. Tenemos que

$$\mathtt{S}_x^{v(x)}\mathtt{S}_{a\setminus\{x\}}\bigwedge y\alpha = \mathtt{S}_x^{v(x)}\bigwedge y\,\mathtt{S}_{a\setminus\{x,y\}}\alpha.$$

Si x e y son la misma variable esto se reduce a

$$\mathtt{S}_y^{v(y)}\bigwedge y\,\mathtt{S}_{a\setminus\{y\}}\alpha = \bigwedge y\,\mathtt{S}_{a\setminus\{y\}}\alpha = \mathtt{S}_a\bigwedge y\alpha.$$

Si son variables distintas queda

$$\bigwedge y\,\mathtt{S}_x^{v(x)}\mathtt{S}_{a\setminus\{x,y\}}\alpha = \bigwedge y\,\mathtt{S}_{a\setminus\{y\}}\alpha = \mathtt{S}_a\bigwedge y\alpha,$$

donde hemos usado la hipótesis de inducción.

Otro hecho que se prueba fácilmente es que si a no contiene variables libres en θ, entonces $\mathtt{S}_a\theta = \theta$.

Deducciones lógicas Ahora ya es fácil definir los axiomas lógicos:

Definición 8.10 Consideramos las fórmulas siguientes (relativas a un lenguaje formal $\mathcal{L}$ dado):

- $u \in \mathrm{K1} \equiv \bigvee \alpha\beta \in \mathrm{Sub}(u)\ (\alpha, \beta \in \mathrm{Form}(\mathcal{L}) \wedge u = \alpha \to (\beta \to \alpha))$,
- $u \in \mathrm{K2} \equiv \bigvee \alpha\beta\gamma \in \mathrm{Sub}(u)\ (\alpha, \beta, \gamma \in \mathrm{Form}(\mathcal{L}) \wedge$
  $u = (\alpha \to (\beta \to \gamma)) \to ((\alpha \to \beta) \to (\alpha \to \gamma)))$,
- $u \in \mathrm{K3} \equiv \bigvee \alpha\beta \in \mathrm{Sub}(u)(\alpha, \beta \in \mathrm{Form}(\mathcal{L}) \wedge$
  $u = (\neg\alpha \to \neg\beta) \to (\beta \to \alpha))$,
- $u \in \mathrm{K4} \equiv \bigvee x \in \mathrm{Rng}(u) \bigvee t\alpha \in \mathrm{Sub}(u)(x \in \mathrm{Var}(\mathcal{L}) \wedge t \in \mathrm{Term}(\mathcal{L})$
  $\wedge\ \alpha \in \mathrm{Form}(\mathcal{L}) \wedge u = \bigwedge x\alpha \to \mathtt{S}_x^t\alpha)$,
- $u \in \mathrm{K5} \equiv \bigvee x \in \mathrm{Rng}(u) \bigvee \alpha\beta \in \mathrm{Sub}(u)(x \in \mathrm{Var}(\mathcal{L}) \wedge \alpha, \beta \in \mathrm{Form}(\mathcal{L})$
  $\wedge\ x \notin \mathrm{Vlib}(\alpha) \wedge u = \bigwedge x(\alpha \to \beta) \to (\alpha \to \bigwedge x\beta))$,

- $u \in \mathrm{K6} \equiv \bigvee x \in \mathrm{Rng}(u) \bigvee t\alpha \in \mathrm{Sub}(u)(x \in \mathrm{Var}(\mathcal{L}) \wedge t \in \mathrm{Term}(\mathcal{L})$
  $\wedge\, \alpha \in \mathrm{Form}(\mathcal{L}) \wedge x \notin \mathrm{Vlib}(t) \wedge u = \bigwedge x(x = t \to \alpha) \leftrightarrow \mathtt{S}_x^t\alpha)$,
- $u \in \mathrm{K7} \equiv \bigvee x \in \mathrm{Rng}(u) \bigvee \alpha \in \mathrm{Sub}(u)(x \in \mathrm{Var}(\mathcal{L}) \wedge \alpha \in \mathrm{Form}(\mathcal{L}))$
  $\wedge\, u = \overset{1}{\bigvee} x\alpha \to \mathtt{S}_x^{x|\alpha}\alpha)$,
- $u \in \mathrm{K8} \equiv \bigvee xy \in \mathrm{Rng}(u) \bigvee \alpha \in \mathrm{Sub}(u)(x, y \in \mathrm{Var}(\mathcal{L}) \wedge \alpha \in \mathrm{Form}(\mathcal{L}))$
  $\wedge\, u = \neg\overset{1}{\bigvee} x\alpha \to x|\alpha = y|(y = y))$,
- $u \in \mathrm{Axl}(\mathcal{L}) \equiv u \in \mathrm{K1} \vee u \in \mathrm{K2} \vee u \in \mathrm{K3} \vee u \in \mathrm{K4} \vee u \in \mathrm{K5}$
  $\vee\, u \in \mathrm{K6} \vee u \in \mathrm{K7} \vee u \in \mathrm{K8}$.

Es claro[5] que todas las fórmulas anteriores son Δ_1.

Definición 8.11 Diremos que d es una *deducción lógica* en un lenguaje formal $\mathcal{L}$ a partir de un conjunto de *premisas* c si se cumple

$$\mathrm{Ded}(d, c) \equiv d \in \mathrm{SucCad}(\mathcal{L}) \wedge \bigwedge i < \ell(d)(d_i \in \mathrm{Form}(\mathcal{L}) \wedge \cdots),$$

donde los puntos suspensivos son la disyunción de las fórmulas siguientes:

1. $d_i \in \mathrm{Axl}(\mathcal{L})$,
2. $d_i \in c$,
3. $\bigvee kl < i(d_k = (d_l \to d_i))$,
4. $\bigvee k < i \bigvee x \in \mathrm{Rng}(d_i)(d_i = \bigwedge x\, d_k)$.

En el caso 1) se dice que d_i es un axioma lógico, en el caso 2) que d_i es una premisa de la deducción, en el caso 3) que d_i se deduce por *modus ponens* de d_k y d_l, y en el caso 4) que d_i se deduce por *generalización* respecto de la variable x de d_k. Escribiremos

$$c \overset{d}{\vdash} \alpha \equiv \mathrm{Ded}(d, c) \wedge \ell(d) > 0 \wedge d_{\ell(d)-1} = \alpha.$$

Claramente se trata de una fórmula Δ_1, mientras que la fórmula

$$c \vdash \alpha \equiv \bigvee d\ c \overset{d}{\vdash} \alpha$$

es Σ_1 (y la leeremos "α es una *consecuencia lógica* de c").

Escribiremos $\alpha_1, \ldots, \alpha_n \vdash \alpha$ en lugar de $\{\alpha_1, \ldots, \alpha_n\} \vdash \alpha$ y simplemente $\vdash \alpha$ en lugar de $\varnothing \vdash \alpha$.

A partir de aquí todos los resultados metamatemáticos sobre deducciones lógicas del capítulo II pueden verse trivialmente como teoremas de $\mathrm{I}\Sigma_1$. Por ejemplo, el teorema 2.7 se formaliza ahora como el teorema siguiente:

[5]Notemos que las variables α, β en K1 están acotadas, pues podemos escribir

$$\bigvee v(v = \mathrm{Sub}(u) \wedge \bigvee \alpha\beta \in v(\cdots)) \text{ o } \bigwedge v(v = \mathrm{Sub}(u) \to \bigvee \alpha\beta \in v(\cdots)),$$

e igualmente en los axiomas siguientes.

Teorema 8.12 $\bigwedge\alpha \in \mathrm{Form}(\mathcal{L})\ \vdash \alpha \to \alpha$.

La demostración es exactamente la misma: la sucesión de cinco fórmulas mostrada allí puede verse como la definición de una sucesión d de longitud 5 que satisface claramente la definición de deducción de $\alpha \to \alpha$. Analicemos, por ejemplo, la prueba del teorema de deducción:

Teorema 8.13 (Teorema de deducción) *Sean α y β fórmulas de $\mathcal{L}$ y sea c un conjunto de fórmulas de $\mathcal{L}$. Si $c \cup \{\alpha\} \vdash \beta$ y existe una deducción de β en la que no se generalice respecto a variables libres en α, entonces $\Gamma \vdash \alpha \to \beta$.*

DEMOSTRACIÓN: Suponemos que existe d tal que $c \cup \{\alpha\} \vdash^d \beta$ de modo que siempre que se aplica el apartado 4) de la definición de deducción se cumple además que $x \notin \mathrm{Vlib}(\alpha)$. Vamos a probar por inducción sobre i que

$$\bigwedge i < \ell(d) \bigvee d'\ c \overset{d'}{\vdash} \alpha \to d_i.$$

Notemos que la fórmula es Σ_1, luego la inducción es legítima. Al aplicar esto a $i = \ell(d) - 1$ obtenemos una deducción de $\alpha \to \beta$.

Lo suponemos cierto para todo $j < i$. Si d_i es un axioma lógico o una premisa, entonces tomamos como d' la sucesión

$$\langle d_i, d_i \to (\alpha \to d_i), \alpha \to d_i) \rangle\,,$$

que es claramente una deducción de $\alpha \to d_i$. La prueba continúa calcando literalmente el argumento metamatemático que vimos en el capítulo II. No merece la pena repetirlo aquí. ∎

Igualmente podemos "calcar" (es decir, formalizar) todas las pruebas de todas las reglas derivadas de inferencia y todos los criterios generales que hemos dado sobre deducciones (razonamiento por reducción al absurdo, etc.).

Definición 8.14 Una *teoría axiomática* T sobre un lenguaje formal $\mathcal{L}$ es una fórmula $u \in \mathrm{Ax}(T)$ tal que $\bigwedge u \in \mathrm{Ax}(T)\ u \in \mathrm{Form}(\mathcal{L})$. Diremos que T es una *teoría axiomática semirrecursiva* (resp. *recursiva*) si la fórmula es Σ_1 (resp.) Δ_1.

Definimos una *demostración* en una teoría axiomática T como

$$\mathrm{Dm}_T(d) \equiv d \in \mathrm{SucCad}(\mathcal{L}) \wedge \bigwedge i < \ell(d)(d_i \in \mathrm{Form}(\mathcal{L}) \wedge \cdots),$$

donde los puntos suspensivos son la disyunción de las fórmulas siguientes:

1. $d_i \in \mathrm{Axl}(\mathcal{L})$,
2. $d_i \in \mathrm{Ax}(T)$,
3. $\bigvee kl < i(d_k = (d_l \to d_i))$,
4. $\bigvee k < i \bigvee x \in \mathrm{Rng}(d_i)(d_i = \bigwedge x\, d_k)$.

Es claro que si T es una teoría axiomática (semi)recursiva entonces $\mathrm{Dm}_T(d)$ es una fórmula Δ_1 (resp. Σ_1), como también lo es la fórmula

$$\overset{d}{\underset{T}{\vdash}}\alpha \equiv \mathrm{Dm}_T(d) \wedge \bigvee n(\ell(d) = n+1 \wedge d_n = \alpha),$$

pues equivale a $\mathrm{Dm}_T(d) \wedge \bigwedge n(\ell(d) = n+1 \to d_n = \alpha)$.

Diremos que α es un *teorema* de T si cumple

$$\underset{T}{\vdash}\alpha \equiv \bigvee d \overset{d}{\underset{T}{\vdash}}\alpha.$$

Notemos que si la teoría axiomática es (semi)recursiva esta fórmula es Σ_1, pero no necesariamente Δ_1. Es claro que toda consecuencia lógica de teoremas de T es un teorema de T.

Disyunciones y conjunciones finitas Si $\mathcal{L}$ es un lenguaje formal y llamamos $0 = \bigvee x\ x \neq x$, se cumple claramente:

1. $\bigwedge \alpha\beta \in \mathrm{Form}(\mathcal{L})\ \alpha \vee \beta \in A$
2. $\bigwedge \alpha\beta \in \mathrm{Form}(\mathcal{L})\ \vdash \alpha \vee \beta \leftrightarrow \beta \vee \alpha$
3. $\bigwedge \alpha\beta\gamma \in \mathrm{Form}(\mathcal{L})\ \vdash \alpha \vee (\beta \vee \gamma) \leftrightarrow (\alpha \vee \beta) \vee \gamma$
4. $\bigwedge \alpha \in \mathrm{Form}(\mathcal{L})\ \vdash \alpha \vee 0 \leftrightarrow \alpha$.

Estas propiedades son análogas a los presupuestos de la sección 6.5 salvo que tenemos equivalencias lógicas en lugar de igualdades. Si $s : I_n \longrightarrow \mathrm{Form}(\mathcal{L})$, podemos definir igualmente

$$\bigvee_{i<0} s_i = 0 \wedge \bigwedge m < \ell(s) \bigvee_{i<m+1} s_i = \bigvee_{i<m} s_i \vee s_m,$$

y los teoremas de la sección 6.5 se adaptan trivialmente cambiando igualdades por equivalencias lógicas. Por ejemplo, el teorema 6.39 se convierte en la equivalencia

$$\vdash \bigvee_{i<m+n} s_i = \bigvee_{i<m} s_i \vee \bigvee_{i<n} s_{m+i}.$$

La propiedad conmutativa generalizada (junto con su prueba) se traduce en este contexto a la equivalencia

$$\vdash \bigvee_{i<n} s_i \leftrightarrow \bigvee_{i<n} s_{\sigma(i)}.$$

Sea ahora x un conjunto finito de fórmulas de $\mathcal{L}$ y sea $n = |x|$. Se cumple entonces que

$$\overset{1}{\bigvee} sn(s : I_n \longrightarrow x \text{ biyectiva} \wedge \bigwedge ij < n(i < n \to s_i < s_j)).$$

En efecto, si $n = |x|$, por inducción sobre m se prueba que

$$m \leq n \to \bigvee^1 s(s : I_m \longrightarrow x \text{ inyectiva} \wedge \bigwedge ij < m(i < j \to s_i < s_j)$$

$$\wedge \bigwedge i < m \bigwedge u \in x \setminus \mathcal{R}s\ s_i < u).$$

En efecto, para $m = 0$ es trivial y, si vale para m y $m + 1 \leq n$, tomamos $s : I_m \longrightarrow x$ según la hipótesis de inducción. Entonces $x \setminus \mathcal{R}s \neq \varnothing$, pues $m < |x|$, luego s no puede ser biyectiva, tomamos $\alpha \in x \setminus \mathcal{R}x$ el mínimo elemento (que existe por 6.28) y formamos $s' = s \cup \{(m, \alpha)\}$. Es claro que s' cumple lo pedido. La unicidad es clara: si $h : I_{m+1} \longrightarrow x$ cumple lo mismo, entonces $h|_m = s$ por la unicidad de la hipótesis de inducción, luego $h|_m = s'|_m$. Por otra parte, $h_m \in x \setminus \mathcal{R}s$ porque h es inyectiva, y es menor que todo elemento de $x \setminus \mathcal{R}h$, luego h_m es el mínimo de $x \setminus \mathcal{R}s$, es decir, $h_m = \alpha$ y así $h = s'$.

Aplicando esto al caso $m = n$ tenemos que s tiene que ser biyectiva (por 6.30) y se cumple lo requerido. Así podemos definir

$$\bigvee_{\alpha \in x} \alpha \equiv \beta \mid \bigvee n(s : I_n \longrightarrow x \text{ biyectiva} \wedge \bigwedge ij < m(i < n \to s_i < s_j) \wedge \beta = \bigvee_{i<n} s_i).$$

Ahora se prueba fácilmente que $\bigvee\limits_{\alpha \in \varnothing} \alpha = 0$, que $\vdash \bigvee\limits_{\alpha \in \{\beta\}} \alpha \leftrightarrow \beta$, así como que si x e y son conjuntos finitos disjuntos de fórmulas de $\mathcal{L}$, se cumple

$$\vdash \bigvee_{\alpha \in x \cup y} \alpha \leftrightarrow \bigvee_{\alpha \in x} \alpha \vee \bigvee_{\alpha \in y} \alpha.$$

Un poco más en general, si $s : x \longrightarrow \mathrm{Form}(\mathcal{L})$, donde x es un conjunto finito, podemos definir

$$\bigvee_{i \in x} s_i \equiv \bigvee_{\alpha \in s[x]} \alpha.$$

Estos resultados justifican fácilmente el uso de notaciones más generales, como

$$\bigvee_{i=1}^{n} \alpha_i \qquad \text{o} \qquad \alpha_1 \vee \cdots \vee \alpha_n,$$

así como las manipulaciones elementales de estas expresiones (siempre en términos de equivalencias lógicas y no de igualdades).

De forma completamente análoga se pueden definir las conjunciones finitas, que podemos expresar con las notaciones

$$\bigwedge_{i=1}^{n} \alpha_i \qquad \text{o} \qquad \alpha_1 \wedge \cdots \wedge \alpha_n.$$

Notemos que la conjunción vacía $\bigwedge\limits_{\alpha \in \varnothing} \alpha$ tiene que definirse como $1 = \bigvee x\ x = x$ para que se cumpla que $\vdash \alpha \wedge 1 \leftrightarrow \alpha$.

8.2 Relación con las teorías metamatemáticas

En la sección precedente hemos definido los lenguajes formales $\ulcorner\mathcal{L}_a\urcorner$ y $\ulcorner\mathcal{L}_{\rm tc}\urcorner$, pero esto no tendría ningún sentido si no hubiéramos definido previamente los lenguajes $\mathcal{L}_a$ o $\mathcal{L}_{\rm tc}$ en los cuales definir $\ulcorner\mathcal{L}_a\urcorner$ y $\ulcorner\mathcal{L}_{\rm tc}\urcorner$: Necesitamos lenguajes formales metamatemáticos definidos informalmente para definir formalmente el concepto de lenguaje formal. Sería catastrófico confundir $\mathcal{L}_a$ con $\ulcorner\mathcal{L}_a\urcorner$, confundir cuándo estamos hablando metamatemáticamente de $\mathcal{L}_a$ y cuándo estamos hablando de $\ulcorner\mathcal{L}_a\urcorner$ en $\mathcal{L}_a$ o $\mathcal{L}_{\rm tc}$. En esta sección estudiaremos con detalle la relación entre los lenguajes metamatemáticos y sus reflejos formales. Ante todo observamos que no es preciso limitarse a los dos ejemplos que estamos considerando.

Definición 8.15 Diremos que un lenguaje formal $\mathcal{L}$ (cuyos signos sean números naturales)[6] es *recursivo* si son recursivas las relaciones $\mathrm{Var}(k)$, $\mathrm{Rel}(k)$ y $\mathrm{Fn}(k)$ que se cumplen, respectivamente, cuando el número k es una variable, un relator o un funtor de $\mathcal{L}$ (entendiendo que las constantes son funtores 0-ádicos), así como la función $\mathrm{Nar}(k)$ que vale cero salvo sobre los relatores y funtores de $\mathcal{L}$, en los cuales es igual a su número de argumentos.

En suma, un lenguaje formal es recursivo si tenemos un criterio práctico para reconocer cuáles son sus variables, relatores y funtores, así como para determinar el número de argumentos que requiere cada uno. Este requisito no lo habíamos planteado explícitamente hasta ahora, pero es obvio que debe cumplirlo todo lenguaje formal que pretenda ser de alguna utilidad. Si no fuéramos capaces de determinar si un número dado es o no un relator o un funtor de $\mathcal{L}$, no podríamos saber si una cadena de signos de $\mathcal{L}$ es o no un término o una fórmula, ni tampoco si una sucesión de cadenas de signos es o no una demostración.

Si $\mathcal{L}$ es un lenguaje formal recursivo, existen fórmulas Δ_1 de $\mathcal{L}_a$, que podemos representar por $x \in \mathrm{Var}(\ulcorner\mathcal{L}\urcorner)$, $x \in \mathrm{Rel}(\ulcorner\mathcal{L}\urcorner)$, $x \in \mathrm{Fn}(\ulcorner\mathcal{L}\urcorner)$, $\mathrm{Nar}(x,n)$, de modo que

$$\mathrm{Var}(k) \text{ syss } \mathbb{N} \vDash 0^{(k)} \in \mathrm{Var}(\ulcorner\mathcal{L}\urcorner),$$

$$\mathrm{Rel}(k) \text{ syss } \mathbb{N} \vDash 0^{(k)} \in \mathrm{Rel}(\ulcorner\mathcal{L}\urcorner),$$

$$\mathrm{Fn}(k) \text{ syss } \mathbb{N} \vDash 0^{(k)} \in \mathrm{Fn}(\ulcorner\mathcal{L}\urcorner),$$

$$\mathrm{Nar}(k) = n \text{ syss } \mathbb{N} \vDash \mathrm{Nar}(0^{(k)}, 0^{(n)}).$$

Llamaremos $\ulcorner\neg\urcorner, \ulcorner\to\urcorner, \ulcorner\bigwedge\urcorner, \ulcorner|\urcorner, \ulcorner=\urcorner$ a los numerales en $\mathcal{L}_a$ de los signos correspondientes de $\mathcal{L}$. Es decir, si el negador de $\mathcal{L}$ es, por ejemplo, el número 10, entonces $\ulcorner\neg\urcorner \equiv 0^{(10)}$. Lo mismo vale para cualquier otro signo de $\mathcal{L}$. Por ejemplo, si una variable x de $\mathcal{L}$ es el número 4, entonces $\ulcorner x\urcorner \equiv 0^{(4)}$, etc.

[6] Tal y como explicábamos en la introducción de este capítulo, si el lector prefiere considerar que los signos de $\mathcal{L}$ no son números naturales, sólo tiene que asignar arbitrariamente a cada uno de ellos un número natural (al que se llama número de Gödel del signo). Entonces, $\mathcal{L}$ es recursivo si dicha asignación puede hacerse de modo que las relaciones "ser el número de Gödel de una variable", etc. sean recursivas.

Diremos que $\mathcal{L}$ es *demostrablemente recursivo* si en $\mathrm{I}\Sigma_1$ puede probarse que los numerales y las fórmulas precedentes definen un lenguaje formal en el sentido de la sección anterior. Esto supone, entre otras cosas, que en $\mathrm{I}\Sigma_1$ se demuestra

$$\bigwedge x \overset{1}{\bigvee} n\, \mathrm{Nar}(x,n) \wedge \bigwedge a(a \text{ es finito} \to \bigvee x \in \mathrm{Var}(\ulcorner\mathcal{L}\urcorner)\ x \notin a).$$

Más en general, podríamos definir lenguajes demostrablemente recursivos en una teoría T, no necesariamente $\mathrm{I}\Sigma_1$, pero no es necesario hacerlo, pues todos los lenguajes formales de interés son demostrablemente recursivos en este sentido, es decir, que las definiciones de los conjuntos de signos de $\mathcal{L}$ son tan simples y explícitas que $\mathrm{I}\Sigma_1$ basta para demostrar que cumplen lo requerido.

De aquí en adelante supondremos tácitamente que todos los lenguajes formales (metamatemáticos) que consideraremos serán demostrablemente recursivos. Ciertamente es el caso de los lenguajes $\mathcal{L}_a$ y $\mathcal{L}_{\mathrm{tc}}$. Más en general, todo lenguaje con un número finito de relatores y funtores es demostrablemente recursivo.

Es inmediato que las relaciones aritméticas determinadas por las fórmulas definidas en la sección anterior se interpretan en el modelo $\mathbb{N}$ de forma natural. Por ejemplo, la relación $n \in \mathrm{Exp}(\mathcal{L})$, es decir, la relación dada por

$$n \in \mathrm{Exp}(\mathcal{L}) \quad \text{syss} \quad \mathbb{N} \vDash 0^{(n)} \in \mathrm{Exp}(\ulcorner\mathcal{L}\urcorner),$$

no es sino la relación que se cumple cuando n codifica una sucesión de números naturales que, vistos como signos de $\mathcal{L}$, forman una expresión de $\mathcal{L}$. Más aún, si consideramos que las sucesiones finitas de números naturales *son* números naturales, entonces cada cadena ζ de signos de $\mathcal{L}$ *es* un número natural, y si llamamos $\ulcorner\zeta\urcorner \equiv \langle\zeta\rangle \equiv \langle\zeta_0), \ldots, \zeta_n\rangle$ al término considerado en la sección 5.6, la equivalencia anterior puede expresarse así:

$$\theta \in \mathrm{Exp}(\mathcal{L}) \quad \text{syss} \quad \mathbb{N} \vDash \ulcorner\theta\urcorner \in \mathrm{Exp}(\ulcorner\mathcal{L}\urcorner).$$

Lo mismo vale para todas las demás relaciones definidas en la sección anterior. Por ejemplo

$$x \text{ es una variable libre en } \theta \quad \text{syss} \quad \mathbb{N} \vDash \ulcorner x\urcorner \in \mathrm{Vlib}(\ulcorner\theta\urcorner).$$

El hecho de que todas las fórmulas consideradas sean Δ_1 se traduce en que todas las relaciones correspondientes son recursivas (lo cual es razonable: existe un algoritmo finito para determinar si un número natural es (o codifica) una sucesión de números naturales que, concretamente, forman una expresión de $\mathcal{L}$, etc.). En particular tenemos que el conjunto de los axiomas lógicos[7] de un lenguaje formal es recursivo.

Más aún, podemos considerar que, por definición, las fórmulas que estamos considerando no son las que hemos descrito, sino las dadas por el teorema 5.30, que son equivalentes en $\mathrm{I}\Sigma_1$ a las que estamos considerando (véase la observación

[7] Alternativamente, el conjunto de los números de Gödel de los axiomas lógicos es recursivo.

posterior) y con este convenio tenemos además que todas ellas son demostrables en Q, es decir, que, por ejemplo:

$$\theta \in \mathrm{Exp}(\mathcal{L}) \quad \text{syss} \quad \underset{\mathrm{Q}}{\vdash} \ulcorner\theta\urcorner \in \mathrm{Exp}(\ulcorner\mathcal{L}\urcorner),$$

$$\theta \notin \mathrm{Exp}(\mathcal{L}) \quad \text{syss} \quad \underset{\mathrm{Q}}{\vdash} \ulcorner\theta\urcorner \notin \mathrm{Exp}(\ulcorner\mathcal{L}\urcorner).$$

Similarmente, podemos considerar que cada sucesión finita d de fórmulas de $\mathcal{L}$ es un número natural, en cuyo caso podemos representar por $\ulcorner d\urcorner$ su numeral correspondiente.[8] Si $c = \{a_1, \ldots, a_n\}$ es un conjunto finito de números naturales, podemos representar $\ulcorner c\urcorner \equiv \{0^{(a_1)}, \ldots, 0^{(a_n)}\}$, y entonces tenemos que

$$c \overset{d}{\vdash} \alpha \quad \text{syss} \quad \mathbb{N} \vDash \ulcorner c\urcorner \overset{\ulcorner d\urcorner}{\vdash} \ulcorner\alpha\urcorner,$$

de modo que d es una deducción de α con premisas c si y sólo si la fórmula que formaliza este concepto es verdadera en $\mathbb{N}$. En particular

$$c \vdash \alpha \quad \text{syss} \quad \mathbb{N} \vDash \ulcorner c\urcorner \vdash \ulcorner\alpha\urcorner.$$

No podemos afirmar que el conjunto de los axiomas (propios) de una teoría axiomática arbitraria es recursivo porque se trata de un conjunto arbitrario. Lo que procede en este punto es dar una definición:

Definición 8.16 Una teoría axiomática T es *(semi)recursiva* si el conjunto de sus axiomas (considerados como números naturales) es (semi)recursivo.

En otras palabras, una teoría es recursiva cuando existe un algoritmo finito para determinar si una fórmula dada es o no uno de sus axiomas. Aunque es evidente que una teoría axiomática no recursiva no tiene ningún interés práctico (sin un medio para reconocer los axiomas tampoco tenemos un medio para reconocer las demostraciones), la posibilidad de manejar conjuntos no recursivos de fórmulas y de tomarlos como axiomas de una determinada teoría axiomática tiene cierto interés teórico, así que, al contrario de lo que hemos hecho con los lenguajes formales, no vamos a dar por supuesto que todas las teorías axiomáticas que consideramos son recursivas o semirrecursivas, sino que explicitaremos esta hipótesis cada vez que la usemos.

Si T es una teoría semirrecursiva sobre un lenguaje formal $\mathcal{L}$, existe una fórmula $x \in \mathrm{Ax}(\ulcorner T\urcorner)$ de tipo Σ_1 en $\mathcal{L}_a$ tal que, para todo número natural α (en particular para toda fórmula α de $\mathcal{L}$, considerada como número natural)

$$\alpha \in \mathrm{Ax}(T) \quad \text{syss} \quad \mathbb{N} \vDash \ulcorner\alpha\urcorner \in \mathrm{Ax}(\ulcorner T\urcorner),$$

donde la parte izquierda es, naturalmente una abreviatura por "α es un axioma (propio) de $\mathcal{L}$". Cambiando $x \in \mathrm{Ax}(\ulcorner T\urcorner)$ por $x \in \mathrm{Form}(\ulcorner\mathcal{L}\urcorner) \wedge x \in \mathrm{Ax}(\ulcorner T\urcorner)$ si es necesario, podemos suponer que además $\vdash \bigwedge u \in \mathrm{Ax}(\ulcorner T\urcorner)\ u \in \mathrm{Form}(\ulcorner\mathcal{L}\urcorner)$. Así la fórmula $x \in \mathrm{Ax}(\ulcorner T\urcorner)$ define una teoría axiomática semirrecursiva en $\mathrm{I}\Sigma_1$ según la definición 8.14.

[8] Equivalentemente, podemos considerar que $\ulcorner d\urcorner$ es el numeral correspondiente al número de Gödel de la demostración d.

Esto nos permite construir la fórmula $\underset{\ulcorner T\urcorner}{\vdash} \alpha$ (de tipo Σ_1) de modo que

$$\underset{T}{\vdash} \alpha \quad \text{syss} \quad \mathbb{N} \vDash \underset{\ulcorner T\urcorner}{\vdash} \ulcorner\alpha\urcorner \quad \text{syss} \quad \underset{\mathrm{Q}}{\vdash} \underset{\ulcorner T\urcorner}{\vdash} \ulcorner\alpha\urcorner.$$

Por lo tanto:

Teorema 8.17 *El conjunto de los teoremas*[9] *de una teoría semirrecursiva es semirrecursivo.*

No obstante, en general no podemos garantizar que el conjunto de los teoremas de una teoría axiomática sea recursivo, ni siquiera cuando la teoría es recursiva. En otras palabras: no disponemos de un algoritmo para decidir si una fórmula dada es o no un teorema de una teoría dada. Más adelante ahondaremos en este asunto.

Por definición, una teoría T es recursiva si y sólo si la relación $\alpha \in \mathrm{Ax}(T)$ es Δ_1, pero eso no significa que la fórmula $x \in \mathrm{Ax}(\ulcorner T\urcorner)$ pueda tomarse Δ_1 en una teoría dada:

Definición 8.18 Una teoría axiomática T es *demostrablemente recursiva* en una teoría axiomática S que extienda a $\mathrm{I}\Sigma_1$ si existe una fórmula $x \in \mathrm{Ax}(\ulcorner T\urcorner)$ de tipo Δ_1^S tal que, para todo número natural α, se cumple

$$\alpha \in \mathrm{Ax}(T) \quad \text{syss} \quad \mathbb{N} \vDash \ulcorner\alpha\urcorner \in \mathrm{Ax}(\ulcorner T\urcorner).$$

En tal caso dicha fórmula define una teoría recursiva en S en el sentido de la definición 8.14. Cuando digamos que una teoría T es *demostrablemente recursiva* nos referiremos a que lo es en $\mathrm{I}\Sigma_1$.

Obviamente, si una teoría T es demostrablemente recursiva en una teoría S que cumpla $\mathbb{N} \vDash S$ entonces es recursiva, pero el recíproco no es cierto en general. Concretamente, que la teoría T sea recursiva significa que existe otra fórmula $x \in \mathrm{Ax}'(\ulcorner T\urcorner)$ de tipo Π_1 de modo que

$$\alpha \in \mathrm{Ax}(T) \quad \text{syss} \quad \mathbb{N} \vDash \ulcorner\alpha\urcorner \in \mathrm{Ax}(\ulcorner T\urcorner) \quad \text{syss} \quad \mathbb{N} \vDash \ulcorner\alpha\urcorner \in \mathrm{Ax}'(\ulcorner T\urcorner).$$

En particular

$$\mathbb{N} \vDash \bigwedge x(x \in \mathrm{Ax}(\ulcorner T\urcorner) \leftrightarrow x \in \mathrm{Ax}'(\ulcorner T\urcorner)).$$

Sin embargo, esto no implica que la equivalencia sea demostrable en $\mathrm{I}\Sigma_1$ o en cualquier otra teoría S sobre $\mathcal{L}_a$. No obstante, si llamamos S a la teoría que resulta de añadir la equivalencia a los axiomas de $\mathrm{I}\Sigma_1$, es claro que $\mathbb{N} \vDash S$ y que T es demostrablemente recursiva sobre S. Por lo tanto, toda teoría axiomática recursiva es demostrablemente recursiva sobre una cierta teoría axiomática S tal que $\mathbb{N} \vDash S$. Sin embargo, se cumple algo más fuerte:

Teorema 8.19 (Craig) *Toda teoría semirrecursiva es equivalente (en el sentido de tener los mismos teoremas) a una teoría demostrablemente recursiva.*

[9] Alternativamente, "de los números de Gödel de los teoremas". No volveremos a hacer más aclaraciones de este tipo.

DEMOSTRACIÓN: Sea $x \in \mathrm{Ax}(\ulcorner T \urcorner) \equiv \bigvee y\, \phi(x,y)$, donde la fórmula ϕ es de tipo Δ_0. Definimos recurrentemente en $\mathrm{I}\Sigma_1$ la función Δ_1 que satisface:

$$R(0,\alpha) = \alpha \wedge \bigwedge n\, R(n+1,\alpha) = R(n,\alpha) \wedge \alpha.$$

Consideramos la fórmula Δ_1 dada por

$$\delta(x) \equiv \bigvee yn\alpha \leq x(x = R(n,\alpha) \wedge \phi(\alpha, y))$$

y sea S la teoría cuyos axiomas son las fórmulas β que cumplen $\mathbb{N} \vDash \delta(\ulcorner \beta \urcorner)$. Es claro entonces que S es demostrablemente recursiva (en $\mathrm{I}\Sigma_1$). Además es equivalente a T, pues si β es un axioma de S entonces existen m, n, α tales que

$$\mathbb{N} \vDash \ulcorner \beta \urcorner = R(0^{(n)}, \ulcorner \alpha \urcorner) \wedge \phi(\ulcorner \alpha \urcorner, 0^{(m)}).$$

Como $\mathbb{N} \vDash \bigvee y\, \phi(\ulcorner \alpha \urcorner, y)$, tenemos que α es un axioma de T y por definición de R resulta que $\beta = \alpha \wedge \cdots \wedge \alpha$. Claramente entonces $\underset{T}{\vdash} \beta$.

Recíprocamente, si α es un axioma de T, tenemos que existe un m tal que $\mathbb{N} \vDash \phi(\ulcorner \alpha \urcorner, 0^{(m)})$. Es claro que, para un n suficientemente grande, la conjunción $\beta \equiv \alpha \wedge \cdots \wedge \alpha$, donde α se repite $n+1$ veces es (vista como número natural) mayor que m y que n. Entonces $\mathbb{N} \vDash \ulcorner \beta \urcorner = R(0^{(n)}, \ulcorner \alpha \urcorner)$, de donde a su vez $\mathbb{N} \vDash \delta(\ulcorner \beta \urcorner)$, es decir, que β es un axioma de S, luego $\underset{S}{\vdash} \alpha$.

Como todo axioma de S es un teorema de T y viceversa, es claro que S y T tienen los mismos teoremas. ∎

Aunque este resultado tiene interés teórico, en la práctica es raramente necesario, pues, por una parte, los resultados que vamos a probar se aplican directamente a teorías semirrecursivas y, por otra parte, todas las teorías de interés práctico son demostrablemente recursivas.

Por ejemplo, la aritmética de Robinson Q es claramente recursiva. Basta definir

$$\alpha \in \mathrm{Ax}(\ulcorner \mathrm{Q} \urcorner) \equiv \alpha = \ulcorner \mathrm{Q1} \urcorner \vee \cdots \vee \alpha = \ulcorner \mathrm{Q7} \urcorner.$$

De hecho, es obvio que toda teoría con un número finito de axiomas es demostrablemente recursiva. También AP es demostrablemente recursiva, pues basta tomar

$$\alpha \in \mathrm{Ax}(\ulcorner \mathrm{AP} \urcorner) \equiv \alpha \in \mathrm{Ax}(\ulcorner \mathrm{Q} \urcorner) \vee \alpha \in \ulcorner I \urcorner$$

donde

$$\alpha \in \ulcorner I \urcorner \equiv \bigvee x\phi \leq \alpha(\phi \in \mathrm{Form}(\ulcorner \mathcal{L}_a \urcorner)$$
$$\wedge\, \alpha = (\mathtt{S}^0_x \phi \wedge \bigwedge x(\phi \to \mathtt{S}^{Sx}_x \phi) \to \bigwedge x\, \phi)).$$

Para demostrar que $\mathrm{I}\Sigma_1$ es demostrablemente recursiva necesitamos definir en $\mathrm{I}\Sigma_1$ el concepto de fórmula Σ_1. Observamos en primer lugar que la fórmula

$$t \in \mathrm{Term}_a(\ulcorner \mathcal{L}_a \urcorner) \equiv t \in \mathrm{Term}(\ulcorner \mathcal{L}_a \urcorner) \wedge \bigwedge i < \ell(x)\ t_i \neq \ulcorner \urcorner$$

es Δ_1, y $\mathbb{N} \vDash \ulcorner t \urcorner \in \mathrm{Term}(\ulcorner \mathcal{L}_a \urcorner)$ syss t es un término sin descriptores de $\mathcal{L}_a$. Más aún, en $\mathrm{I}\Sigma_1$ (o KP) se demuestra trivialmente que $t \in \mathrm{Term}_a(\ulcorner \mathcal{L}_a \urcorner)$ si y sólo si está definido por una sucesión de expresiones que sólo contiene términos.

Ahora definimos:

$$\alpha \in \Delta_0 \equiv \bigvee sm(s \in \mathrm{SucCad}(\ulcorner \mathcal{L}_a \urcorner) \wedge \ell(s) = m+1 \wedge s_m = \alpha \wedge \bigwedge i \leq m(\cdots)),$$

donde los puntos suspensivos son la disyunción de las fórmulas siguientes:

1. $\bigvee t_1 t_2 (t_1, t_2 \in \mathrm{Term}_a(\mathcal{L}_a) \wedge s_i = (t_1 \ulcorner = \urcorner t_2))$,
2. $\bigvee t_1 t_2 x (t_1, t_2 \in \mathrm{Term}_a(\mathcal{L}_a) \wedge x \in \mathrm{Var}(\mathcal{L}_a) \wedge x \notin \mathrm{Vlib}(t_1 = t_2)$
   $\wedge\, s_i = \bigvee x(x + t_1 = t_2))$,
3. $\bigvee j < i \; s_i = \neg s_j$,
4. $\bigvee jk < i \; s_i = (s_j \to s_k)$
5. $\bigvee j < i \bigvee xyz \leq \alpha (x, y, z \in \mathrm{Var}(\mathcal{L}_a) \wedge s_i = \bigwedge x(\bigvee z \; z + x = y \to s_j)$.

De este modo, para todo número natural α, se cumple que

$$\alpha \text{ es una fórmula } \Delta_0 \quad \text{syss} \quad \mathbb{N} \vDash \ulcorner \alpha \urcorner \in \Delta_0.$$

En principio la fórmula $\alpha \in \Delta_0$ es Σ_1, pero los mismos argumentos empleados para $x \in \mathrm{Exp}(\mathcal{L})$ (simplificados, de hecho) prueban que es Δ_1. Es inmediato que en $\mathrm{I}\Sigma_1$ se demuestra que toda fórmula Δ_0 es de la forma $t_1 = t_2$, $t_1 \leq t_2$ (para ciertos términos sin descriptores t_1, t_2), o bien $\neg\alpha$, $\alpha \to \beta$, $\bigwedge x \leq y \, \alpha$, para ciertas fórmulas α, β de tipo Δ_0. Recíprocamente, toda fórmula construida de este modo es $\ulcorner \Delta_0 \urcorner$.

Seguidamente definimos la fórmula Δ_1:

$$\alpha \in \Sigma_1 \equiv \bigvee x\beta \leq \alpha (\beta \in \Delta_0 \wedge x \in \mathrm{Var}(\ulcorner \mathcal{L}_a \urcorner) \wedge \alpha = \bigvee x\beta).$$

Es inmediato que, para todo número natural α, se cumple que

$$\alpha \text{ es una fórmula } \Sigma_1 \text{ syss } \mathbb{N} \vDash \ulcorner \alpha \urcorner \in \Sigma_1.$$

Análogamente se define una fórmula $\alpha \in \Pi_1$, también de tipo Δ_1.

Ahora podemos definir $\alpha \in \ulcorner \mathrm{I}\Sigma_1 \urcorner$ exactamente igual que $\alpha \in \ulcorner \mathrm{AP} \urcorner$ salvo que en la parte correspondiente al esquema de inducción cambiamos $\phi \in \mathrm{Form}(\ulcorner \mathcal{L}_a \urcorner)$ por $\phi \in \Sigma_1$. Así tenemos probado que $\mathrm{I}\Sigma_1$ es demostrablemente recursiva.

Ejercicio: Definir en $\mathrm{I}\Sigma_1$ fórmulas Δ_1 que formalicen los conceptos de fórmulas Σ_1 y Π_1 respecto de la jerarquía de Lévy y usarlas para definir $\ulcorner \mathrm{KP} \urcorner$ en $\mathrm{I}\Sigma_1$. En particular se concluye que KP es demostrablemente recursiva.

Ahora estamos en condiciones de probar un caso particular del teorema de incompletitud de Gödel. La prueba no es constructiva, pero en el capítulo siguiente demostraremos resultados más generales con pruebas constructivas:

Teorema 8.20 *Sea T una teoría axiomática semirrecursiva sobre $\mathcal{L}_a$ tal que $\mathbb{N} \vDash T$. Entonces existe una sentencia Σ_1 o Π_1 verdadera en $\mathbb{N}$ y no demostrable (ni, por supuesto, refutable) en T.*

DEMOSTRACIÓN: Por el teorema de Craig podemos suponer que T es recursiva. Sea A un conjunto semirrecursivo no recursivo (en la sección 7.6 hemos mostrado varios ejemplos). Por definición de relación semirrecursiva existe una fórmula $\phi(x)$ de tipo Σ_1 tal que A está formado por los números n que cumplen $\mathbb{N} \vDash \phi(0^{(n)})$. Veamos que existe un n tal que no $\underset{T}{\vdash} \phi(0^{(n)})$ y no $\underset{T}{\vdash} \neg\phi(0^{(n)})$. En caso contrario, la función

$$f(n) = \mu d(\underset{T}{\overset{d}{\vdash}} \phi(0^{(n)}) \vee \underset{T}{\overset{d}{\vdash}} \neg\phi(0^{(n)}))$$

sería recursiva, como también lo sería la función $\chi_A(n) = \chi_R(n, f(n))$, donde R es la relación recursiva dada por

$$R(n,d) \quad \text{syss} \quad \mathbb{N} \vDash \underset{T}{\overset{0^{(d)}}{\vdash}} \phi(0^{(n)}),$$

luego A sería recursivo. Por lo tanto, existe un n tal que $\phi(0^{(n)})$ o bien $\neg\phi(0^{(n)})$ cumple lo exigido. ∎

Esto significa que ninguna teoría semirrecursiva sobre $\mathcal{L}_a$ (y en particular esto vale para AP) es capaz de demostrar todas las afirmaciones verdaderas sobre los números naturales (a menos que también demuestre afirmaciones falsas). Más concretamente, siempre existe una sentencia Σ_1 o Π_1 que no puede demostrarse ni refutarse en una teoría dada (que admita a $\mathbb{N}$ como modelo). Por el teorema 7.29 dicha sentencia puede reducirse a que una ecuación diofántica concreta tenga o no tenga solución.

La hipótesis de semirrecursividad es razonable, pues, como ya hemos comentado, si una teoría no es semirrecursiva no podemos asegurar si una sucesión de fórmulas es o no una demostración, por lo que carece de utilidad práctica. Más aún, es una hipótesis necesaria, pues la teoría que tiene por axiomas todas las sentencias de $\mathcal{L}_a$ verdaderas en $\mathbb{N}$ es trivialmente completa. Pero esto nos da una información adicional:

Teorema 8.21 *El conjunto de las sentencias de $\mathcal{L}_a$ verdaderas en $\mathbb{N}$ no es semirrecursivo.*

DEMOSTRACIÓN: Si lo fuera, la teoría que los tiene por axiomas sería incompleta por el teorema anterior, pero obviamente es completa. ∎

Esto quiere decir que no existe ningún algoritmo que permita determinar si una sentencia arbitraria de $\mathcal{L}_a$ es verdadera o falsa en $\mathbb{N}$. Más precisamente, el conjunto de las sentencias Π_1 verdaderas en $\mathbb{N}$ no es recursivo. Si lo fuera, el conjunto de las sentencias Σ_1 o Π_1 sería semirrecursivo, y la teoría con tales axiomas contradiría el teorema 8.20. Así pues, no existe un algoritmo que permita determinar si cualquier sentencia Π_1 dada es verdadera o falsa en $\mathbb{N}$.

En particular, para toda teoría T sobre $\mathcal{L}_a$ que extienda a la aritmética de Robinson Q y que cumpla $\mathbb{N} \vDash T$, existe una sentencia Π_1 verdadera en $\mathbb{N}$ y no demostrable en T (no puede ser Σ_1 porque T es Σ_1-completa, por el teorema 5.29).

8.3 La Σ_1-completitud de Q

El objetivo de esta sección es mostrar que en $\mathrm{I}\Sigma_1$ puede formalizarse y demostrarse una versión débil del teorema de Σ_1-completitud de la aritmética de Robinson Q, es decir, el teorema 5.29. Conviene trabajar en un contexto ligeramente más general:

Definición 8.22 Diremos que una teoría axiomática T sobre un lenguaje formal $\mathcal{L}$ *interpreta* a Q, IA, $\mathrm{I}\Sigma_1$ o AP si interpreta a $\mathcal{L}_a$ en el sentido de la definición 3.28 y permite demostrar las traducciones de (las clausuras universales de) los axiomas de la teoría correspondiente.

Según el teorema 3.30, esto implica que en T se demuestran las traducciones de todas las sentencias demostrables en la teoría que interpreta.

En lugar de considerar Q, consideraremos una teoría axiomática semirrecursiva T sobre un lenguaje $\mathcal{L}$ que interprete a Q. Esto significa que tenemos unas expresiones $x \in \mathbb{N}$, 0, Sx, $x+y$, $x \cdot y$ de $\mathcal{L}$ de modo que en T se demuestra:

1. $0 = (x|x = x) \in \mathbb{N}$,
2. $\bigwedge x \in \mathbb{N}\ Sx \in \mathbb{N}$,
3. $\bigwedge xy \in \mathbb{N}\ x + y \in \mathbb{N}$,
4. $\bigwedge xy \in \mathbb{N}\ x \cdot y \in \mathbb{N}$,

así como las traducciones $\overline{Q1}, \ldots, \overline{Q7}$ de todos los axiomas de Q. Podemos suponer sin pérdida de generalidad que las variables de $\mathcal{L}$ son las mismas que las de $\mathcal{L}_a$.

Notemos que todo cuanto digamos vale en particular en el caso en que T es la propia Q y $x \in \mathbb{N} \equiv x = x$.

Ahora, en $\mathrm{I}\Sigma_1$ podemos considerar el lenguaje formal $\ulcorner\mathcal{L}\urcorner$ y la teoría axiomática semirrecursiva $\ulcorner T\urcorner$ en el sentido explicado en la sección anterior. El teorema 5.57 nos da un término $0^{(x)}$ de tipo Δ_1 que cumple las propiedades siguientes:

$$0^{(0)} = 0 \wedge \bigwedge x \in \mathbb{N}\ 0^{(x+1)} = \ulcorner S\urcorner\frown 0^{(x)}.$$

Una simple inducción prueba que $\bigwedge x \in \mathbb{N}\ 0^{(x)} \in \mathrm{Term}(\ulcorner\mathcal{L}\urcorner)$.

Notemos que el término metamatemático $0^{(x)}$ tiene a x como variable libre, pero como término matemático es un designador (una cadena de signos formada únicamente por la constante $\ulcorner 0\urcorner$ y varios funtores $\ulcorner S\urcorner$), y así, podemos probar $\bigwedge x \in \mathbb{N}\ \mathrm{Vlib}(0^{(x)}) = \varnothing$. El lector debería asegurarse de que comprende que no hay contradicción en esto que decimos.

También es fácil probar lo siguiente:

Teorema 8.23 $\bigwedge n \in \mathbb{N} \underset{\ulcorner T\urcorner}{\vdash} 0^{(n)} \in \mathbb{N}$.

DEMOSTRACIÓN: Se trata de una fórmula Σ_1, luego podemos probarla por inducción sobre n. Para $n = 0$ se reduce a $\underset{\ulcorner T\urcorner}{\vdash} \ulcorner 0 \in \mathbb{N}\urcorner$, que se cumple porque $\underset{T}{\vdash} 0 \in \mathbb{N}$.

Supongamos que $\underset{\ulcorner T\urcorner}{\vdash} 0^{(n)} \in \mathbb{N}$. Entonces, como $\underset{\ulcorner T\urcorner}{\vdash} \ulcorner \bigwedge x \in \mathbb{N}\, Sx \in \mathbb{N}\urcorner$, usando la versión formalizada de EG es claro que $\underset{\ulcorner T\urcorner}{\vdash} S0^{(n)} \in \mathbb{N}$, pero esto es exactamente lo mismo que $\underset{\ulcorner T\urcorner}{\vdash} 0^{(n+1)} \in \mathbb{N}$. ∎

A continuación demostramos la versión formal de la generalización del metateorema 5.2 a este contexto. Notemos que lo que eran cinco esquemas teoremáticos que comprendían infinitos teoremas cada uno, se convierten ahora en cinco teoremas concretos:

Teorema 8.24 *Se cumple:*

1. $\bigwedge mn \in \mathbb{N}\ \underset{\ulcorner T\urcorner}{\vdash} 0^{(m+n)} = 0^{(m)} + 0^{(n)}$,
2. $\bigwedge mn \in \mathbb{N}\ \underset{\ulcorner T\urcorner}{\vdash} 0^{(mn)} = 0^{(m)} \cdot 0^{(n)}$,
3. $\bigwedge mn \in \mathbb{N}\ (m \neq n \to \underset{\ulcorner T\urcorner}{\vdash} 0^{(m)} \neq 0^{(n)})$,
4. $\bigwedge mn \in \mathbb{N}\ (m \leq n \to \underset{\ulcorner T\urcorner}{\vdash} 0^{(m)} \leq 0^{(n)})$.
5. $\bigwedge mn \in \mathbb{N}\ (n < m \to \underset{\ulcorner T\urcorner}{\vdash} \neg 0^{(m)} \leq 0^{(n)})$.

DEMOSTRACIÓN: Probamos 1) por inducción sobre n. Esto es correcto porque la fórmula $\phi(n) \equiv \bigwedge m \in \mathbb{N}\ \underset{\ulcorner T\urcorner}{\vdash} 0^{(m+n)} = 0^{(m)} + 0^{(n)}$ es ciertamente Σ_1. Para $n = 0$ hay que probar que

$$\bigwedge m \in \mathbb{N}\ \underset{\ulcorner T\urcorner}{\vdash} 0^{(m)} = 0^{(m)} + 0.$$

Esto se demuestra aplicando (entre otras cosas) la versión formalizada de EG al axioma $\ulcorner\overline{\text{Q4}}\urcorner$.

Similarmente, si suponemos $\phi(n)$, es decir, que tenemos una prueba en $\ulcorner T\urcorner$ de $0^{(m)} + 0^{(n)} = 0^{(m+n)}$, aplicando las reglas de inferencia oportunas la demostración puede prolongarse hasta otra de $S0^{(m+n)} = S(0^{(m)} + 0^{(n)})$ y, usando $\ulcorner\overline{\text{Q5}}\urcorner$, hasta una demostración de $S0^{(m+n)} = 0^{(m)} + S0^{(n)}$, pero, por definición de $0^{(x)}$, esta sentencia es la misma que $0^{(m+n+1)} = 0^{(m)} + 0^{(n+1)}$, luego se cumple $\phi(n+1)$.

El apartado 2) es análogo. En 3) suponemos $m < n$ y probamos por inducción sobre i que

$$\bigwedge i \leq m \underset{\ulcorner T\urcorner}{\vdash} (0^{(m)} = 0^{(n)} \to 0^{(m-i)} = 0^{(n-i)}),$$

que es una fórmula Σ_1. El paso de i a $i+1$ se hace aplicando $\ulcorner\overline{\text{Q2}}\urcorner$. Aplicando esto a $i = m$ llegamos a que

$$\underset{\ulcorner T\urcorner}{\vdash} (0^{(m)} = 0^{(n)} \to 0 = 0^{(n-m)}),$$

pero, por definición de $0^{(x)}$, tenemos que $0^{(n-m)} = S0^{(n-m-1)}$, por lo que

$$\underset{\ulcorner T\urcorner}{\vdash} (0^{(m)} = 0^{(n)} \to 0 = S0^{(n-m-1)}),$$

y usando $\ulcorner\overline{\text{Q1}}\urcorner$ llegamos a que

$$\underset{\ulcorner T\urcorner}{\vdash} 0^{(m)} \neq 0^{(n)},$$

como queríamos probar. Los apartados 4) y 5) no ofrecen dificultad. ∎

El teorema 5.3 se convierte ahora en:

Teorema 8.25 *Se cumple:*

1. $\underset{\ulcorner T\urcorner}{\vdash} \bigwedge xy \in \mathbb{N}(x + y = 0 \to x = 0 \wedge y = 0)$,
2. $\underset{\ulcorner T\urcorner}{\vdash} \bigwedge xy \in \mathbb{N}(xy = 0 \to x = 0 \vee y = 0)$,
3. $\underset{\ulcorner T\urcorner}{\vdash} \bigwedge x \in \mathbb{N}\ 0 \leq x$,
4. $\bigwedge n \in \mathbb{N}\ \underset{\ulcorner T\urcorner}{\vdash} \bigwedge x \in \mathbb{N}(x + 1 \leq 0^{(n+1)} \to x \leq 0^{(n)})$,
5. $\bigwedge n \in \mathbb{N}\ \underset{\ulcorner T\urcorner}{\vdash} \bigwedge x \in \mathbb{N}((x + 1) + 0^{(n)} = x + 0^{(n+1)})$.

Demostración: Las tres primeras afirmaciones son triviales. Por ejemplo, como $\underset{\text{Q}}{\vdash} \bigwedge xy(x + y = 0 \to x = 0 \wedge y = 0)$ y T interpreta a Q, tenemos también que $\underset{T}{\vdash} \bigwedge xy \in \mathbb{N}(x + y = 0 \to x = 0 \wedge y = 0)$, y esto implica 1).

El argumento metamatemático de 5.3 para el apartado 4) se generaliza y formaliza sin dificultad, mientras que 5) requiere una simple inducción, que en la prueba de 5.3 era metamatemática y ahora es una inducción formalizada respecto de una fórmula Σ_1. ∎

Pasamos ahora al teorema 5.4:

Teorema 8.26 *Se cumple*

1. $\bigwedge n \in \mathbb{N} \underset{\ulcorner T\urcorner}{\vdash} \bigwedge x \in \mathbb{N}(x \leq 0^{(n)} \leftrightarrow x = 0^{(0)} \vee x = 0^{(1)} \vee \cdots \vee x = 0^{(n)})$,
2. $\bigwedge n \in \mathbb{N} \underset{\ulcorner T\urcorner}{\vdash} \bigwedge x \in \mathbb{N}(0^{(n)} \leq x \leftrightarrow 0^{(n)} = x \vee 0^{(n+1)} \leq x)$,
3. $\bigwedge n \in \mathbb{N} \underset{\ulcorner T\urcorner}{\vdash} \bigwedge x \in \mathbb{N}(x \leq 0^{(n)} \vee 0^{(n)} \leq x)$.

DEMOSTRACIÓN: 1) Para probar que

$$\bigwedge n \in \mathbb{N} \underset{\ulcorner T \urcorner}{\vdash} \bigwedge x(\bigvee_{i\leq n} x = 0^{(i)} \to x \leq 0^{(n)})$$

razonamos por inducción sobre n que

$$\bigwedge n \in \mathbb{N} \bigwedge m \in \mathbb{N}(n \leq m \to \underset{\ulcorner T \urcorner}{\vdash} \bigwedge x(\bigvee_{i\leq n} x = 0^{(i)} \to x \leq 0^{(m)}).$$

Para $n = 0$ hay que probar que $\bigwedge m \in \mathbb{N} \underset{\ulcorner T \urcorner}{\vdash} \bigwedge x(x = 0 \to x \leq 0^{(m)})$, lo cual no ofrece dificultad.

Supuesto que $\bigwedge m \in \mathbb{N}(n \leq m \to \underset{\ulcorner T \urcorner}{\vdash} \bigwedge x(\bigvee_{i\leq n} x = 0^{(i)} \to x \leq 0^{(m)})$, suponemos que $n + 1 \leq m$, con lo que tenemos $\underset{\ulcorner T \urcorner}{\vdash} \bigwedge x(\bigvee_{i\leq n} x = 0^{(i)} \to x \leq 0^{(m)})$ por hipótesis de inducción, y basta observar que $\underset{\ulcorner T \urcorner}{\vdash} x = 0^{(n+1)} \to x \leq 0^{(m)}$, con lo que

$$\underset{\ulcorner T \urcorner}{\vdash} \bigvee_{i\leq n} x = 0^{(i)} \vee x = 0^{(n+1)} \to x \leq 0^{(m)}.$$

Esto es lo mismo que $\underset{\ulcorner T \urcorner}{\vdash} \bigvee_{i\leq n+1} x = 0^{(i)} \to x \leq 0^{(m)}$.

Demostramos la implicación contraria por inducción sobre n. Para $n = 0$ se reduce a

$$\underset{\ulcorner T \urcorner}{\vdash} \bigwedge x \in \mathbb{N}(x \leq 0 \leftrightarrow x = 0),$$

lo cual se sigue del apartado 1) del teorema anterior, teniendo en cuenta que $x \leq 0 \equiv \bigvee z \in \mathbb{N}\ z + x = 0$.

Supuesto cierto para n, vamos a construir una demostración en $\ulcorner T \urcorner$ usando el teorema de deducción, es decir, vamos a suponer $x \in \mathbb{N} \wedge x \leq 0^{(n+1)}$ como premisa. Distinguimos entonces dos casos: si $x = 0$, entonces, sabemos que

$$x = 0 \to x = 0 \vee \bigvee_{i=1}^{n} x = 0^{(i)} \to \bigvee_{i\leq n} x = 0^{(i)}.$$

Suponemos en segundo lugar (siempre razonando en $\ulcorner T \urcorner$) que $x = y + 1$, con lo que el apartado d) del teorema anterior nos da que $y \leq 0^{(n)}$, luego la hipótesis de inducción nos da $\bigvee_{i\leq n} y = 0^{(i)}$.

En este punto probamos por inducción sobre j que

$$\underset{\ulcorner T \urcorner}{\vdash} \bigvee_{i\leq j} y = 0^{(i)} \to \bigvee_{i\leq j+1} Sy = 0^{(i)}.$$

En efecto, para $j = 0$ se reduce a

$$\underset{\ulcorner T \urcorner}{\vdash} (y = 0 \to Sy = 0 \vee Sy = 0^{(1)}),$$

que se prueba sin dificultad. Si vale para j, observamos que $\bigvee_{i\leq j+1} y = 0^{(i)}$ es $\bigvee_{i\leq j} y = 0^{(i)} \vee y = 0^{(j+1)}$, luego basta probar que

$$\underset{\ulcorner T\urcorner}{\vdash} \bigvee_{i\leq j} y = 0^{(i)} \to \bigvee_{i\leq j+2} Sy = 0^{(i)}, \qquad \underset{\ulcorner T\urcorner}{\vdash} y = 0^{(j+1)} \to \bigvee_{i\leq j+2} Sy = 0^{(i)}.$$

Por la hipótesis de inducción, y puesto que $\vdash y = 0^{(j+1)} \to Sy = 0^{(j+2)}$, basta probar que

$$\underset{\ulcorner T\urcorner}{\vdash} \bigvee_{i\leq j+1} Sy = 0^{(i)} \to \bigvee_{i\leq j+2} Sy = 0^{(i)}, \qquad \underset{\ulcorner T\urcorner}{\vdash} Sy = 0^{(j+2)} \to \bigvee_{i\leq j+2} Sy = 0^{(i)},$$

pero ambos resultados son triviales, pues $\bigvee_{i\leq j+2} Sy = 0^{(i)}$ no es sino la disyunción $\bigvee_{i\leq j+1} Sy = 0^{(i)} \vee Sy = 0^{(j+2)}$.

Por lo tanto, volviendo a nuestra prueba, como tenemos $\bigvee_{i\leq n} y = 0^{(i)}$, podemos concluir $\bigvee_{i\leq n+1} x = 0^{(i)}$, como queríamos probar.

La formalización de los otros dos apartados a partir de los correspondientes en 5.4 no ofrece ninguna dificultad. ∎

Finalmente podemos probar el resultado principal de esta sección:

Teorema 8.27 (Formalización de la Σ_1-completitud de Q) *Si T es una teoría axiomática semirrecursiva que interpreta a* Q*, para cada sentencia α de $\mathcal{L}_a$ de tipo Σ_1, se cumple que*

$$\underset{\mathrm{I}\Sigma_1}{\vdash} (\alpha \to \underset{\ulcorner T\urcorner}{\vdash} \ulcorner\bar{\alpha}\urcorner),$$

donde $\bar{\alpha}$ es la traducción de α a $\mathcal{L}$.

Demostración: Vamos a probar en realidad una versión más general de este resultado y más próxima al enunciado original de 5.6. En dicho enunciado (y en su demostración) aparecen expresiones de la forma $\alpha(0^{(k_1)}, \ldots, 0^{(k_n)})$, y para formalizarlas necesitamos el concepto de sustitución simultánea definido en 8.9. Más concretamente, fijamos un conjunto finito de variables distintas de $\mathcal{L}_a$, digamos $x_0, \ldots, x_n, x, y, z$, y definimos

$$d \equiv \{\ulcorner x_0\urcorner, \ldots, \ulcorner x_n\urcorner, \ulcorner x\urcorner, \ulcorner y\urcorner, \ulcorner z\urcorner\}, \quad e \equiv \{0^{(x_0)}, \ldots, 0^{(x_n)}, 0^{(x)}, 0^{(y)}, 0^{(z)}\},$$

$$v = \{(\ulcorner x_0\urcorner, 0^{(x_0)}), \ldots, (\ulcorner x_n\urcorner, 0^{(x_n)}), (\ulcorner x\urcorner, 0^{(x)}), (\ulcorner y\urcorner, 0^{(y)}), (\ulcorner z\urcorner, 0^{(z)})\},$$

donde los numerales son los de $\mathcal{L}$, no los de $\mathcal{L}_a$.

Tomamos[10] además $y \equiv \mathcal{P}d$, que cumple los requisitos de la definición de $\mathtt{S}_a\theta$. Lo que vamos a probar es que para cada fórmula $\alpha(x_0, \ldots, x_n)$ de tipo Σ_1 cuyas variables estén entre las indicadas,

$$\underset{\mathrm{I}\Sigma_1}{\vdash} \bigwedge x_0 \cdots x_n \, (\alpha \to \underset{\ulcorner T \urcorner}{\vdash} \mathtt{S}_d{}^{\ulcorner}\bar{\alpha}^{\urcorner}).$$

Si α es una sentencia, entonces sabemos que $\mathtt{S}_d{}^{\ulcorner}\bar{\alpha}^{\urcorner} = \bar{\alpha}$, por lo que este hecho se reduce al enunciado del teorema.

Empezamos demostrando que si $t(x_0, \ldots, x_n)$ es un término sin descriptores de $\mathcal{L}_a$ entonces

$$\underset{\mathrm{I}\Sigma_1}{\vdash} \bigwedge x x_0 \cdots x_n \, (x = t \to \underset{\ulcorner T \urcorner}{\vdash} \mathtt{S}_d{}^{\ulcorner} x = \bar{t}^{\urcorner})$$

o, equivalentemente,

$$\underset{\mathrm{I}\Sigma_1}{\vdash} \bigwedge x x_0 \cdots x_n \, (x = t \to \underset{\ulcorner T \urcorner}{\vdash} 0^{(x)} = \mathtt{S}_d{}^{\ulcorner}\bar{t}^{\urcorner}).$$

Por inducción (metamatemática) sobre t. Si $t \equiv x_i$, lo que hay que probar es que

$$\underset{\mathrm{I}\Sigma_1}{\vdash} \bigwedge x x_0 \cdots x_n \, (x = x_i \to \underset{\ulcorner T \urcorner}{\vdash} 0^{(x)} = 0^{(x_i)}),$$

pero esto es trivial, pues se trata de la regla de inferencia I.

Si $t \equiv 0$, entonces hay que probar

$$\underset{\mathrm{I}\Sigma_1}{\vdash} \bigwedge x x_0 \cdots x_n \, (x = 0 \to \underset{\ulcorner T \urcorner}{\vdash} 0^{(x)} = 0),$$

que también es un caso particular de I.

Si $t \equiv St_0$, por hipótesis de inducción, si $y = t$, se cumple $\underset{\ulcorner T \urcorner}{\vdash} 0^{(y)} = \mathtt{S}_d{}^{\ulcorner}\bar{t}^{\urcorner}$, pero de aquí se sigue claramente (por la versión formalizada de ETI) que $\underset{\ulcorner T \urcorner}{\vdash} {}^{\ulcorner}S^{\urcorner}0^{(y)} = {}^{\ulcorner}S^{\urcorner}\mathtt{S}_d{}^{\ulcorner}\bar{t}^{\urcorner}$, y esto es exactamente lo mismo que $\underset{\ulcorner T \urcorner}{\vdash} 0^{(y+1)} = \mathtt{S}_d{}^{\ulcorner}\overline{St}^{\urcorner}$. Por lo tanto, si suponemos que $x = St = y + 1$, se cumple que $\underset{\ulcorner T \urcorner}{\vdash} 0^{(x)} = \mathtt{S}_d{}^{\ulcorner}\overline{St}^{\urcorner}$.

Similarmente, si suponemos que, si $y = t_1$ y $z = t_2$ implican $\underset{\ulcorner T \urcorner}{\vdash} 0^{(y)} = \mathtt{S}_d{}^{\ulcorner}\bar{t}_1{}^{\urcorner}$) y $\underset{\ulcorner T \urcorner}{\vdash} 0^{(z)} = \mathtt{S}_d{}^{\ulcorner}\bar{t}_2{}^{\urcorner}$), respectivamente, entonces

$$\underset{\ulcorner T \urcorner}{\vdash} 0^{(y)} + 0^{(z)} = \mathtt{S}_d{}^{\ulcorner}\bar{t}_1{}^{\urcorner} + \mathtt{S}_d{}^{\ulcorner}\bar{t}_2{}^{\urcorner},$$

pero por 8.24 sabemos que $\underset{\ulcorner T \urcorner}{\vdash} 0^{(y+z)} = 0^{(y)} + 0^{(z)}$, luego, por la versión formalizada de TI, si $x = y + z$ tenemos que $\underset{\ulcorner T \urcorner}{\vdash} 0^{(x)} = \mathtt{S}_d{}^{\ulcorner}\overline{t_1 + t_2}^{\urcorner}$. El caso para $t = t_1 \cdot t_2$ es análogo, y esto termina la prueba.

[10] Notemos que podemos definir y enumerando explícitamente sus elementos, con lo que es claramente un término Δ_1.

Consideremos ahora una fórmula α de $\mathcal{L}_a$ de tipo Δ_0 y veamos que

$$\underset{\mathrm{I}\Sigma_1}{\vdash} \bigwedge x_0 \cdots x_n \, ((\alpha \to \underset{\ulcorner T \urcorner}{\vdash} \mathtt{S}_d^{\ulcorner \bar{\alpha} \urcorner}) \wedge (\neg\alpha \to \underset{\ulcorner T \urcorner}{\vdash} \mathtt{S}_d^{\ulcorner \neg\bar{\alpha} \urcorner})).$$

Razonamos por inducción sobre la longitud de α. Dados dos términos aritméticos t_1 y t_2 (siempre razonando en $\mathrm{I}\Sigma_1$) suponemos $x = t_1$, $y = t_2$, con lo que, según acabamos de probar,

$$\underset{\ulcorner T \urcorner}{\vdash} 0^{(x)} = \mathtt{S}_d^{\ulcorner \bar{t}_1 \urcorner} \wedge \underset{\ulcorner T \urcorner}{\vdash} 0^{(y)} = \mathtt{S}_d^{\ulcorner \bar{t}_2 \urcorner}.$$

Si $t_1 = t_2$, entonces $x = y$, mientras que si $t_1 \neq t_2$, entonces $x \neq y$, y por la regla de identidad en el primer caso y 8.24 en el segundo, tenemos que

$$\underset{\ulcorner T \urcorner}{\vdash} 0^{(x)} = 0^{(y)} \quad \text{o bien} \quad \underset{\ulcorner T \urcorner}{\vdash} 0^{(x)} \neq 0^{(y)},$$

luego por las reglas de simetría y transitividad de la igualdad (formalizadas en $\mathrm{I}\Sigma_1$) concluimos que

$$\underset{\ulcorner T \urcorner}{\vdash} \mathtt{S}_d^{\ulcorner \bar{t}_1 \urcorner} = \mathtt{S}_d^{\ulcorner \bar{t}_2 \urcorner} \quad \text{o bien} \quad \underset{\ulcorner T \urcorner}{\vdash} \mathtt{S}_d^{\ulcorner \bar{t}_1 \urcorner} \neq \mathtt{S}_d^{\ulcorner \bar{t}_2 \urcorner}$$

según el caso. Esto equivale a

$$\underset{\ulcorner T \urcorner}{\vdash} \mathtt{S}_d^{\ulcorner \overline{t_1 = t_2} \urcorner} \quad \text{o bien} \quad \underset{\ulcorner T \urcorner}{\vdash} \mathtt{S}_d^{\ulcorner \overline{t_1 \neq t_2} \urcorner}$$

y prueba el resultado para $\alpha \equiv t_1 = t_2$. Para $\alpha \equiv t_1 \leq t_2$ la prueba es similar.

Por el planteamiento de la inducción, el caso $\alpha \equiv \neg\beta$ es trivial. El caso en que $\alpha \equiv \beta \to \gamma$ se prueba siguiendo literalmente el argumento de 5.6. Veamos con más detalle el caso $\alpha \equiv \bigwedge x_i \leq x_j \, \beta$. Entonces

$$\mathtt{S}_d^{\ulcorner \bar{\alpha} \urcorner} = \bigwedge^{\ulcorner x_i \urcorner} \leq 0^{(x_j)} \, \mathtt{S}_{d\setminus\{\ulcorner x_i \urcorner\}}^{\ulcorner \beta \urcorner}.$$

Suponemos $\bigwedge x_i \leq x_j \, \beta$, con lo que, por hipótesis de inducción, para todo $x_i \leq x_j$ se cumple

$$\underset{\ulcorner T \urcorner}{\vdash} \mathtt{S}_d^{\ulcorner \bar{\beta} \urcorner},$$

pero sabemos que $\mathtt{S}_d^{\ulcorner \bar{\beta} \urcorner} = \mathtt{S}_{\ulcorner x_i \urcorner}^{0^{(x_i)}} \mathtt{S}_{d\setminus\{\ulcorner x_i \urcorner\}}^{\ulcorner \bar{\beta} \urcorner}$, por lo que, por la versión formalizada de la regla de Introducción del Igualador:

$$\underset{\ulcorner T \urcorner}{\vdash} \ulcorner x_i \urcorner = 0^{(x_i)} \to \mathtt{S}_{d\setminus\{\ulcorner x_i \urcorner\}}^{\ulcorner \bar{\beta} \urcorner}.$$

Por 8.26 tenemos que

$$\underset{\ulcorner T \urcorner}{\vdash} \bigwedge^{\ulcorner x_i \urcorner} \in \mathbb{N} \, (\ulcorner x_i \urcorner \leq 0^{(x_j)} \leftrightarrow \ulcorner x_i \urcorner = 0^{(0)} \vee \ulcorner x_i \urcorner = 0^{(1)} \vee \cdots \vee \ulcorner x_i \urcorner = 0^{(x_j)}).$$

Una simple inducción sobre x_j prueba ahora que

$$\underset{\ulcorner T \urcorner}{\vdash} \bigwedge^{\ulcorner x_i \urcorner} \in \mathbb{N} \, (\ulcorner x_i \urcorner \leq 0^{(x_j)} \to \mathtt{S}_{d\setminus\{\ulcorner x_i \urcorner\}}^{\ulcorner \bar{\beta} \urcorner}),$$

pero esto es lo mismo que $\bigwedge\ulcorner x_i\urcorner\ \mathtt{S}_{d\setminus\{\ulcorner x_i\urcorner\}}(\ulcorner x_i\urcorner \in \mathbb{N} \to (\ulcorner x_i\urcorner \leq 0^{(x_j)} \to \ulcorner\bar{\beta}\urcorner))$ y que

$$\mathtt{S}_d\bigwedge\ulcorner x_i\urcorner \in \mathbb{N}\ (\ulcorner x_i\urcorner \leq 0^{(x_j)} \to \ulcorner\beta\urcorner),$$

que no es sino $\mathtt{S}_d\ulcorner\bar{\alpha}\urcorner$.

Si suponemos $\neg\bigwedge x_i \leq x_j\beta$, existe un $x_i \leq x_j$ tal que $\neg\beta$ y, por hipótesis de inducción, $\underset{\ulcorner T\urcorner}{\vdash} \mathtt{S}_d\ulcorner\neg\bar{\beta}\urcorner$, que es lo mismo que $\mathtt{S}^{0^{(x_i)}}_{\ulcorner x_i\urcorner}\mathtt{S}_{d\setminus\{\ulcorner x_i\urcorner\}}\ulcorner\neg\bar{\beta}\urcorner$. Más aún,

$$\underset{\ulcorner T\urcorner}{\vdash} 0^{(x_i)} \in \mathbb{N} \wedge 0^{(x_i)} \leq 0^{(x_j)} \wedge \mathtt{S}^{0^{(x_i)}}_{\ulcorner x_i\urcorner}\mathtt{S}_{d\setminus\{\ulcorner x_i\urcorner\}}\ulcorner\neg\bar{\beta}\urcorner,$$

que es lo mismo que

$$\underset{\ulcorner T\urcorner}{\vdash} \mathtt{S}^{0^{(x_i)}}_{\ulcorner x_i\urcorner}(\ulcorner x_i\urcorner \in \mathbb{N} \wedge \ulcorner x_i\urcorner \leq 0^{(x_j)} \wedge \mathtt{S}_{d\setminus\{\ulcorner x_i\urcorner\}}\ulcorner\neg\bar{\beta}\urcorner).$$

Por la versión formalizada de IP concluimos que

$$\underset{\ulcorner T\urcorner}{\vdash} \bigvee\ulcorner x_i\urcorner(\ulcorner x_i\urcorner \in \mathbb{N} \wedge \ulcorner x_i\urcorner \leq 0^{(x_j)} \wedge \mathtt{S}_{d\setminus\{\ulcorner x_i\urcorner\}}\ulcorner\neg\bar{\beta}\urcorner).$$

La versión formalizada de un razonamiento lógico elemental nos da

$$\underset{\ulcorner T\urcorner}{\vdash} \neg\bigwedge\ulcorner x_i\urcorner(\ulcorner x_i\urcorner \in \mathbb{N} \to (\ulcorner x_i\urcorner \leq 0^{(x_j)} \to \mathtt{S}_{d\setminus\{\ulcorner x_i\urcorner\}}\ulcorner\neg\bar{\beta}\urcorner)),$$

que es lo mismo que

$$\underset{\ulcorner T\urcorner}{\vdash} \mathtt{S}_d\neg\bigwedge\ulcorner x_i\urcorner \in \mathbb{N}\ (\ulcorner x_i\urcorner \leq 0^{(x_j)} \to \ulcorner\neg\bar{\beta}\urcorner),$$

y esta fórmula es $\mathtt{S}_d\ulcorner\neg\bar{\alpha}\urcorner$.

Esto termina la prueba del resultado para fórmulas Δ_0. Por último, consideremos una fórmula Σ_1, de la forma $\alpha \equiv \bigvee x_i\beta$, donde β es Δ_0. Si suponemos α, existe un x_i tal que β, luego, según lo que hemos probado, $\underset{\ulcorner T\urcorner}{\vdash} \mathtt{S}_d\ulcorner\bar{\beta}\urcorner$, que es lo mismo que $\underset{\ulcorner T\urcorner}{\vdash} \mathtt{S}^{0^{(x_i)}}_{\ulcorner x_i\urcorner}\mathtt{S}_{d\setminus\{\ulcorner x_i\urcorner\}}\ulcorner\bar{\beta}\urcorner$. De hecho, tenemos que

$$\underset{\ulcorner T\urcorner}{\vdash} 0^{(x_i)} \in \mathbb{N} \wedge \mathtt{S}^{0^{(x_i)}}_{\ulcorner x_i\urcorner}\mathtt{S}_{d\setminus\{\ulcorner x_i\urcorner\}}\ulcorner\bar{\beta}\urcorner,$$

que es lo mismo que

$$\underset{\ulcorner T\urcorner}{\vdash} \mathtt{S}^{0^{(x_i)}}_{\ulcorner x_i\urcorner}(\ulcorner x_i\urcorner \in \mathbb{N} \wedge \mathtt{S}_{d\setminus\{\ulcorner x_i\urcorner\}}\ulcorner\bar{\beta}\urcorner).$$

Por IP resulta que $\underset{\ulcorner T\urcorner}{\vdash} \bigvee\ulcorner x_i\urcorner\ \mathtt{S}_{d\setminus\{\ulcorner x_i\urcorner\}}(\ulcorner x_i\urcorner \in \mathbb{N} \wedge \ulcorner\bar{\beta}\urcorner)$, y esto es lo mismo que $\underset{\ulcorner T\urcorner}{\vdash} \mathtt{S}_d\bigvee\ulcorner x_i\urcorner \in \mathbb{N}\ \ulcorner\bar{\beta}\urcorner$, es decir, que $\underset{\ulcorner T\urcorner}{\vdash} \mathtt{S}_d\ulcorner\bar{\alpha}\urcorner$. ∎

Como aplicación demostramos las propiedades siguientes sobre el concepto formalizado de demostración:

Teorema 8.28 (Condiciones de Hilbert-Bernays) *Sea T una teoría axiomática semirrecursiva sobre un lenguaje formal $\mathcal{L}$ que interprete a* Q *y sean ϕ, ψ fórmulas de $\mathcal{L}$. Entonces:*

1. *Si $\vdash_T \phi$, entonces $\vdash_{\mathrm{Q}} \vdash_{\ulcorner T \urcorner} \ulcorner \phi \urcorner$,*
2. $\vdash_{\mathrm{I}\Sigma_1} (\vdash_{\ulcorner T \urcorner} \ulcorner \phi \urcorner \to \vdash_{\ulcorner T \urcorner} \ulcorner \vdash_{\ulcorner T \urcorner} \ulcorner \phi \urcorner \urcorner)$,
3. $\vdash_{\mathrm{I}\Sigma_1} (\vdash_{\ulcorner T \urcorner} \ulcorner \phi \urcorner \wedge \vdash_{\ulcorner T \urcorner} \ulcorner \phi \to \psi \urcorner \to \vdash_{\ulcorner T \urcorner} \ulcorner \psi \urcorner)$.

Demostración: Como T es semirrecursiva, tenemos que la fórmula $\vdash_{\ulcorner T \urcorner} \alpha$ es[11] Σ_1 en $\mathrm{I}\Sigma_1$.

1) Si $\vdash_T \phi$, entonces $\mathbb{N} \vDash \vdash_{\ulcorner T \urcorner} \ulcorner \phi \urcorner$ y, al tratarse de una sentencia Σ_1, podemos aplicar el teorema 5.29, que nos da la conclusión.

2) Basta aplicar el teorema anterior a la sentencia $\vdash_{\ulcorner T \urcorner} \ulcorner \phi \urcorner$ de $\mathcal{L}_a$.

3) es trivial, pues se trata simplemente de la formalización en $\mathrm{I}\Sigma_1$ de la regla *modus ponens.* ∎

[11] Para los apartados 1) y 2) hemos de suponer (cosa que siempre podemos hacer) que es, por definición, una fórmula Σ_1. Si sólo suponemos que es equivalente en $\mathrm{I}\Sigma_1$ a una fórmula de tipo Σ_1, entonces hemos de cambiar Q por $\mathrm{I}\Sigma_1$ en ambos apartados.

Capítulo IX

Incompletitud

En el capítulo anterior hemos probado el teorema 8.20, que pone en evidencia serias limitaciones al estudio de los números naturales: existen afirmaciones aritméticas relativamente simples (de tipo Π_1) que son verdaderas en el modelo natural, pero no son demostrables en AP. De hecho, el resultado no se aplica únicamente a AP, sino a cualquier teoría semirrecursiva sobre $\mathcal{L}_a$ cuyos axiomas sean verdaderos en el modelo natural. En este capítulo generalizaremos este teorema haciéndolo extensivo a cualquier teoría axiomática razonable capaz de probar los resultados más básicos sobre los números naturales (los axiomas de la aritmética de Robinson Q).

En suma, veremos que es imposible construir una teoría axiomática capaz de dar respuesta a cualquier problema matemático y, más aún, probaremos que la consistencia de una teoría axiomática suficientemente "potente" no puede ser demostrada (salvo en el seno de una teoría cuya consistencia sea aún más dudosa que la de la primera). Estos resultados, debidos a Gödel, muestran que la fundamentación de la matemática no puede llevarse a cabo en los términos a los que aspiraba Hilbert, es decir, especificando una teoría general de conjuntos de la que pueda probarse que es consistente y completa.

9.1 El primer teorema de incompletitud

Si una teoría T sobre un lenguaje formal $\mathcal{L}$ interpreta a Q (es decir, interpreta a $\mathcal{L}_a$ y permite probar los axiomas de Q), representaremos con la misma notación $0^{(n)}$ los numerales de $\mathcal{L}_a$ y sus traducciones a $\mathcal{L}$. En particular, si θ es una expresión de $\mathcal{L}$, representaremos indistintamente mediante $\ulcorner\phi\urcorner$ al numeral de $\mathcal{L}_a$ que ya tenemos definido o al numeral de $\mathcal{L}$ correspondiente.

Tenemos que algunas fórmulas de $\mathcal{L}$ (al menos las fórmulas aritméticas) pueden interpretarse como afirmaciones sobre los números naturales, y por otra parte las fórmulas de $\mathcal{L}$ pueden considerarse números naturales, lo cual abre la posibilidad de que algunas fórmulas se puedan interpretar como afirmaciones sobre ellas mismas. El teorema siguiente muestra que esto es posible en un contexto muy general:

Teorema 9.1 *Sea T una teoría axiomática sobre un lenguaje $\mathcal{L}$ que interprete a* Q *y sea $\psi(x)$ una fórmula (aritmética) de $\mathcal{L}$ con x como única variable libre. Entonces existe una sentencia (aritmética) ϕ de $\mathcal{L}$ tal que*

$$\underset{T}{\vdash}(\phi \leftrightarrow \psi(\ulcorner\phi\urcorner)).$$

Demostración: Consideremos la siguiente fórmula de $\mathcal{L}_a$:

$$\sigma(x,y) \equiv x \in \mathrm{Form}(\ulcorner\mathcal{L}\urcorner) \wedge \bigvee u(\mathrm{Vlib}(x) = \{u\} \wedge y = \mathtt{S}_u^{0^{(x)}}x).$$

Claramente es de tipo Σ_1 y la relación aritmética que define en $\mathbb{N}$ es de hecho una función parcial F en el sentido del teorema 5.31. Su dominio está formado por los números naturales $\delta(u)$ que son fórmulas de $\mathcal{L}$ con una única variable libre u, y entonces $F(\delta(u)) \equiv \mathtt{S}_u^{0^{(\delta(u))}}\delta(u) \equiv \delta(\ulcorner\delta(u)\urcorner)$. Sea $\alpha(x,y)$ la fórmula Σ_1 dada por el teorema 5.31, de modo que, para toda fórmula $\delta(u)$, se cumple

$$\underset{\mathrm{Q}}{\vdash}\bigwedge y(\alpha(\ulcorner\delta(u)\urcorner, y) \leftrightarrow y = \ulcorner\delta(\ulcorner\delta(u)\urcorner)\urcorner),$$

luego, como T representa a Q, también tenemos que:

$$\underset{T}{\vdash}\bigwedge y \in \mathbb{N}(\alpha_{\mathcal{L}}(\ulcorner\delta(u)\urcorner, y) \leftrightarrow y = \ulcorner\delta(\ulcorner\delta(u)\urcorner)\urcorner),$$

donde $\alpha_{\mathcal{L}}$ es la traducción de α a $\mathcal{L}$. Tomemos concretamente

$$\delta(u) \equiv \bigvee y \in \mathbb{N}(\alpha_{\mathcal{L}}(u,y) \wedge \psi(y))$$

y consideremos la sentencia[1] $\phi \equiv \delta(\ulcorner\delta(u)\urcorner)$. Entonces, en T se cumplen las equivalencias siguientes:

$$\phi \equiv \delta(\ulcorner\delta(u)\urcorner) \equiv \bigvee y \in \mathbb{N}(\alpha_{\mathcal{L}}(\ulcorner\delta(u)\urcorner, y) \wedge \psi(y))$$

$$\leftrightarrow \bigvee y \in \mathbb{N}(y = \ulcorner\delta(\ulcorner\delta(u)\urcorner)\urcorner \wedge \psi(y)) \leftrightarrow \bigvee y \in \mathbb{N}(y = \ulcorner\phi\urcorner \wedge \psi(y)) \leftrightarrow \psi(\ulcorner\phi\urcorner).$$

■

Nota Observemos que la prueba se adapta fácilmente al caso en que la fórmula dada tenga un parámetro $\psi(x,p)$, en cuyo caso obtenemos una fórmula $\phi(p)$ tal que $\underset{T}{\vdash}\bigwedge p(\phi(p) \leftrightarrow \psi(\ulcorner\phi\urcorner, p))$.

En efecto, en estas condiciones definimos F sobre los números naturales $\delta(u,p)$ de modo que $F(\delta(u,p)) = \delta(\ulcorner\delta(u,y)\urcorner, p)$. Así obtenemos igualmente una fórmula $\alpha(x,y)$ con la que formamos $\delta(u,p) \equiv \bigvee y \in \mathbb{N}(\alpha_{\mathcal{L}}(u,y) \wedge \psi(y,p))$. Basta tomar $\phi(p) \equiv \delta(\ulcorner\delta(u,y)\urcorner, p)$. ■

[1] Notemos que si ψ es aritmética, es decir si es la traducción de una fórmula ψ_0 de $\mathcal{L}_a$, entonces $\delta(u)$ es la traducción de $\bigvee y(\alpha(u,y) \wedge \psi_0(y))$, luego también es aritmética. Más aún, es (la traducción de una fórmula) Σ_1 si ψ lo es. Como $\ulcorner\delta(u)\urcorner$ es un término Δ_0 (es un numeral), lo mismo vale para ϕ.

Observaciones Notemos que la fórmula $\alpha(x,y)$ que hemos construido significa "y es la sentencia que resulta de sustituir en la fórmula x su única variable libre por el numeral $0^{(x)}$", de modo que $\delta(u)$ significa "la fórmula que resulta de sustituir en la fórmula x su única variable libre por el numeral $0^{(x)}$ tiene la propiedad $\psi(y)$", y por consiguiente ϕ significa "la fórmula que resulta de sustituir en $\delta(u)$ la variable u por el numeral $\ulcorner\delta(u)\urcorner$ tiene la propiedad ψ", pero sucede que la fórmula que cumple dicha descripción es precisamente ϕ, luego así hemos conseguido una fórmula ϕ que hable de sí misma y diga cualquier cosa prefijada que queramos que diga.

También es conveniente observar que la prueba del teorema anterior es totalmente constructiva: la fórmula α se construye explícitamente a partir de la fórmula σ, y deshaciendo las definiciones sucesivas que intervienen en ella podemos expresarla explícitamente como una fórmula de tipo Σ_1, y entonces la demostración del teorema 5.6 nos proporciona una demostración explícita en Q de la equivalencia

$$\vdash_{\mathrm{Q}} \bigwedge y(\alpha(\ulcorner\delta(u)\urcorner, y) \leftrightarrow y = \ulcorner\phi\urcorner),$$

de la que a su vez se sigue por razonamientos lógicos elementales la equivalencia del enunciado. En otras palabras, podríamos programar a un ordenador para que, dándole una descripción razonable de la teoría T y la fórmula ψ, nos proporcionara una fórmula ϕ junto con una demostración en T de la equivalencia del enunciado. ∎

El interés del teorema anterior se pone de manifiesto cuando lo usamos para construir sentencias que afirmen de sí mismas hechos "comprometidos":

Definición 9.2 Si T es una teoría semirrecursiva que interprete a Q, podemos considerar la fórmula $\psi(\alpha) \equiv \neg \vdash_{\ulcorner T\urcorner} \alpha$ de $\mathcal{L}_a$, así como su traducción a $\mathcal{L}$, que representaremos con la misma notación. Por el teorema anterior, existe una sentencia aritmética G del lenguaje $\mathcal{L}$ de T tal que

$$\vdash_T (G \leftrightarrow \neg \vdash_{\ulcorner T\urcorner} \ulcorner G\urcorner).$$

A cualquier sentencia de $\mathcal{L}_a$ que cumpla esta propiedad se le llama *sentencia de Gödel* para la teoría T. Se trata de una sentencia que afirma de sí misma que no es demostrable en T. Notemos que G es Π_1 en T, porque la fórmula derecha de la equivalencia anterior es trivialmente Π_1.

Aunque no es exactamente lo mismo, el hecho de que G afirme su propia indemostrabilidad implica que ciertamente es indemostrable:

Teorema 9.3 (Primer teorema de incompletitud de Gödel) *Sea T una teoría semirrecursiva que interprete a* Q*, en la que no pueda probarse la traducción de ninguna sentencia Σ_1 de $\mathcal{L}_a$ que sea falsa en el modelo natural. Entonces la sentencia de Gödel de T no es demostrable ni refutable en T.*

Demostración: Notemos que por hipótesis hay fórmulas no demostrables en T, luego en particular estamos suponiendo que T es consistente. Vamos a probar en primer lugar que si T es semirrecursiva, consistente y extiende a Q, entonces las sentencias de Gödel de T no son demostrables en T.

En efecto, si $\underset{T}{\vdash} G$, entonces $\mathbb{N} \vDash \underset{\ulcorner T\urcorner}{\vdash} \ulcorner G\urcorner$ y, por Σ_1-completitud, $\underset{\mathrm{Q}}{\vdash} \underset{\ulcorner T\urcorner}{\vdash} \ulcorner G\urcorner$, luego $\underset{T}{\vdash} \underset{\ulcorner T\urcorner}{\vdash} \ulcorner G\urcorner$, donde la última fórmula es la traducción a $\mathcal{L}$ de la anterior, luego, por definición de sentencia de Gödel, $\underset{T}{\vdash} \neg G$, y resulta que T es contradictoria.

Así pues, G no es demostrable, luego $\mathbb{N} \vDash \neg \underset{\ulcorner T\urcorner}{\vdash} \ulcorner G\urcorner$, luego $\underset{\ulcorner T\urcorner}{\vdash} \ulcorner G\urcorner$ es una sentencia Σ_1 de $\mathcal{L}_a$ falsa en el modelo natural, luego su traducción a $\mathcal{L}$ no es demostrable en T, pero, por definición de sentencia de Gödel dicha traducción es equivalente a $\neg G$ en T, luego $\neg G$ no es demostrable en T. ∎

Observaciones El hecho más destacable de la demostración del teorema anterior es que es totalmente constructiva: dada una teoría semirrecursiva T que interprete a Q, sabemos construir explícitamente una sentencia G con la propiedad de que tenemos un algoritmo que aplicado a una hipotética demostración de G nos daría una demostración de $\neg G$, luego G no es demostrable a menos que T sea contradictoria.

A efectos de interpretar el teorema, es preferible que nos centremos en la sentencia $\tilde{G} \equiv \neg \underset{\ulcorner T\urcorner}{\vdash} \ulcorner G\urcorner$, que es equivalente a G en T, por lo que todo lo dicho vale para ella: $\tilde{G}$ no es demostrable en T a menos que la teoría T sea contradictoria. La ventaja de considerar $\tilde{G}$ es que se trata de la traducción al lenguaje de T de una sentencia de $\mathcal{L}_a$ (la que representamos con la misma notación), por lo que tiene sentido afirmar que $\mathbb{N} \vDash \tilde{G}$. En la prueba del teorema anterior hemos visto que esto es cierto, es decir, que $\tilde{G}$ es una sentencia aritmética verdadera en su interpretación natural. En cambio, su traducción a T no es demostrable en T.

Por lo tanto, el teorema de incompletitud puede parafrasearse diciendo que en toda teoría semirrecursiva consistente existe una sentencia $\tilde{G}$ aritmética (concretamente, de tipo Π_1) que es verdadera pero no demostrable (a menos que la teoría sea contradictoria) y no es refutable (a menos que la teoría permita demostrar sentencias falsas). Por el teorema 7.29, dicha sentencia es equivalente a que cierta ecuación diofántica no tenga solución.

En particular vemos que el teorema de Σ_1-completitud de Q (teorema 5.29) no puede generalizarse a fórmulas de tipo Π_1 ni para Q ni para ninguna teoría que interprete (en particular que extienda) a Q.

En realidad las hipótesis con las que Gödel demostró su teorema eran ligeramente distintas de las que hemos considerado aquí. Gödel enunció su teorema para teorías aritméticas recursivas ω-consistentes, donde la ω-consistencia significa que no existe ninguna fórmula aritmética $\phi(x)$ tal que

$$\underset{\ulcorner T\urcorner}{\vdash} \bigvee x \neg\phi(x) \qquad \text{y a la vez} \qquad \underset{\ulcorner T\urcorner}{\vdash} \phi(0^{(n)})$$

para todo número natural n. En efecto, bajo esta hipótesis también podemos demostrar que las sentencias de Gödel no son refutables. En realidad, si en lugar de trabajar en una teoría aritmética trabajamos meramente en una teoría que

interpreta a Q, necesitamos refinar ligeramente la construcción de las sentencias de Gödel. Recordemos que las hemos construido a partir de la fórmula

$$\underset{\ulcorner T\urcorner}{\vdash} \alpha \equiv \bigvee d \overset{d}{\underset{\ulcorner T\urcorner}{\vdash}} \alpha.$$

Si la teoría T es recursiva, la fórmula tras $\bigvee d$ es Δ_1, luego podemos sustituirla por otra (equivalente en $\mathrm{I}\Sigma_1$) en las condiciones del teorema 5.30. Si construimos G a partir de dicha fórmula, no sólo contamos con todos los hechos que hemos usado para probar el teorema anterior, sino que además, como sabemos que (por la mera consistencia de T) la sentencia G no es demostrable en T, ningún número natural d codifica una demostración de G, es decir, ningún d cumple

$$\mathbb{N} \vDash \overset{0^{(d)}}{\underset{\ulcorner T\urcorner}{\vdash}} \ulcorner G \urcorner,$$

y de ahí podemos pasar a que $\underset{\mathrm{Q}}{\vdash} \neg \overset{0^{(d)}}{\underset{\ulcorner T\urcorner}{\vdash}} \ulcorner G\urcorner$, luego también $\underset{T}{\vdash} \neg \overset{0^{(d)}}{\underset{\ulcorner T\urcorner}{\vdash}} \ulcorner G\urcorner$.

Por otra parte, si G fuera refutable en T, entonces, por la definición de sentencia de Gödel,

$$\underset{T}{\vdash} \underset{\ulcorner T\urcorner}{\vdash} \ulcorner G\urcorner,$$

o, lo que es lo mismo,

$$\underset{T}{\vdash} \bigvee d \in \mathbb{N} \overset{d}{\underset{\ulcorner T\urcorner}{\vdash}} \ulcorner G\urcorner.$$

En resumen, si llamamos $\phi(x) \equiv \neg \overset{x}{\underset{\ulcorner T\urcorner}{\vdash}} \ulcorner G\urcorner$, llegamos a que T no es ω-consistente. ∎

No obstante, es irrelevante considerar una hipótesis adicional u otra, porque lo cierto es que cualquiera de ellas es superflua:

Teorema 9.4 (Teorema de incompletitud (versión de Rosser)) *Toda teoría semirrecursiva consistente que interprete a* Q *es incompleta.*

Demostración: Sea T una teoría en las hipótesis del enunciado. Entonces la fórmula $\underset{\ulcorner T\urcorner}{\vdash} \alpha$ es Σ_1 en $\mathrm{I}\Sigma_1$, luego en esta teoría equivale a una fórmula[2] $\bigvee d\, \sigma(d, \alpha)$, donde σ es de tipo Δ_0. Tenemos entonces que, para toda fórmula α del lenguaje $\mathcal{L}$ de T, se cumple:

$$\underset{T}{\vdash} \alpha \quad \text{syss} \quad \mathbb{N} \vDash \bigvee d\, \sigma(d, \ulcorner \alpha \urcorner).$$

[2]Si la teoría T es recursiva, podemos tomar $\sigma(d, \alpha) \equiv \overset{d}{\underset{\ulcorner T\urcorner}{\vdash}} \alpha$, que no es Δ_0, sino Δ_1, pero si la tomamos en las condiciones del teorema 5.30, toda la demostración vale sin cambio alguno, y entonces d tiene una interpretación directa: es (un número natural que codifica) una demostración de α.

Sea R una sentencia dada por el teorema 9.1 de modo que[3]

$$\vdash_T R \leftrightarrow \bigwedge d \in \mathbb{N}\ (\sigma(d, \ulcorner R \urcorner) \to \bigvee e < d\ \sigma(e, \neg \ulcorner R \urcorner)).$$

Vamos a probar que R no es demostrable ni refutable en T. Supongamos que se cumple $\vdash_T R$. Entonces existe un d tal que $\mathbb{N} \vDash \sigma(0^{(d)}, \ulcorner R \urcorner)$, luego por Σ_1-completitud $\vdash_T \sigma(0^{(d)}, \ulcorner R \urcorner)$. Consecuentemente $\vdash_T \bigvee e < 0^{(d)}\ \sigma(e, \neg \ulcorner R \urcorner)$.

Consideramos ahora todos los números $e < d$. Si existe alguno que cumpla $\mathbb{N} \vDash \sigma(0^{(e)}, \neg \ulcorner R \urcorner)$, entonces $\vdash_T \neg R)$ y tenemos que T es contradictoria. En caso contrario, por Σ_1-completitud (recordemos que σ es Δ_0) $\vdash_T \neg\sigma(0^{(e)}, \neg \ulcorner R \urcorner)$, pero en Q (luego en T) se prueba

$$d = 0^{(0)} \vee \cdots \vee d = 0^{(e-1)},$$

luego podemos concluir que $\vdash_T \bigwedge e < 0^{(d)}\ \neg\sigma(e, \neg \ulcorner R \urcorner)$, y tenemos nuevamente una contradicción en T.

Supongamos ahora que $\vdash_T \neg R$. Entonces existe un e tal que $\mathbb{N} \vDash \sigma(0^{(e)}, \neg \ulcorner R \urcorner)$, de donde a su vez $\vdash_T \sigma(0^{(e)}, \neg \ulcorner R \urcorner)$. Por otro lado, por la construcción de R tenemos que

$$\vdash_T \bigvee d \in \mathbb{N}\ (\sigma(d, \ulcorner R \urcorner) \wedge \bigwedge e < d\ \neg\sigma(e, \neg \ulcorner R \urcorner)).$$

Razonando en T, tomamos un d que cumpla esto. En Q (luego en T) se demuestra $0^{(e)} < d \vee d \leq 0^{(e)}$, pero el primer caso lleva a una contradicción, luego tiene que ser $d \leq 0^{(e)}$. Nuevamente tenemos dos posibilidades: si algún número $i \leq e$ cumple $\mathbb{N} \vDash \sigma(0^{(i)}, \ulcorner R \urcorner)$ entonces $\vdash_T R$ y tenemos una contradicción. En caso contrario, para todo $i \leq e$ se demuestra $\neg\sigma(0^{(i)}, \ulcorner R \urcorner)$, pero $d \leq 0^{(e)}$ implica

$$d = 0^{(0)} \vee \cdots \vee d = 0^{(e)},$$

luego podemos concluir que $\neg\sigma(d, \ulcorner R \urcorner)$, y nuevamente llegamos a una contradicción. ∎

El lector debería convencerse de que la prueba del teorema anterior también es constructiva: dada una teoría semirrecursiva que interprete a Q, podemos construir explícitamente la sentencia R, y podemos programar a un ordenador para que si le damos una prueba de R o de $\neg R$ en T, nos devuelva la prueba de una contradicción en T.

De este modo, si una teoría axiomática cumple los requisitos mínimos de ser semirrecursiva y consistente (sin lo cual sería inútil en la práctica) y de demostrar unas mínimas propiedades sobre los números naturales (los axiomas de Q),

[3]Notemos que R afirma de sí misma algo así como "Si puedo ser demostrada, también puedo ser refutada (con una refutación menor que mi demostración)", por lo que indirectamente R afirma que no es demostrable supuesto que la teoría T sea consistente. Notemos también que, como la sentencia de Gödel, R es de tipo Π_1.

entonces es incompleta: ninguna teoría axiomática aceptable para un matemático puede resolver cualquier problema aritmético, en el sentido de que siempre habrá sentencias aritméticas que no podrán ser demostradas ni refutadas.

Más aún: el teorema siguiente muestra que no sólo no podemos demostrar o refutar cualquier sentencia, sino que no siempre podemos saber si una sentencia dada es demostrable o no en una teoría dada:

Teorema 9.5 *Si T es una teoría consistente que interpreta a* Q*, el conjunto de sus teoremas no es recursivo.*

DEMOSTRACIÓN: Supongamos que el conjunto de los teoremas de T es recursivo. Entonces llamamos T' a la teoría cuyos axiomas son los teoremas de T. Así T' tiene los mismos teoremas que T (y en particular es consistente e interpreta a Q), pero además es una teoría recursiva. Equivalentemente, podemos suponer que T es recursiva.

Sea $\phi(x)$ una fórmula Σ_1, que podemos tomar en las condiciones de 5.30, que defina al conjunto de los teoremas de T, es decir, que tenemos que

$$\text{si } \underset{T}{\vdash} \alpha \quad \text{entonces} \quad \underset{T}{\vdash} \phi(\ulcorner\alpha\urcorner) \quad \text{y si no } \underset{T}{\vdash} \alpha \quad \text{entonces} \quad \underset{T}{\vdash} \neg\phi(\ulcorner\alpha\urcorner).$$

El teorema 9.1 nos da una sentencia C tal que

$$\underset{T}{\vdash} C \leftrightarrow \neg\phi(\ulcorner C\urcorner).$$

Nuevamente, C es una sentencia que afirma que no puede ser demostrada. Si esto fuera cierto, es decir, si no $\underset{T}{\vdash} C$, entonces $\underset{T}{\vdash} \neg\phi(\ulcorner C\urcorner)$, pero por construcción de C esto equivale a $\underset{T}{\vdash} C$, y tenemos una contradicción. Por lo tanto, $\underset{T}{\vdash} C$, luego $\underset{T}{\vdash} \phi(\ulcorner C\urcorner)$, pero esto implica $\underset{T}{\vdash} \neg C$, y llegamos a que T es contradictoria. ∎

Tenemos así un ejemplo conceptualmente muy simple de conjunto semirrecursivo no recursivo: el conjunto de los teoremas de cualquier teoría semirrecursiva consistente que interprete a Q (por ejemplo el conjunto de los teoremas de la propia Q, o de AP). En general disponemos de un algoritmo finito para enumerar todos los teoremas de una teoría axiomática recursiva T (sólo tenemos que ir enumerando los números naturales, comprobando si cada uno de ellos codifica una demostración de la teoría y, en caso afirmativo, añadir a la lista de teoremas la conclusión de tal demostración). De este modo, si queremos confirmar que una determinada fórmula es un teorema de T, en teoría siempre tenemos el "método" (muy poco práctico) de esperar a ver si aparece en la lista, pero ahora acabamos de probar que, si una fórmula no es un teorema de una teoría semirrecursiva consistente T, no existe ningún algoritmo que nos permita confirmar en un tiempo finito que no es un teorema (lo cual no impide que en casos concretos logremos averiguarlo por uno u otro medio, pero no existe un procedimiento general que nos garantice una respuesta en cualquier caso).

El teorema anterior todavía puede generalizarse más hasta eliminar toda alusión explícita a los números naturales:

Teorema 9.6 (Teorema de Church) *No existe ningún criterio que permita distinguir en un tiempo finito si cualquier fórmula dada de un lenguaje formal es un teorema lógico, consistente o contradictoria.*

DEMOSTRACIÓN: Más concretamente, vamos a probar que si $\mathcal{L}$ es un lenguaje formal que tenga al menos un relator diádico distinto del igualador, entonces el conjunto de los teoremas lógicos del lenguaje $\mathcal{L}$ es semirrecursivo, pero no recursivo. Esto implica que no es posible distinguir mediante ningún algoritmo las fórmulas que son teoremas lógicos de las que no lo son, y por lo tanto tampoco podemos determinar si una fórmula es consistente (es decir, si su negación no es un teorema lógico) o contradictoria (si su negación es un teorema lógico).

En el capítulo siguiente (10.12) introduciremos la teoría de conjuntos NBG*, de la que probaremos que es una teoría aritmética, recursiva, consistente y finitamente axiomatizable sobre el lenguaje $\mathcal{L}_{\mathrm{tc}}$. Podemos suponerla una teoría sobre $\mathcal{L}$ sin más que identificar el relator $\in$ con cualquier relator diádico de $\mathcal{L}$ y no tener en cuenta para nada cualquier otro signo eventual. Si llamamos C a la conjunción de todos sus axiomas (podemos suponer que C es una sentencia), tenemos que

$$\underset{\mathrm{NBG}^*}{\vdash} \alpha \quad \text{syss} \quad \vdash C \to \alpha.$$

La función dada por $f(\alpha) = (C \to \alpha)$ es claramente recursiva. Si el conjunto de los teoremas lógicos de $\mathcal{L}$ fuera recursivo, también lo sería su función característica χ, así como la composición $g(\alpha) = \chi(f(\alpha))$, que es la función característica del conjunto de los teoremas de NBG*. Por lo tanto, este conjunto sería recursivo, en contradicción con el teorema anterior. ∎

Ejercicio: Probar que si una fórmula contiene únicamente relatores monádicos (aparte del igualador) entonces existe un algoritmo finito para determinar si es o no consistente o si es o no un teorema lógico. AYUDA: Probar que si una fórmula con n relatores monádicos tiene un modelo, entonces tiene un modelo con a lo sumo 2^n elementos. Para ello basta establecer una relación de equivalencia en el modelo dado en la que dos objetos de su universo son equivalentes si las n relaciones que interpretan los relatores coinciden en ellos. Las clases de equivalencia se convierten fácilmente en un modelo de la fórmula dada, y son a lo sumo 2^n. Nótese además que sólo hay un número finito de modelos "esencialmente distintos" con a lo sumo 2^n elementos y todos ellos pueden ser calculados en la práctica.

9.2 El teorema de Tarski

La hipótesis de semirrecursividad en el teorema de incompletitud es claramente necesaria. Por ejemplo, podemos considerar la extensión T de la aritmética de Peano que resulta de tomar como axiomas todas las sentencias verdaderas en su modelo natural. Es inmediato que T cumple todos los requisitos del primer teorema de incompletitud salvo quizá la semirrecursividad y, de hecho, no puede ser semirrecursiva, pues es trivialmente completa.

En particular podemos concluir que el conjunto de las afirmaciones verdaderas sobre números naturales no es recursivo o, dicho de otro modo, que no existe ningún algoritmo para determinar si una determinada afirmación sobre números naturales es verdadera o falsa. (Ya habíamos llegado a esta conclusión como consecuencia del teorema 8.20.) El teorema de Tarski va más allá:

Teorema 9.7 (Teorema de Tarski) *Sea T una teoría axiomática que interprete a* Q *y sea M un modelo de T.*

1. *No existe ninguna fórmula $V(x)$ con x como única variable libre y tal que para toda sentencia ϕ se cumpla*

$$M \models \phi \quad \textit{syss} \quad M \models V(\ulcorner\phi\urcorner).$$

2. *En particular la relación monádica dada por $V(\phi)$ syss ϕ es una sentencia verdadera en M no es recursiva.*

3. *El conjunto de los números naturales que codifican sentencias de $\mathcal{L}_a$ verdaderas en $\mathbb{N}$ no es aritmético.*

4. *Tampoco puede existir una fórmula $V(x)$ tal que para toda sentencia ϕ se cumpla $\underset{T}{\vdash}(\phi \leftrightarrow V(\ulcorner\phi\urcorner))$.*

Demostración: Supongamos que existe la fórmula $V(x)$ según 1). Entonces el teorema 9.1 nos da una sentencia τ tal que

$$\underset{T}{\vdash} \tau \leftrightarrow \neg V(\ulcorner\tau\urcorner).$$

Notemos que τ significa "yo soy falsa" (en M).

Si $M \models \tau$, entonces $M \models V(\ulcorner\tau\urcorner)$, pero, por la propiedad que define a τ tendremos también que $M \models \neg\tau$, lo cual es absurdo.

Si $M \models \neg\tau$, entonces $M \models \neg V(\ulcorner\tau\urcorner)$, luego por hipótesis $M \models \tau$, y tenemos de nuevo un imposible. Así pues, no existe tal V.

Para probar 2) observamos simplemente que si la relación $M \models \phi$ fuera semirrecursiva, por el teorema 5.30 existiría una fórmula V tal que

$$M \models \phi \quad \text{syss} \quad \underset{T}{\vdash} V(\ulcorner\phi\urcorner), \quad M \models \neg\phi \quad \text{syss} \quad \underset{T}{\vdash} \neg V(\ulcorner\phi\urcorner),$$

de donde se deduce que V cumple las condiciones de 1).

3) Es simplemente el caso particular de 1) que resulta de considerar como T la propia Q y como M el modelo natural $\mathbb{N}$.

Es claro que una fórmula que cumpla 4) también cumple 1), para cualquier modelo M de T, pero, suponiendo meramente que T es consistente, podemos obtener una prueba directa puramente sintáctica (que no involucre modelos). En efecto, construimos igualmente la sentencia τ, y ahora tenemos que $\underset{T}{\vdash}(\tau \leftrightarrow \neg\tau)$, de donde se sigue que T es contradictoria. ∎

Observaciones Vemos, pues, que el conjunto de las sentencias aritméticas verdaderas en el modelo natural no sólo no es recursivo (es decir, no sólo no es posible determinar en la práctica cuáles son exactamente sus elementos), sino que ni siquiera es aritmético (no puede ser definido mediante una fórmula del lenguaje $\mathcal{L}_a$).

Más en general, el apartado 1) del teorema dice que el conjunto de las sentencias de una teoría dada verdaderas en un modelo no es definible en dicho modelo mediante una fórmula del lenguaje de la teoría.

Por su parte, el apartado 4) afirma que, en una teoría T sobre un lenguaje $\mathcal{L}$, no es posible asignar un "significado" a cada sentencia $\alpha \in \mathrm{Form}(\ulcorner\mathcal{L}\urcorner)$ de modo que, para cada sentencia concreta (es decir, metamatemática) α, el significado asociado a $\ulcorner\alpha\urcorner$ sea precisamente α. No obstante, ya hemos visto que este impedimento teórico puede burlarse parcialmente restringiendo la definición a determinadas clases de sentencias (Σ_n, o Π_n) y más adelante veremos que también es posible burlarlo limitando el universo en que las sentencias deben ser interpretadas.

Respecto a la prueba del teorema de Tarski, hemos de destacar que se apoya esencialmente en la conocida paradoja de Epiménides, o paradoja del mentiroso. La versión clásica se remonta a Epiménides, que, para rebatir la fama de mentirosos que en la antigüedad tenían los cretenses afirmaba: "Todos los cretenses mienten"; ahora bien, Epiménides era cretense, por lo que cualquiera que le oyera tenía que admitir que, por lo menos, los cretenses dicen la verdad en algunas ocasiones, ya que si suponemos que los cretenses mienten siempre entonces la afirmación de Epiménides sería cierta, pero eso es contradictorio con que la afirme un cretense.

Depurando el argumento (siguiendo a Gödel), supongamos que un día, a las 12:01, Juan afirma: "Todo lo que hoy ha dicho Juan entre las 12:00 y las 12:02 es falso", y no dice nada más en dicho intervalo. Esta afirmación sería sin duda verdadera o falsa si la hubiera pronunciado cualquiera que no fuera Juan, o incluso si la hubiera pronunciado Juan en otro momento. Pero cuando la pronuncia Juan a las 12:01 se vuelve contradictoria.

Es la misma contradicción a la que llegamos si negamos la conclusión del teorema de Tarski: si en una teoría aritmética T podemos definir la noción de "sentencia verdadera en un modelo dado", entonces podemos construir una sentencia que diga "yo soy falsa" y tenemos la paradoja. No obstante, hemos de insistir en que la prueba no es un sofisma, sino que, muy al contrario, es totalmente constructiva: si alguien pudiera definir una fórmula $V(x)$ que cumpla el apartado 1) del teorema de Tarski, entonces sabríamos escribir una sentencia τ de la que podríamos probar tanto que es verdadera como que es falsa. Por consiguiente estamos seguros de que la fórmula V no existe.

9.3 El segundo teorema de incompletitud

El teorema de incompletitud de Gödel ha dado pie a muchas falacias en virtud de las cuales la mente humana no es susceptible de análisis lógico. El

argumento general es que, dada cualquier teoría axiomática suficientemente rica, el teorema de incompletitud nos permite conocer una sentencia verdadera pero que no es demostrable en la teoría en cuestión, de modo que nosotros sabemos más de lo que puede contener cualquier teoría axiomática. Variantes de este argumento se han empleado también contra la inteligencia artificial, es decir, para argumentar que un ordenador nunca podrá pensar como un ser humano. Como veremos enseguida, todo esto no tiene ningún fundamento.

Ciertamente estamos ante una paradoja: no habría problema en admitir la existencia de afirmaciones verdaderas sobre números naturales que no puedan ser demostradas en una teoría dada, pero algo muy distinto es que sepamos encontrarlas explícitamente, es decir, que podamos señalar sentencias concretas de las que sepamos demostrar que son verdaderas pero no demostrables. Consideremos, por ejemplo, la sentencia de Gödel G para la aritmética de Peano, AP. Como sabemos que AP es consistente, sabemos que $\mathbb{N} \models G$, pero también que G no puede probarse a partir de los axiomas de Peano. Ahora bien, para probar el teorema de incompletitud, ¿hemos usado alguna propiedad extraña sobre los números naturales, algo que no se deduzca de los axiomas de Peano? La respuesta es negativa, pero entonces, ¿cómo hemos podido llegar a probar algo que no se deduce de los axiomas de Peano?

Esta paradoja desaparece en cuanto nos damos cuenta de que el teorema de incompletitud no dice que la sentencia de Gödel sea verdadera y no demostrable. Sólo dice que SI la teoría axiomática interpreta a Q y es semirrecursiva y consistente, entonces G es verdadera y no demostrable. La recursividad es una propiedad muy fácil de comprobar y que satisface cualquier teoría razonable, y también es fácil constatar que una teoría interpreta a Q, así que la cuestión se reduce a que SI la teoría T es consistente, ENTONCES G no es demostrable. Notemos que el recíproco es trivialmente cierto. Esto es lo que realmente hemos demostrado para una teoría semirrecursiva T que interprete a Q. Para desarrollar esta idea conviene introducir una definición:

Definición 9.8 Sea T una teoría axiomática semirrecursiva que interprete a Q. Llamaremos

$$\mathrm{Consis}\,\ulcorner T\urcorner \equiv \neg \underset{\ulcorner T\urcorner}{\vdash} \ulcorner 0 \neq 0\urcorner,$$

que es una sentencia Π_1 de en $\mathcal{L}_a$. Representaremos igualmente su traducción al lenguaje de T. Es claro que T es consistente si y sólo si $\mathbb{N} \models \mathrm{Consis}\,\ulcorner T\urcorner$.

De este modo, lo esencial del teorema de incompletitud (una vez constatadas las hipótesis de que la teoría T es semirrecursiva e interpreta a Q) puede enunciarse así:

$$\mathbb{N} \models (\mathrm{Consis}\,\ulcorner T\urcorner \leftrightarrow \tilde{G}),$$

donde, como antes, llamamos $\tilde{G} \equiv \neg \underset{\ulcorner T\urcorner}{\vdash} \ulcorner G\urcorner$.

Tendríamos una auténtica paradoja si esto —que es lo que realmente hemos demostrado— no pudiera probarse a partir de los axiomas de Peano. En tal caso sí tendríamos que preguntarnos qué hemos usado sobre los números naturales

que no se deduzca de los axiomas de Peano. Sin embargo, lo cierto es que el teorema de incompletitud, visto así, como una afirmación puramente aritmética expresada por la sentencia anterior, sí puede ser demostrado exclusivamente a partir de los axiomas de Peano, de hecho se puede probar en $\mathrm{I}\Sigma_1$. Esto es precisamente lo que afirma el segundo teorema de incompletitud de Gödel, que, según lo que acabamos de explicar, no es más que la formalización en la teoría T del primer teorema de incompletitud.

Teorema 9.9 (Segundo teorema de incompletitud de Gödel) *Si T es una teoría semirrecursiva que interpreta a $\mathrm{I}\Sigma_1$ y G es una sentencia de Gödel para T, entonces*

$$\underset{T}{\vdash} (\mathrm{Consis}\,\ulcorner T\urcorner \leftrightarrow G).$$

En particular, si T es consistente no $\underset{T}{\vdash} \mathrm{Consis}\,\ulcorner T\urcorner$.

Demostración: En realidad vamos a probar lo que afirmábamos en los comentarios previos al teorema, que es ligeramente más fuerte: si llamamos $\tilde{G} \equiv \neg \underset{\ulcorner T\urcorner}{\vdash} \ulcorner G\urcorner$, demostraremos que

$$\underset{\mathrm{I}\Sigma_1}{\vdash} (\mathrm{Consis}\,\ulcorner T\urcorner \leftrightarrow \tilde{G}),$$

entendiendo ambos términos como sentencias de $\mathcal{L}_a$. Entonces, la traducción de esta equivalencia al lenguaje de T es demostrable en T, pero en T se cumple que $\tilde{G}$ es equivalente a G (por definición de sentencia de Gödel), luego tenemos la equivalencia del enunciado.

Como $\underset{T}{\vdash} (0 \neq 0 \to G)$, la propiedad 1) del teorema 8.28 nos da que

$$\underset{\mathrm{I}\Sigma_1}{\vdash} \underset{\ulcorner T\urcorner}{\vdash} \ulcorner 0 \neq 0 \to G\urcorner,$$

luego la propiedad 3) nos da que

$$\underset{\mathrm{I}\Sigma_1}{\vdash} \underset{\ulcorner T\urcorner}{\vdash} \ulcorner 0 \neq 0\urcorner \to \underset{\ulcorner T\urcorner}{\vdash} \ulcorner G\urcorner,$$

luego, aplicando la regla de Negación del Implicador, concluimos que

$$\underset{\mathrm{I}\Sigma_1}{\vdash} \tilde{G} \to \mathrm{Consis}\,\ulcorner T\urcorner.$$

Recíprocamente, por la propiedad 2) del teorema 8.28,

$$\underset{\mathrm{I}\Sigma_1}{\vdash} (\underset{\ulcorner T\urcorner}{\vdash} \ulcorner G\urcorner \to \underset{\ulcorner T\urcorner}{\vdash} \ulcorner \underset{\ulcorner T\urcorner}{\vdash} \ulcorner G\urcorner\urcorner),$$

que es lo mismo que

$$\underset{\mathrm{I}\Sigma_1}{\vdash} (\neg\tilde{G} \to \underset{\ulcorner T\urcorner}{\vdash} \ulcorner \underset{\ulcorner T\urcorner}{\vdash} \ulcorner G\urcorner\urcorner).$$

Por definición de sentencia de Gödel, $\vdash_T \vdash_{\ulcorner T\urcorner} \ulcorner G\urcorner \to \neg G$, luego por 1)

$$\vdash_{\mathrm{I}\Sigma_1} \vdash_{\ulcorner T\urcorner} \ulcorner \vdash_{\ulcorner T\urcorner} \ulcorner G\urcorner \to \neg G\urcorner$$

y por 3)

$$\vdash_{\mathrm{I}\Sigma_1} (\neg\tilde{G} \to \vdash_{\ulcorner T\urcorner} \ulcorner\neg G\urcorner).$$

Por otro lado, por definición de $\tilde{G}$ tenemos trivialmente que

$$\vdash_{\mathrm{I}\Sigma_1} (\neg\tilde{G} \to \vdash_{\ulcorner T\urcorner} \ulcorner G\urcorner),$$

luego por la formalización de la regla de contradicción

$$\vdash_{\mathrm{I}\Sigma_1} (\neg\tilde{G} \to \vdash_{\ulcorner T\urcorner} \ulcorner 0 \neq 0\urcorner),$$

es decir, $\vdash_{\mathrm{I}\Sigma_1} (\neg\tilde{G} \to \neg\,\mathrm{Consis}\ulcorner T\urcorner)$. ∎

Así pues, acabamos de probar que no hay nada de "misterioso" en el primer teorema de incompletitud de Gödel que no pueda demostrarse en $\mathrm{I}\Sigma_1$.

Observaciones Mientras la sentencia de Gödel tiene una interpretación autorreferente, su forma equivalente $\mathrm{Consis}\ulcorner T\urcorner$ tiene una interpretación mucho más simple: la consistencia de la teoría axiomática considerada. Así se ve más claramente que la paradoja que describíamos antes no es tal: no hemos probado que la sentencia $\mathrm{Consis}\ulcorner T\urcorner$ sea verdadera y no demostrable, sino que, si es verdadera (es decir, si T es consistente) entonces no es demostrable. Nuevamente, por el teorema 7.29 sabemos que $\mathrm{Consis}\ulcorner T\urcorner$ es equivalente a que una determinada ecuación diofántica no tenga solución. En otras palabras, para cada teoría axiomática semirrecursiva existe un polinomio P (de varias variables) con coeficientes enteros tal que la existencia de una solución entera de la ecuación $P = 0$ equivale a la existencia de una demostración de una contradicción en la teoría.

En algunos casos sencillos podemos asegurar que $\mathrm{Consis}\ulcorner T\urcorner$ es verdadera. Por ejemplo, la sentencia $\mathrm{Consis}\ulcorner \mathrm{AP}\urcorner$ es un ejemplo de afirmación verdadera sobre los números naturales que no es demostrable a partir de los axiomas de Peano. Esto es posible porque la prueba de la consistencia de AP (es decir, la observación de que el conjunto de los números naturales constituye un modelo de AP junto con las definiciones y teoremas relativos a modelos) involucra esencialmente razonamientos sobre colecciones infinitas y relaciones y funciones sobre colecciones infinitas que van más allá de lo que podemos formalizar en AP. No obstante, en el capítulo siguiente presentaremos una versión más potente de la aritmética de Peano que sí permite considerar sucesiones infinitas de conjuntos infinitos, y en ella la prueba de la consistencia de AP se formaliza sin dificultad (teorema 10.7).

De este modo, la aritmética de Peano es una teoría axiomática de la que sí está justificado decir que es más limitada que la mente humana: Sabemos más

sobre los números naturales de lo que puede probarse a partir de los axiomas de Peano. Concretamente, la definición (metamatemática) de la relación $\mathbb{N} \vDash \alpha$ requiere considerar sucesiones finitas de conjuntos infinitos, mientras que en AP, aunque es posible hablar de conjuntos infinitos a través de fórmulas (los conjuntos infinitos son en realidad "clases propias" en AP) sólo es posible definir sucesiones finitas de conjuntos finitos, y ésa es la razón última (en la práctica) por lo que la definición de $\mathbb{N} \vDash \alpha$ no es formalizable en AP.

Pasemos ahora al extremo opuesto: sea T una teoría axiomática de conjuntos. Hasta ahora hemos descrito varias teorías débiles, la más fuerte de las cuales es ZF$-$AI, que con el axioma "todo conjunto es finito" es equivalente a AP, pero veremos que si añadimos a ZF el axioma que postula la existencia de conjuntos infinitos obtenemos una teoría muchísimo más potente, hasta el punto de que en ella se puede formalizar fácilmente cualquier razonamiento (meta)matemático que podamos considerar racionalmente "convincente". Esto hace que si fuera posible dar un argumento convincente de que T es consistente, no habría ninguna dificultad en convertirlo en una demostración matemática en T de la sentencia Consis$\ulcorner T \urcorner$ (exactamente igual que cualquier matemático sabe convertir en teoremas de T todos sus razonamientos válidos). El segundo teorema de incompletitud nos daría entonces que T es contradictoria. Más concretamente, nos permitiría construir explícitamente una contradicción en T. Con esto hemos probado algo muy importante:

> *Si la teoría de conjuntos es consistente, no existe ningún argumento que pueda convencernos de que así es.*

Equivalentemente, si T es una teoría de conjuntos, la sentencia Consis$\ulcorner T \urcorner$ es un ejemplo de una afirmación sobre números naturales tal que, si es verdadera, jamás conseguiremos demostrar que lo es. Ahora estamos ante una teoría axiomática "más potente" que la mente humana, en el sentido de que en ella pueden formalizarse todos los razonamientos que nosotros consideramos convincentes (los razonamientos metamatemáticos) y muchos razonamientos más sobre objetos extraños, como puedan ser conjuntos no numerables, de los que no sabríamos hablar consistentemente sin la guía de la teoría axiomática de conjuntos.

Observemos que no es difícil demostrar Consis$\ulcorner T \urcorner$ en una teoría adecuada. Por ejemplo, basta llamar T' a la teoría que resulta de añadirle a T el axioma Consis$\ulcorner T \urcorner$ y, ciertamente, en T' se puede probar la consistencia de la teoría de conjuntos, pero la prueba no nos convence de nada. En general, no hay ningún problema en que la consistencia de una teoría T pueda probarse en otra teoría más fuerte T'. Lo que afirma el segundo teorema de incompletitud es que T' ha de ser *necesariamente* más fuerte que T. Así, puesto que la teoría de conjuntos T es más fuerte que nuestra capacidad de razonamiento metamatemático, sucede que no existen razonamientos metamatemáticos que prueben la consistencia de T. Cualquier demostración de esta consistencia (como el caso trivial que acabamos de considerar) partirá necesariamente de algún principio cuya consistencia es, a su vez, dudosa.

Esto supone una seria limitación a la fundamentación de la matemática. El programa de fundamentación de Hilbert pedía una teoría axiomática de conjun-

tos cuya consistencia y completitud pudieran ser demostradas mediante técnicas metamatemáticas finitistas. Los teoremas de incompletitud muestran que este programa es irrealizable: la completitud es imposible y la consistencia es indemostrable. Esto no quiere decir que sea imposible fundamentar satisfactoriamente las matemáticas. En el capítulo siguiente veremos varias teorías axiomáticas de conjuntos que cumplen este objetivo, es decir, proporcionan una noción precisa de lo que debemos entender por una demostración matemática rigurosa. Cualquiera de estas teorías constituye de hecho una fundamentación de la matemática en el sentido de que es la referencia que de hecho toman los matemáticos para precisar en qué consiste su trabajo.[4] Es cierto que no podemos probar que ninguna de estas teorías es aceptable (consistente), pero los matemáticos vienen trabajando en ellas más de un siglo sin que nadie haya encontrado ninguna contradicción. Si unimos a esto la imposibilidad teórica marcada por el segundo teorema de incompletitud, concluimos que no hay motivos para sospechar de que la teoría axiomática de conjuntos no sea todo lo sólida que parece ser. Por otra parte, la completitud que exigía Hilbert no es realmente necesaria para el trabajo del matemático. En ninguna rama del conocimiento se considera necesario tener una garantía de poder responder a cualquier pregunta. Es cierto que las matemáticas parecían ser la única ciencia donde se hubiera podido tener tal garantía, pero el primer teorema de incompletitud no ha hecho sino acercarla a otras ramas del saber, como la física o la biología. Sin duda es imposible saber exactamente cómo, cuándo y dónde apareció el primer organismo vivo sobre la Tierra, pero esto no quita para que podamos determinar con gran precisión el proceso que dio lugar a la aparición de la vida. ∎

Veamos una aplicación del segundo teorema de incompletitud:

La interpretación natural de una sentencia de Gödel equivale a su no demostrabilidad, y hemos demostrado que, bajo las débiles hipótesis del teorema, las sentencias de Gödel son verdaderas (afirman que no pueden ser demostradas y no pueden ser demostradas). Por simple curiosidad, podemos preguntarnos qué sucede si aplicamos el teorema 9.1 para construir una sentencia que afirma su propia demostrabilidad. Específicamente, dada una teoría semirrecursiva T que interprete a Q, sabemos construir una sentencia H tal que

$$\underset{T}{\vdash}(H \leftrightarrow \underset{\ulcorner T\urcorner}{\vdash} \ulcorner H \urcorner).$$

Las sentencias con esta propiedad se llaman *sentencias de Henkin* y no es evidente en principio si son verdaderas o falsas o —lo que en este caso es lo mismo—, si son demostrables o no. La respuesta es que todas son verdaderas, pero la prueba se basa en el segundo teorema de incompletitud.

[4]No hay que entender aquí que la teoría axiomática de conjuntos explique la naturaleza de las matemáticas. Éste es un problema mucho más amplio. La teoría de conjuntos se limita precisar un patrón de rigor suficiente para que el matemático pueda trabajar sin vacilaciones. No obstante, es posible considerar argumentos informales, por ejemplo de carácter geométrico, que merecen el mismo calificativo de "matemáticas" y que no pueden ser considerados teoremas formales.

Teorema 9.10 (Teorema de Löb) *Sea T una teoría semirrecursiva que interprete a $\mathrm{I}\Sigma_1$ y sea H una sentencia tal que*

$$\underset{T}{\vdash} (\underset{\ulcorner T\urcorner}{\vdash} \ulcorner H\urcorner \to H).$$

Entonces $\underset{T}{\vdash} H$.

Demostración: Sea T^* la extensión de T que resulta de añadirle el axioma $\neg H$. Si no $\underset{T}{\vdash} H$, entonces T^* es consistente. Formalizando (en $\mathrm{I}\Sigma_1$) este sencillo resultado obtenemos que $\underset{T^*}{\vdash} \neg \underset{\ulcorner T\urcorner}{\vdash} \ulcorner H\urcorner \to \mathrm{Consis}\,\ulcorner T^*\urcorner$.

Por el segundo teorema de incompletitud tenemos que no $\underset{T^*}{\vdash} \mathrm{Consis}\,\ulcorner T^*\urcorner$, luego no $\underset{T^*}{\vdash} \neg \underset{\ulcorner T\urcorner}{\vdash} \ulcorner H\urcorner$, de donde no $\underset{T}{\vdash} \neg H \to \neg \underset{\ulcorner T\urcorner}{\vdash} \ulcorner H\urcorner$, es decir, que no se cumple $\underset{T}{\vdash} \underset{\ulcorner T\urcorner}{\vdash} \ulcorner H\urcorner \to H$, como había que probar. ∎

9.4 Incompletitud y aritmética no estándar

Los teoremas de incompletitud nos permiten construir y estudiar más claramente modelos no estándar de la aritmética. En efecto, sea T una teoría recursiva y consistente que interprete a Q. Llamemos $S(x) \equiv \overset{x}{\underset{T}{\vdash}} \ulcorner 0 \neq 0\urcorner$, de modo que $S(x)$ significa "x es (el número de Gödel de) una demostración de $0 \neq 0$ en T". Por definición:

$$\mathrm{Consis}\,\ulcorner T\urcorner \equiv \neg \bigvee x\, S(x).$$

Si, por ejemplo, T es una teoría de conjuntos suficientemente potente como para formalizar todo razonamiento metamatemático aceptable, (y suponemos que es consistente) todo número natural n tiene la propiedad $\mathbb{N} \models \neg S(0^{(n)})$ (es decir, ningún número natural demuestra una contradicción en T), pero, al mismo tiempo, es imposible demostrar que esto es así: no existe ningún argumento concluyente que justifique que todo número natural tiene dicha propiedad, pues en tal caso dicho argumento podría formalizarse en T y tendríamos

$$\underset{T}{\vdash} \mathrm{Consis}\,\ulcorner T\urcorner,$$

pero el segundo teorema de incompletitud implicaría entonces que T es contradictoria, luego nuestro argumento "concluyente" de la consistencia de T no podría ser tan "concluyente". En suma tenemos una propiedad S que es recursiva (podemos comprobar explícitamente en un tiempo finito si un número natural la cumple o no) y es verdad que todos los números naturales cumplen la propiedad $\neg S$, pero este hecho no es verdadero por ninguna razón. Es verdadero, pero no hay ningún argumento que justifique que lo es. Es verdadero simplemente porque "sucede", porque si vamos mirando los números naturales uno por uno, lo cierto es que ninguno deja de cumplir la propiedad $\neg S$, pero sin que haya ningún argumento que obligue a que así sea.

Volviendo al caso general en que T es cualquier teoría recursiva consistente, puesto que $\mathrm{Consis}\, T$ no es demostrable en T, el teorema 3.6 nos da que la teoría T' que resulta de añadirle a T el axioma $\neg\,\mathrm{Consis}\,\ulcorner T\urcorner$ o, equivalentemente,

$$\bigvee x \in \mathbb{N}\, S(x),$$

es consistente. Si suponemos que T interpreta a $\mathrm{I}\Sigma_1$ en T' podemos demostrar

$$\overset{1}{\bigvee} x \in \mathbb{N}(S(x) \wedge \bigwedge y < x\, \neg S(y)),$$

es decir, que existe un mínimo número natural que cumple $\neg S(x)$. Esto nos permite definir

$$c \equiv x \mid (x \in \mathbb{N} \wedge S(x) \wedge \bigwedge y < x\, \neg S(y)).$$

La interpretación natural[5] de c es que se trata de la mínima demostración en T de que $0 \neq 0$. Como estamos suponiendo que T es consistente, c es una descripción impropia en su interpretación natural. Así, como convenimos que las descripciones impropias denotan el cero, tenemos que la interpretación natural de c es el número 0. Sin embargo, en T' tenemos que c es una descripción propia, por lo que la regla de las descripciones propias nos da que $\underset{T'}{\vdash} S(c)$.

Por otra parte, puesto que T es consistente, todo n cumple $\mathbb{N} \vDash \neg S(0^{(n)})$, luego $\underset{T'}{\vdash} \neg S(0^{(n)})$.

Así pues, en T' podemos probar[6] que c es un número natural que cumple una propiedad que también sabemos probar que no cumple ni 0, ni 1, ni 2, etc. Si M es un modelo de T', entonces $M(c)$ es un número natural no estándar. A diferencia de lo que sucedía en el capítulo IV, ahora sabemos definir explícitamente números no estándar en una teoría T: un ejemplo es el mínimo número de Gödel de la demostración de una contradicción en T (o, más concretamente, de $0 \neq 0$). Por supuesto, necesitamos postular su existencia con un axioma, pero no necesitamos introducir una constante no definida para referirnos a él.

Observemos que c no es el mínimo número no estándar (de hecho no existe tal mínimo). En efecto, puesto que podemos probar que $c \neq 0$, de aquí deducimos que existe un número d tal que $c = d'$. Es claro que d también es no estándar, en el sentido de que para todo número natural n sabemos probar que $d \neq 0^{(n)}$.

Notemos que no existe un único modelo M para la aritmética no estándar, sino que existen infinitos modelos "no isomorfos" dos a dos, en el sentido de que satisfacen sentencias diferentes. Por ejemplo, a partir de AP podemos formar dos teorías axiomáticas consistentes, pero mutuamente contradictorias,

$$\mathrm{AP} + \mathrm{Consis}\,\ulcorner \mathrm{AP}\urcorner \qquad \text{y} \qquad \mathrm{AP} + \mathrm{Consis}\,\ulcorner \neg \mathrm{AP}\urcorner.$$

Si las llamamos T_0 y T_1, respectivamente, cualquiera de ellas puede extenderse a su vez a dos teorías consistentes y mutuamente contradictorias:

$$T_0 + \mathrm{Consis}\,\ulcorner T_0\urcorner, \quad T_0 + \mathrm{Consis}\,\ulcorner \neg T_0\urcorner, \quad T_1 + \mathrm{Consis}\,\ulcorner T_1\urcorner, \quad T_1 + \mathrm{Consis}\,\ulcorner \neg T_1\urcorner,$$

[5] Notemos que c es la traducción al lenguaje de T de un término de $\mathcal{L}_a$, por lo que podemos hablar de la interpretación de dicho término en el modelo natural de $\mathcal{L}_a$.

[6] Tenemos así un ejemplo de teoría consistente ω-contradictoria.

que podemos llamar T_{00}, T_{01}, T_{10} y T_{11}, respectivamente, e igualmente podemos formar otras ocho teorías T_{000}, T_{001}, T_{010}, ... Cada una de las teorías construidas de este modo tiene su propio modelo, y dos cualesquiera de estos modelos satisfacen sentencias mutuamente contradictorias.

Vamos a probar ahora que, en realidad, no necesitamos elegir una nueva sentencia a cada paso para formar nuevas teorías consistentes, sino que podemos usar siempre la misma. Necesitamos un resultado previo:

Teorema 9.11 *Si T es una teoría consistente que interpreta a* Q*, existe una fórmula $\phi(x)$ de tipo Σ_1 tal que, para todo número natural k, la teoría*

$$T + \bigwedge x \in \mathbb{N}(\phi(x) \leftrightarrow x = 0^{(k)})$$

es consistente.

Demostración: Sea $\mathcal{L}$ el lenguaje de T. Consideremos la fórmula Σ_1

$$\chi(p,k) \equiv p \in \mathrm{Form}(\ulcorner\mathcal{L}\urcorner) \wedge \bigvee x d(\mathrm{Vlib}(p) = \{x\} \wedge \underset{\ulcorner T\urcorner}{\overset{d}{\vdash}} \neg\bigwedge x \in \mathbb{N}(p \leftrightarrow x = 0^{(k)})$$

$$\wedge \bigwedge d' < d \bigwedge k' < d' \neg \underset{\ulcorner T\urcorner}{\overset{d'}{\vdash}} \neg\bigwedge x \in \mathbb{N}(p \leftrightarrow x = 0^{(k')})).$$

Así, si $\mathbb{N} \vDash \chi(\ulcorner\psi(x)\urcorner, 0^{(k)})$ y $\mathbb{N} \vDash \chi(\ulcorner\psi(x)\urcorner, 0^{(k')})$, con $k \neq k'$, existen d y d' que demuestran, respectivamente,

$$\underset{T}{\vdash} \neg\bigwedge x \in \mathbb{N}(\psi(x) \leftrightarrow x = 0^{(k)}), \qquad \underset{T}{\vdash} \neg\bigwedge x \in \mathbb{N}(\psi(x) \leftrightarrow x = 0^{(k')}).$$

No perdemos generalidad si suponemos $d' < d$, y claramente $k' < d'$ (pues d' demuestra una fórmula que contiene el numeral $0^{(k')}$), pero esto contradice la minimalidad de d que postula χ.

Vemos, pues, que χ define una función parcial. El teorema 5.31 nos da una fórmula $\alpha(u,v)$ de tipo Σ_1 tal que, si se cumple $\mathbb{N} \vDash \chi(\ulcorner\psi(x)\urcorner, 0^{(k)})$, entonces

$$\underset{T}{\vdash} \bigwedge v(\alpha(\ulcorner\psi(x)\urcorner, v) \leftrightarrow v = 0^{(k)}).$$

Por el teorema 9.1 (véase la nota posterior), existe una fórmula $\phi(x)$ tal que

$$\underset{T}{\vdash} \bigwedge x(\phi(x) \leftrightarrow \alpha(\ulcorner\phi(x)\urcorner, x)).$$

Esta equivalencia prueba que ϕ es Σ_1.

Vamos a probar que no existe ningún k tal que $\mathbb{N} \vDash \chi(\ulcorner\phi(x)\urcorner, 0^{(k)})$. Si existiera, consideremos el d cuya existencia postula χ, de modo que es el mínimo número natural que demuestra $\underset{T}{\vdash} \neg\bigwedge x \in \mathbb{N}(\phi(x) \leftrightarrow x = 0^{(k)})$.

Por otra parte tenemos que

$$\underset{T}{\vdash} \bigwedge v(\alpha(\ulcorner\phi(x)\urcorner, v) \leftrightarrow v = 0^{(k)}),$$

de donde

$$\underset{T}{\vdash} \bigwedge x(\phi(x) \leftrightarrow x = 0^{(k)}),$$

y resulta que T es contradictoria. Concluimos que, para todo k,

$$\text{no } \underset{T}{\vdash} \neg \bigwedge x \in \mathbb{N}(\phi(x) \leftrightarrow x = 0^{(k)}),$$

pues en caso contrario podríamos tomar una mínima demostración d que probaría $\mathbb{N} \vDash \chi(\ulcorner \phi(x) \urcorner, 0^{(k)})$. Esto implica a su vez que la teoría del enunciado es consistente. ∎

En definitiva, fijada una teoría semirrecursiva consistente que interprete a Q, hemos construido una fórmula $\phi(x)$ con la propiedad de que es consistente con T que 0 sea el único número que la cumple, pero también que 1 sea el único número que la cumple, o que lo sea 2, etc. Podemos refinar aún más la conclusión:

Teorema 9.12 *Si T es una teoría semirrecursiva consistente que interpreta a* Q, *existe una fórmula*[7] *$\psi(x)$ de tipo Σ_1 tal que, si representamos $+\psi \equiv \psi$ y $-\psi \equiv \neg\psi$, entonces la teoría que resulta de añadir a T los axiomas*

$$\pm\psi(0), \quad \pm\psi(1), \quad \pm\psi(2), \quad \pm\psi(3), \quad \ldots$$

es consistente, para cualquier elección de los signos.

Demostración: Puesto que una hipotética prueba de una contradicción usaría sólo un número finito de axiomas, basta probar que la teoría que resulta de añadir los axiomas $\pm\psi(0^{(0)}), \ldots, \pm\psi(0^{(n)})$ es consistente, para cada número natural n. Sea $\phi(x)$ la fórmula dada por el teorema anterior y sea

$$\psi(x) \equiv \bigvee s \in \mathbb{N}(\mathrm{Suc}(s) \wedge \phi(s) \wedge x < \ell(s) \wedge s_x = 1).$$

Claramente es (la traducción de) una fórmula Σ_1 de $\mathcal{L}_a$. Sea ahora una sucesión arbitraria s de n ceros y unos y consideremos la teoría T' dada por

$$T + \bigwedge x(\phi(x) \leftrightarrow x = \ulcorner s \urcorner).$$

Por el teorema anterior sabemos que T' es consistente, y es claro entonces que la s que aparece en la definición de ψ es necesariamente $\ulcorner s \urcorner$, es decir, que

$$\underset{T'}{\vdash} (\psi(x) \leftrightarrow x < 0^{(n)} \wedge \ulcorner s \urcorner_x = 1)$$

Así, para cada $i < n$, puesto que $\underset{T'}{\vdash} \ulcorner s \urcorner_{0^{(i)}} = 0^{(s(i))}$, es claro que $\underset{T'}{\vdash} \psi(0^{(i)})$ o $\underset{T'}{\vdash} \neg\psi(0^{(i)})$ según si $s(i)$ es cero o uno. Si la correspondiente extensión de T con estas sentencias fuera contradictoria, también lo sería T'. ∎

Así pues, aplicando el teorema a AP, concluimos que existe una fórmula $\phi(x)$ de tipo Σ_1 que la cumplirán los números naturales que la cumplan, pero de tal

[7] Las fórmulas con esta propiedad se llaman fórmulas *flexibles*.

modo que los axiomas de Peano no permiten probar nada al respecto, sino que es consistente con ellos suponer que los números que queramos la cumplen y que cualesquiera otros no la cumplen. Y esto no es una deficiencia particular de los axiomas de Peano, sino que lo mismo sucede (cambiando la formula ϕ por otra adecuada) para toda teoría axiomática que cumpla los requisitos mínimos usuales.

Cada extensión de AP con una determinación particular de una fórmula flexible nos da un modelo diferente de AP, en el sentido de que dos cualesquiera de ellos difieren al menos en que uno satisface una sentencia que el otro no satisface.

Tercera parte

Teorías de conjuntos

Capítulo X

Clases y conjuntos

Ya hemos tenido ocasión de enfrentarnos al hecho de que, en las teorías axiomáticas de conjuntos, surge la necesidad de distinguir entre "clases" y "conjuntos", debido a que no todas las colecciones de conjuntos que podemos especificar mediante una fórmula del lenguaje formal de la teoría (por ejemplo, la de todos los conjuntos, la de todos los ordinales, etc.) pueden considerarse como la extensión de un conjunto sin caer en contradicciones. En realidad se trata de un fenómeno habitual, sobre todo en las teorías axiomáticas que no son teorías de conjuntos propiamente dichas, en las que no "se espera" que las colecciones de objetos puedan representarse como conjuntos. Es el caso de AP, en la que, a través de las fórmulas oportunas, podemos hablar de "los números pares", "los números primos", etc. sin que exista en la teoría nada que podamos considerar "el conjunto de los números pares", etc., pues en la teoría sólo hay números naturales, y ningún número natural es "el conjunto de todos los números pares".

Sin embargo, a los matemáticos en general no les gustan los conceptos metamatemáticos, y no les gusta tener que hablar de fórmulas como sucedáneos de conjuntos para referirse a las colecciones de conjuntos con las que trabajan. La notación conjuntista que hemos introducido para las clases (que no es sino una forma de tratar con fórmulas arbitrarias de modo que parezca que se está tratando con conjuntos, aunque el "parezca" se traduce en que, a la hora de la verdad, determinados razonamientos válidos para conjuntos no valen para clases propias) las hace tolerables en la medida en que no se vean obligados a tratar mucho con ellas, pero la idea de que esas "clases propias" no sean más que una forma velada de tratar con fórmulas metamatemáticas no les acaba de gustar.

En este capítulo veremos cómo formalizar la noción de clase propia, tanto en el caso de la teoría de conjuntos ZF^* como en AP, sólo que en el caso de AP a las clases de números naturales (es decir, las colecciones infinitas de números naturales definibles mediante una fórmula de $\mathcal{L}_a$) se las suele llamar conjuntos, pero son el equivalente de las clases propias de ZF^*.

Cabe aclarar que esto no supone eliminar la "molesta" necesidad de tratar con colecciones de objetos definibles mediante fórmulas pero que no son la extensión de un conjunto o de una clase formalizada, sino que simplemente eleva

un nivel su presencia en la teoría y las aleja del quehacer cotidiano de los matemáticos. Aprovechamos para explicar por fin qué significa "de primer orden" cuando hablamos de la lógica de primer orden.

10.1 La aritmética de segundo orden

Un lenguaje formal de primer orden tiene (o puede tener) constantes, que pretenden representar objetos fijos (por lo menos, cuando se fija un modelo) y variables, que pretenden recorrer objetos arbitrarios (incluso cuando se fija un modelo). Sin embargo, sus relatores y funtores son en este sentido como las constantes, pues cada relator pretende representar una relación fija (al menos cuando se fija un modelo) y cada funtor pretende representar una función fija. Con los cuantificadores podemos decir "para todo objeto x" o "existe un objeto x", pero no podemos decir "para toda relación R" o "para toda función f".

Un lenguaje de segundo orden es un lenguaje formal que, además de *variables de primer orden* (las variables que recorren objetos arbitrarios, como en la lógica de primer orden) tiene también *variables de segundo orden,* que pretenden recorrer relaciones o funciones arbitrarias (incluso fijado un modelo).

Del mismo modo que cada relator y cada funtor debe tener definido un rango que determina el número de sus argumentos, al definir un lenguaje de segundo orden hay que especificar para cada variable de segundo orden si es una variable de relación o de función y asignarle un rango. Por ello se pueden definir lenguajes distintos según cuántos tipos de variables de segundo orden se quiera incluir en ellos.

No es nuestro propósito desarrollar aquí una teoría general sobre lenguajes de segundo orden (para ello véase, por ejemplo, el capítulo VII de [LF]), sino únicamente describir uno en concreto para entender mejor en qué consiste que la lógica de primer orden sea "de primer orden". Nuestro lenguaje tendrá únicamente variables de relación de segundo orden y de rango 1, es decir, que las variables de segundo orden recorrerán relaciones monádicas "arbitrarias". Luego discutiremos cómo debe entenderse esto, pero de momento nos amparamos una vez más en que conocido principio de que para definir un lenguaje formal no necesitamos decir nada sobre la posible interpretación de sus signos:

Definición El *lenguaje de la aritmética de segundo orden* está determinado por los signos siguientes:

1. Variables de primer orden $x_0, x_1, x_2, \ldots$
2. Variables de segundo orden $X_0, X_1, X_2, \ldots$
3. La constante 0,
4. Los funtores S (monádico) y $+, \cdot$ (diádicos),
5. El igualador $=$ (relator diádico)
6. Los mismos signos lógicos (conectores, cuantificador, descriptor) que los lenguajes de primer orden.

Convendremos en usar letras minúsculas para referirnos a variables de primer orden y letras mayúsculas para referirnos a variables de segundo orden.

Ahora podemos definir los *términos de primer orden,* los *términos de segundo orden* y las *fórmulas* mediante las reglas siguientes:

1. Toda variable de primer (resp. segundo) orden es un término de primer (resp. segundo) orden.
2. Si t_1 y t_2 son términos de primer orden, entonces St_1, $t_1 + t_2$, $t_1 \cdot t_2$ son términos de primer orden.
3. Si t_1 y t_2 son términos del mismo orden, entonces $t_1 = t_2$ es una fórmula.
4. Si T es un término de segundo orden y t es un término de primer orden, entonces $T(t)$ es una fórmula.
5. Si α y β son fórmulas, también lo son $\neg\alpha$, $\alpha \to \beta$, $\bigwedge x\alpha$ y $\bigwedge X\alpha$.
6. Si α es una fórmula, $x|\alpha$ es un término de primer orden y $X|\alpha$ es un término de segundo orden.

El concepto de variable libre se define de forma obvia. La *aritmética de Peano de segundo orden* es la teoría determinada por los axiomas siguientes:

- Los axiomas de la aritmética de Robinson Q,
- El principio de inducción:

$$\bigwedge X(X(0) \wedge \bigwedge x(X(x) \to X(S(x))) \to \bigwedge x\ X(x)).$$

- El axioma de extensionalidad:

$$\bigwedge XY(\bigwedge x(X(x) \leftrightarrow Y(x)) \to X = Y)$$

- El esquema de comprensión:

$$\bigvee X\bigwedge x(X(x) \leftrightarrow \phi(x)),$$

  para toda fórmula $\phi(x)$, tal vez con más variables libres.

En realidad, esta definición presupone un concepto que no hemos definido, y es el de "teoría". Para que estos "axiomas" puedan considerarse realmente axiomas en algún sentido, es necesario especificar un calculo deductivo de segundo orden o, alternativamente, una semántica de segundo orden (que nos permitiera hablar de modelos de la teoría). Lo que dicha hipotética "semántica de segundo orden" debería establecer es que, por ejemplo, cuando escribimos $\bigwedge X$ en el principio de inducción, eso deba interpretarse como "para toda propiedad X", es decir, que la condición que sigue se cumpla para toda posible propiedad de los objetos del hipotético modelo.

Observemos que esto es mucho más fuerte que lo que expresa el principio de inducción de AP, pues éste postula que el principio de inducción se cumple para toda posible propiedad de los números naturales definida por una fórmula concreta, lo cual tiene un significado totalmente preciso. En cambio, el principio de inducción de segundo orden debería ser verdadero en un modelo si el principio de inducción es aplicable a cualquier posible propiedad concebible sobre los objetos del modelo, sin exigir que sea definible por una fórmula del modelo, y en principio habría que admitir propiedades definibles en términos del propio modelo.

Podría decirse que las "propiedades arbitrarias" sobre las que se supone que debe variar la variable de segundo orden X están precisadas por el esquema de comprensión, que viene a decir que, en definitiva, son también propiedades definidas por fórmulas, pero a eso hay que hacer dos observaciones: en primer lugar el axioma de comprensión sólo postula que las fórmulas definen propiedades, pero no que toda propiedad sea definible mediante una fórmula, y en segundo lugar, aunque de algún modo pudiéramos restringirnos a propiedades definidas por fórmulas de segundo orden, esto no es menos problemático, pues en la fórmula ϕ pueden aparecer cuantificadores $\bigwedge Y$ o $\bigvee Y$ de interpretación dudosa.

En suma, definir una semántica de segundo orden acorde con la "intención" con la que hemos definido el lenguaje de la aritmética de segundo orden exigiría aceptar que podemos dar un significado objetivo a afirmaciones que involucran posibles propiedades arbitrarias, no especificadas, de los números naturales (o, lo que es lo mismo, subconjuntos arbitrarios del conjunto de los números naturales, y ya hemos argumentado que "la totalidad de los subconjuntos de $\mathbb{N}$" es un concepto matemático de contenido dudoso, un concepto que es cuestionable que pueda ser manejado sin el soporte de una teoría axiomática de conjuntos). Alternativamente, en la definición de modelo de la aritmética de segundo orden podríamos exigir que haya que especificar una clase bien definida de propiedades (o, equivalentemente, de subconjuntos de $\mathbb{N}$) en la que permitimos que varíen las variables de segundo orden.

Pero si estamos dispuestos a aceptar este enfoque, resulta que, como vamos a ver seguidamente, podemos conseguir lo mismo sin necesidad de salirnos del entorno que ya conocemos de la lógica de primer orden. Esto hace que, aunque suene paradójico, lo que habitualmente se conoce como "aritmética de Peano de segundo orden" es una teoría axiomática de primer orden:

Definición 10.1 Llamaremos *lenguaje formal de la aritmética de segundo orden* al lenguaje formal de primer orden $\mathcal{L}_a^2$ cuyos signos primitivos son una constante 0, un funtor monádico S, dos funtores diádicos $+$ y $\cdot$, un relator monádico cto, y un relator diádico $\in$ (aparte del igualador). Abreviaremos $x' \equiv Sx$ y $\mathrm{Nat}\, x \equiv \neg\, \mathrm{cto}\, x$.

De este modo, el lenguaje $\mathcal{L}_a^2$ nos permite hablar de dos clases de objetos: números naturales y conjuntos. Vamos a ver cómo con estos ingredientes de primer orden podemos imitar el lenguaje de segundo orden que hemos definido antes. Para ello dividimos las variables de $\mathcal{L}_a^2$ en dos subconjuntos de infinitas

variables cada uno. A las de uno de ellos las llamaremos *variables de primer orden,* y las representaremos por letras minúsculas, y a las del otro *variables de segundo orden,* y las representaremos por letras mayúsculas.

Diremos que una expresión de $\mathcal{L}_a^2$ es un *término de primer orden,* un *término de segundo orden* o una *fórmula estructurada* si se puede probar que lo es a partir de las propiedades siguientes:[1]

1. Las variables de primer (resp. segundo) orden son términos de primer (resp. segundo) orden.
2. 0 es un término de primer orden.
3. Si t_1 y t_2 son términos de primer orden, también lo son St_1, t_1+t_2 y $t_1\cdot t_2$.
4. Si t_1 y t_2 son términos de primer orden $t_1 = t_2$ es una fórmula estructurada.
5. Si T_1 y T_2 son términos de segundo orden, $T_1 = T_2$ es una fórmula estructurada.
6. Si t_1 es un término de primer orden y T_2 es un término de segundo orden, entonces $t_1 \in T_2$ es una fórmula estructurada.
7. Si α y β son fórmulas estructuradas, también lo son las fórmulas $\neg\alpha$, $\alpha \to \beta$, $\bigwedge x(\mathrm{Nat}\, x \to \alpha)$, y $\bigwedge X(\mathrm{cto}\, X \to \alpha)$,
8. Si α es una fórmula estructurada, $x|(\mathrm{Nat}\, x \wedge \alpha)$ es un término de primer orden, mientras que
$$X|((\overset{1}{\bigvee} X(\mathrm{cto}\, X \wedge \alpha) \wedge \alpha) \vee (\neg \overset{1}{\bigvee} X(\mathrm{cto}\, X \wedge \alpha) \wedge \bigwedge u(\mathrm{Nat}\, u \to u \notin X)))$$
es un término de segundo orden.

Más en general, consideraremos estructuradas todas las fórmulas lógicamente equivalentes a fórmulas estructuradas. Es claro entonces que, en este sentido amplio, si α y β son fórmulas estructuradas también lo son $\alpha \vee \beta$, $\alpha \wedge \beta$, $\alpha \leftrightarrow \beta$, $\bigvee x\,(\mathrm{Nat}\, x \wedge \alpha)$, $\bigvee X\,(\mathrm{cto}\, X \wedge \alpha)$.

La idea es que los términos de orden 1 son los que pretenden denotar números naturales, los de orden 2 los que pretenden denotar conjuntos, y las fórmulas estructuradas son las que sólo admiten que el igualador relacione términos del mismo orden, el relator de pertenencia sólo se aplique para expresar que un número natural pertenece a un conjunto y las variables ligadas están restringidas a números o conjuntos según su tipo.[2]

[1] Más precisamente, si existe una sucesión de expresiones $\theta_0, \ldots, \theta_n$ que termina en la expresión considerada, y una sucesión $a_1, \ldots a_n$ de números 0, 1, 2 (que representan respectivamente las fórmulas, los términos de primer orden y los términos de segundo orden) de modo que α_i es una variable de primer orden y $a_i = 1$, o bien... etc.

[2] Naturalmente existen expresiones de $\mathcal{L}_a^2$ que no tienen esta estructura, como $X + 0 = X$, y que no se corresponden con expresiones de la "auténtica" lógica de segundo orden, pero simplemente no vamos a usar para nada tales expresiones.

Como sólo vamos a considerar expresiones estructuradas, convendremos en que

$$\bigwedge x\alpha \equiv \bigwedge x(\mathrm{Nat}\, x \to \alpha), \quad \bigwedge X\alpha \equiv \bigwedge x(\mathrm{cto}\, X \to \alpha),$$
$$\bigvee x\alpha \equiv \bigvee x(\mathrm{Nat}\, x \wedge \alpha), \quad \bigvee X\alpha \equiv \bigvee X(\mathrm{cto}\, X \wedge \alpha),$$
$$x|\alpha \equiv x|(\mathrm{Nat}\, x \wedge \alpha),$$
$$X|\alpha \equiv X|((\overset{1}{\bigvee} X(\mathrm{cto}\, X \wedge \alpha) \wedge \alpha) \vee (\neg \overset{1}{\bigvee} X(\mathrm{cto}\, X \wedge \alpha) \wedge \bigwedge u(\mathrm{Nat}\, u \to u \notin X))).$$

El último convenio se debe a lo siguiente: según es habitual, vamos a tomar como axioma (semi)lógico $0 = x|x = x$, y de los axiomas que vamos a considerar sobre $\mathcal{L}_a^2$ se deducirá la existencia de un único conjunto vacío. Teniendo esto en cuenta, una descripción de la forma $X|\alpha$ va a ser siempre propia, pues o bien existe un único conjunto X que cumple α, y entonces dicho X es el único que cumple la descripción, o bien no es así, en cuyo caso $\varnothing$ es el único X que cumple la descripción. Esto hace que, por la regla de las descripciones propias (no impropias), cuando $\neg \overset{1}{\bigvee} X\alpha$, se cumple $X|\alpha = \varnothing$, mientras que, por la regla de las descripciones impropias, si $\neg \overset{1}{\bigvee} x\alpha$, entonces $x|\alpha = 0$.

La *Aritmética de Peano de segundo orden* (AP_2) es la teoría axiomática (de primer orden) sobre $\mathcal{L}_a^2$ determinada por los axiomas siguientes:

$$\begin{aligned}
&\text{(AP1)} && \mathrm{Nat}\, 0\\
&\text{(AP2)} && \textstyle\bigwedge x\ \mathrm{Nat}\, x'\\
&\text{(AP3)} && \textstyle\bigwedge x\ x' \neq 0\\
&\text{(AP4)} && \textstyle\bigwedge xy(x' = y' \to x = y)\\
&\text{(AP5)} && \textstyle\bigwedge x\ x + 0 = x\\
&\text{(AP6)} && \textstyle\bigwedge xy\ x + y' = (x + y)'\\
&\text{(AP7)} && \textstyle\bigwedge x\ x \cdot 0 = 0\\
&\text{(AP8)} && \textstyle\bigwedge xy\ x \cdot y' = xy + x\\
&\text{(Inducción)} && \textstyle\bigwedge X(0 \in X \wedge \bigwedge n \in X\ n' \in X \to \bigwedge n\, n \in X)\\
&\text{(Extensionalidad)} && \textstyle\bigwedge XY(\bigwedge u(u \in X \leftrightarrow u \in Y) \to X = Y),\\
&\text{(Comprensión)} && \textstyle\bigwedge X_1 \cdots X_r \bigwedge x_1 \cdots x_s \bigvee X \bigwedge n(n \in X \leftrightarrow \phi(n)),
\end{aligned}$$

para toda fórmula estructurada[3] ϕ de $\mathcal{L}_a^2$ con variables libres $n, X_1, \ldots, X_r$, $x_1, \ldots, x_s$, todas ellas distintas de X.

Llamaremos *expresiones aritméticas* de $\mathcal{L}_a^2$ a las que cumplen la definición de fórmula estructurada salvo la regla 7) en el caso de $\bigwedge X$ y la regla 8) en el caso de $X|$, es decir, las fórmulas estructuradas en las que no aparecen variables ligadas de segundo orden.

[3]En principio, tomamos como axiomas de comprensión los determinados por fórmulas estructuradas en sentido estricto, pero es claro que entonces pueden probarse los casos correspondientes a fórmulas estructuradas en el sentido amplio de ser lógicamente equivalentes a éstas.

La subteoría de AP_2 que resulta de restringir el axioma de comprensión a fórmulas aritméticas se conoce como ACA_0 (por Axioma de Comprensión Aritmética).

Del esquema de comprensión y el axioma de inducción se deduce inmediatamente el esquema de inducción

$$\bigwedge X_1 \cdots X_r \bigwedge x_1 \cdots x_s \bigvee X \bigwedge n(\phi(0) \wedge \bigwedge n(\phi(n) \to \phi(n')) \to \bigwedge n\, \phi(n))$$

para toda fórmula estructurada (en el caso de AP_2) y toda formula aritmética (en el caso de ACA_0).

Observemos que podemos definir una interpretación obvia de AP en ACA_0 (y en particular en AP_2) identificando las variables de $\mathcal{L}_a$ con las variables de primer orden de $\mathcal{L}_a^2$ y usando la fórmula $\mathrm{Nat}\, x$ como universo de la interpretación. Las traducciones de las fórmulas de $\mathcal{L}_a$ son precisamente las fórmulas aritméticas que no contienen variables de segundo orden. Esto implica que la traducción de todo teorema de AP es un teorema de ACA_0 y, por lo tanto, de AP_2. Con el convenio que estamos empleando de sobrentender el relator $\mathrm{Nat}\, x$ en las variables expresadas por letras minúsculas, una expresión de $\mathcal{L}_a$ se representa exactamente igual que su traducción a $\mathcal{L}_a^2$ (si escribimos todas sus variables en minúsculas).

Los axiomas de comprensión y extensionalidad implican que la descripción

$$\{x \mid \phi(x)\} \equiv X | \bigwedge x(x \in X \leftrightarrow \phi(x))$$

es propia salvo en casos triviales (suponiendo que ϕ es aritmética si trabajamos en ACA_0). Más concretamente, si las variables libres de ϕ son $X_1, \ldots X_r$, $x, x_1, \ldots, x_s$, tenemos que

$$\bigwedge X_1 \cdots X_r \bigwedge x x_1 \cdots x_x (x \in \{x | \phi(x)\} \leftrightarrow \phi(x)\}).$$

En particular podemos definir

$$\mathbb{N} \equiv \{x | x = x\}, \quad \varnothing \equiv \{x | x \neq x\},$$

$$A \cap B \equiv \{x \mid x \in A \wedge x \in B\}, \quad A \cup B \equiv \{x \mid x \in A \vee x \in B\},$$

$$A \setminus B \equiv \{x \mid x \in A \wedge x \notin B\}, \quad A \subset B \equiv \bigwedge x \in A\, x \in B, \quad \text{etc.}$$

Eliminación de descriptores Observemos que podemos restringir el esquema de comprensión (tanto de AP_2 como de ACA_0) a los casos correspondientes a fórmulas sin descriptores. En efecto, con esta restricción podemos probar igualmente la existencia de un (único) conjunto vacío, y es claro entonces que

$$y = x | \alpha \leftrightarrow (\bigwedge x(\alpha \leftrightarrow x = y)) \vee (\neg \overset{1}{\bigvee} x \alpha \wedge y = 0),$$

$$Y = X | \alpha \leftrightarrow (\bigwedge X(\alpha \leftrightarrow X = Y) \vee (\neg \overset{1}{\bigvee} X \alpha \wedge \bigwedge u\, u \notin Y).$$

Con estas equivalencias la prueba del teorema 3.32 se adapta trivialmente para probar que toda fórmula estructurada (resp. aritmética) de $\mathcal{L}_a$ es equivalente en ACA_0 a una fórmula estructurada (resp. aritmética) sin descriptores, por lo que todos los casos particulares del esquema de comprensión correspondientes a fórmulas con descriptores pueden probarse a partir de los correspondientes a fórmulas sin descriptores. ∎

La razón por la que hemos introducido la subteoría ACA_0 es que verifica el teorema siguiente:

Teorema 10.2 *Una sentencia de $\mathcal{L}_a$ es demostrable en* AP *si y sólo si su traducción a $\mathcal{L}_a^2$ es demostrable en* ACA_0.

Demostración: Una implicación ya la tenemos demostrada. Fijemos un modelo[4] (que podemos suponer numerable) $M \vDash \mathrm{AP}$. Sea $\phi_0, \phi_1, \ldots$ una enumeración de las fórmulas de $\mathcal{L}_a$ que tienen a x_0 como variable libre. Consideramos todos los pares $(\phi_i, \mathbf{a})$, donde $\mathbf{a}$ es una sucesión de elementos de M cuya longitud es igual al número de variables libres de ϕ_i distintas de x_0. Diremos que

$$(\phi_i, \mathbf{a}) \sim (\phi_j, \mathbf{b}) \quad \text{syss para todo } a \text{ de } M,\ M \vDash \phi_i[a, \mathbf{a}] \quad \text{syss} \quad M \vDash \phi_j[a, \mathbf{b}].$$

Representaremos por $[\phi_i, \mathbf{a}]$ la clase de equivalencia respecto de la relación anterior, es decir, el conjunto de todos los pares $(\phi_j, \mathbf{b})$ relacionados con $(\phi_i, \mathbf{a})$. Llamaremos $\bar{M}$ al conjunto[5] formado por todos los elementos de M más todas las clases $[\phi_i, \mathbf{a}]$. Vamos a convertir a $\bar{M}$ en el universo de un modelo de ACA_0.

Definimos la relación $\bar{M}(\mathrm{cto})$ como la que se verifica sobre los elementos de la forma $[\phi_i, \mathbf{a}]$, de modo que

$$M \vDash \mathrm{cto}[C] \quad \text{syss} \quad C \equiv [\phi_i, \mathbf{a}], \text{ para ciertos } \phi_i, \mathbf{a},$$

$$M \vDash \mathrm{Nat}[n] \quad \text{syss} \quad n \text{ está en } M.$$

Definimos $\bar{M}(0) \equiv M(0)$ y definimos las funciones $\bar{M}(S)$, $\bar{M}(+)$ y $\bar{M}(\cdot)$ como extensiones arbitrarias de las funciones correspondientes de M, es decir, como funciones que sobre M actúan como las interpretaciones de los funtores correspondientes de $\mathcal{L}_a$ y sobre los demás elementos de $\bar{M}$ toman, por ejemplo, el valor $M(0)$, aunque esto es irrelevante. Como descripción impropia fijamos $M(0)$. Finalmente, la relación $\bar{M}(\in)$ se cumple sólo en el caso:

$$\bar{M}(\in)(a, [\phi_i, \mathbf{a}]) \quad \text{syss} \quad M \vDash \phi_i[a, \mathbf{a}],$$

que no depende de la elección del representante $(\phi_i, \mathbf{a})$ de la clase. Ahora una simple inducción sobre la longitud de una expresión prueba que, para todo término $t(x_1, \ldots, x_n)$ de $\mathcal{L}_a$ y todos los $a_1, \ldots, a_n$ en M, se cumple

$$M(t)[a_1, \ldots, a_n] \equiv \bar{M}(\bar{t})[a_1, \ldots, a_n],$$

[4] En [CS 9.21] damos una demostración constructiva de este mismo resultado, sin usar modelos.

[5] No perdemos generalidad si suponemos que ningún conjunto $[\phi_i, \mathbf{a}]$ es uno de los objetos del universo de M, por ejemplo, porque podemos suponer que los objetos de M son simplemente números naturales.

y para toda fórmula $\alpha(x_1,\dots,x_n)$ se cumple

$$M \vDash \alpha[a_1,\dots,a_n] \quad \text{syss} \quad \bar{M} \vDash \bar{\alpha}[a_1,\dots,a_n],$$

donde $\bar{t}$ y $\bar{\alpha}$ son las traducciones a $\mathcal{L}_a^2$ de t y α respectivamente. En particular, si α es una sentencia tenemos que

$$M \vDash \alpha \quad \text{syss} \quad \bar{M} \vDash \bar{\alpha}.$$

Ahora es inmediato que $\bar{M}$ satisface los axiomas AP1–AP8. Para probar que

$$\bar{M} \vDash \bigwedge X(0 \in X \wedge \bigwedge n \in X\ n' \in X \to \bigwedge n\, n \in X)$$

tomamos un objeto C de $\bar{M}$ que cumpla $M \vDash \mathrm{cto}[C]$, de modo que $C \equiv [\phi_i, \mathbf{a}]$, y suponemos

$$\bar{M} \vDash 0 \in [C] \wedge \bigwedge n \in [C]\, n' \in [C].$$

Esto significa que $M \vDash (\phi_i(0) \wedge \bigwedge n(\phi_i(n) \to \phi_i(n')))[\mathbf{a}]$ y, como M cumple el principio de inducción, esto implica que $M \vDash \bigwedge n\, \phi_i(n)[\mathbf{a}]$, pero esto equivale a que todo n de $\bar{M}$ que cumpla $\bar{M} \vDash \mathrm{Nat}[n]$ cumple $\bar{M} \vDash [n] \in [C]$, es decir, a que $\bar{M} \vDash \bigwedge n\, n \in [C]$ y con esto queda probado el principio de inducción.

El axioma de extensionalidad se demuestra sin dificultad: fijamos C, D en $\bar{M}$ que cumplan $\bar{M} \vDash \mathrm{cto}[C]$ y $\bar{M} \vDash \mathrm{cto}[D]$, de modo que $C \equiv [\phi_i, \mathbf{a}]$ y $D \equiv [\phi_j, \mathbf{b}]$. Es claro que podemos sustituir ϕ_j por la fórmula de cambiar sus variables libres por otras nuevas de modo que no haya ninguna en común con ϕ_i salvo x_0, y la clase $[\phi_j, \mathbf{b}]$ sigue siendo la misma.

Suponemos $\bar{M} \vDash \bigwedge u(u \in [C] \leftrightarrow u \in [D])[\mathbf{a}, \mathbf{b}]$, lo cual equivale a que

$$M \vDash \bigwedge u(\phi_i(u) \leftrightarrow \phi_j(u))[\mathbf{a}, \mathbf{b}]$$

o también a que para todo a de M se cumpla $M \vDash \phi_i[a, \mathbf{a}]$ syss $M \vDash \phi_j[a, \mathbf{b}]$, es decir, a que $(\phi_i, \mathbf{a}) \sim (\phi_j, \mathbf{b})$, o también a que $C \equiv D$, luego $\bar{M} \vDash [C] = [D]$, como había que probar.

Sólo falta probar que $\bar{M}$ satisface el esquema de comprensión aritmética. Lo probamos por inducción sobre la fórmula aritmética ϕ (que podemos suponer sin descriptores). Notemos que los términos de primer orden sin descriptores no pueden contener variables de segundo orden, por lo que son traducciones de términos de $\mathcal{L}_a$, mientras que los términos de segundo orden son simplemente las variables de segundo orden.

Si $\phi \equiv t_1(x_0, x_1, \dots, x_n) = t_2(x_0, x_1, \dots, x_n)$, fijamos una sucesión $\mathbf{a}$ en M de interpretaciones de las variables $x_1, \dots, x_n$ y tomamos $C \equiv [t_1 = t_2, \mathbf{a}]$. Entonces, para todo a en M, se cumple

$$\bar{M} \vDash [a] \in [C] \quad \text{syss} \quad M \vDash (t_1 = t_2)[a, \mathbf{a}],$$

luego $\bar{M} \vDash \bigwedge x_0(x_0 \in [C] \leftrightarrow t_1 = t_2)[\mathbf{a}]$, y esto implica el caso de comprensión correspondiente a $t_1 = t_2$.

El caso correspondiente a una fórmula $\phi \equiv X = Y$ es trivial, pues el conjunto $\{x \mid X = Y\}$ es necesariamente $\mathbb{N}$ o $\varnothing$, y su existencia queda asegurara por el axioma de comprensión aplicado a las fórmulas $x = x$ o $x \neq x$.

Si $\phi \equiv t(x_0, y_1, \ldots, y_n) \in Y$, fijamos una sucesión $\mathbf{a}$ en M de interpretaciones de las variables $x_1, \ldots, x_n$ y un conjunto $C \equiv [\phi_i, \mathbf{b}]$ que interprete a Y. No perdemos generalidad si suponemos que las variables libres de ϕ_i distintas de x_0 son $x_1, \ldots, x_r$, ninguna de las cuales coincide con $y_0, \ldots, y_n$. Consideramos entonces $D \equiv [\mathbf{S}^t_{x_0}\phi_i, \mathbf{a}, \mathbf{b}]$. De este modo, un a en M, fijada una valoración v que interprete las variables y_i con $\mathbf{a}$ y las y_i con $\mathbf{b}$, se cumple

$$\bar{M} \vDash x_0 \in X[v^{a\ D}_{x_0 X}] \quad \text{syss} \quad M \vDash \mathbf{S}^t_{x_0}\phi_i[v^a_{x_0}] \quad \text{syss} \quad M \vDash \phi_i[v_{x_0}^{M(t)[v^a_{x_0}]}]$$

$$\text{syss} \quad \bar{M} \vDash u \in Y[v_{Yu}^{CM(t)[v^a_{x_0}]}] \quad \text{syss} \quad \bar{M} \vDash t \in Y[v^{Ca}_{Yx_0}].$$

Por lo tanto,

$$\bar{M} \vDash \bigwedge x_0(x_0 \in X \leftrightarrow t \in Y)[v^{DC}_{XY}],$$

luego

$$\bar{M} \vDash \bigvee X \bigwedge x_0(x_0 \in X \leftrightarrow t \in Y)[v^C_Y],$$

lo que nos da el caso correspondiente del axioma de comprensión.

Si existen conjuntos $C \equiv [\phi_i, \mathbf{a}]$ y $D \equiv [\phi_j, \mathbf{b}]$ (donde podemos suponer que las fórmulas ϕ_i y ϕ_j sólo tienen la variable x_0 en común) que cumplen el axioma de comprensión para las fórmulas ϕ y ψ respectivamente, entonces es claro que $[\neg\phi_i, \mathbf{a}]$ lo verifica para $\neg\phi$ y que $[\phi_i \to \phi_j, \mathbf{a}, \mathbf{b}]$ lo verifica para $\phi \to \psi$, así como que $[\bigwedge x_k\phi_i, \mathbf{a}]$ lo verifica para $\bigwedge x_k\phi_i$. Por lo tanto, $\bar{M}$ satisface el axioma de comprensión aritmética y, por consiguiente, es un modelo de ACA_0.

Ahora ya podemos concluir: Si la traducción de una sentencia ϕ de $\mathcal{L}_a$ no es demostrable demostrable en AP, entonces el teorema de completitud de Gödel (teorema 4.17) nos da que existe un modelo (que podemos suponer numerable) M de AP tal que $M \vDash \neg\phi$, pero hemos visto que entonces $\bar{M} \vDash \neg\bar{\phi}$ y, como $\bar{M}$ es un modelo de ACA_0, concluimos que la traducción $\bar{\phi}$ no es un teorema de ACA_0. ∎

En particular, el teorema anterior nos da que ACA_0 es consistente (y, de hecho, la prueba nos muestra cómo construirle un modelo,[6] partiendo, por ejemplo, del modelo natural de AP). No podemos decir lo mismo de AP_2.

Así pues, ACA_0 y AP demuestran exactamente los mismos teoremas aritméticos (es decir, teoremas expresables en el lenguaje $\mathcal{L}_a$), al igual que sucede con ZF−AI. Sin embargo, ACA_0 permite formalizar y demostrar resultados que no son formalizables en AP, como el teorema de completitud semántica, que es demostrable, de hecho, en una subteoría de ACA_0 [LF 7.37].

[6] No obstante, la prueba no es constructiva pues, para empezar, no tenemos forma de saber cuándo se cumple la relación $(\phi_i, \mathbf{a}) \sim (\phi_j, \mathbf{b})$ que determina el universo del modelo. En particular, el teorema anterior no indica cómo obtener una demostración en AP de una sentencia a partir de una demostración de su traducción en ACA_0. Sin embargo, el teorema admite una prueba constructiva [LF 7.6].

En la práctica, no obstante, hay bastante diferencia, ya que ACA_0 lleva "implementada de fábrica" una relación de pertenencia, por lo que no necesitamos el esfuerzo que hemos tenido que hacer en AP (en $\mathrm{I}\Sigma_1$, de hecho) para definir una relación de pertenencia que satisfaga los axiomas conjuntistas usuales. En ACA_0, por el contrario, basta demostrar que los pares ordenados $\langle x, y\rangle$ se comportan adecuadamente y a partir de ahí podemos definir el producto cartesiano de conjuntos

$$A \times B \equiv \{z \mid \bigvee xy(x \in A \wedge y \in B \wedge z = \langle x, y\rangle)\},$$

y a su vez podemos definir de forma obvia los conceptos conjuntistas de relación, función, etc. y demostrar sus propiedades básicas. Por ejemplo, el teorema 5.57, que en AP es un esquema teoremático, en ACA_0 puede expresarse como un único teorema, en el que en lugar de partir de dos fórmulas $x \in X$ y ψ, ahora partimos de dos conjuntos X y G, sin ninguna alusión a fórmulas o cualquier otro concepto metamatemático, ni en el enunciado ni en la demostración.

Teorema $\bigwedge GXa(G : \mathbb{N} \times X \longrightarrow X \wedge a \in X \to \overset{1}{\bigvee} F(F : \mathbb{N} \longrightarrow X$
$\wedge\, F(0) = a \wedge \bigwedge n\, F(n) = G(n, F(n))))$.

Todos los argumentos de la sección 6.6 se formalizan mucho más fácilmente en ACA_0 (en parte también porque allí trabajábamos en $\mathrm{I}\Sigma_1$ en lugar de en AP, y ahora no tenemos que preocuparnos de rastrear la complejidad de las fórmulas que usamos). En particular podemos definir los conjuntos numéricos $\mathbb{Z}$ y $\mathbb{Q}$. Incluso podemos definir cómodamente los conceptos algebraicos de grupo, anillo, cuerpo, etc. y demostrar resultados generales.

Observemos también que ACA_0 nos ha permitido reducir el esquema de inducción de AP a un único axioma. Tal vez el lector objete que se trata de una ventaja muy relativa, puesto que para que este principio de inducción sirva para algo necesitamos acompañarlo del esquema de comprensión aritmética, que es nuevamente un esquema axiomático. Sin embargo, no es lo mismo: mientras que hemos demostrado que AP no es finitamente axiomatizable, resulta que ACA_0 sí que lo es:

Teorema 10.3 ACA_0 *es finitamente axiomatizable.*

Demostración: Vamos a probar que las sentencias que listamos más abajo (todas las cuales son teoremas de ACA_0) permiten probar (en ACA_0 sin el axioma de comprensión aritmética) todos los casos particulares de dicho esquema, luego añadidos a los axiomas restantes de ACA_0 constituyen una axiomatización finita de esta teoría.[7] Notemos que del primero de ellos se deduce que

$$\langle x, y\rangle \equiv z|2z = (x + y)(x + y + 1) + 2x$$

[7]Alternativamente, podemos tomar como axiomas todos los axiomas de ACA_0 menos el esquema de comprensión aritmética, más el conjunto finito casos particulares de dicho esquema necesarios para demostrar las sentencias anteriores.

es una descripción propia. Escribiremos

$$\langle x\rangle \equiv x, \qquad \langle x_1,\dots,x_n\rangle \equiv \langle \langle x_1,\dots,x_{n-1}\rangle, x_n\rangle.$$

Así, por ejemplo, $\langle x,y,z\rangle$ es una abreviatura por $\langle\langle x,y\rangle, z\rangle$.

Los axiomas alternativos al esquema de comprensión aritmética son los siguientes:

Par 1	$\bigwedge xy\overset{1}{\bigvee}z\ 2z = (x+y)(x+y+1)+2x$
Par 2	$\bigwedge xyzw(\langle x,y\rangle = \langle z,w\rangle \to x = z \wedge y = w)$
Conjunto unitario	$\bigwedge x\bigvee A\bigwedge n(n\in A \leftrightarrow n = x)$
Sucesor	$\bigvee A\bigwedge xy(\langle x,y\rangle \in A \leftrightarrow y = x')$
Suma	$\bigvee A\bigwedge xyz(\langle x,y,z\rangle \in A \leftrightarrow x = y+z)$
Producto	$\bigvee A\bigwedge xyz(\langle x,y,z\rangle \in A \leftrightarrow x = yz)$
Intersección	$\bigwedge XY\bigvee Z\bigwedge u(u\in Z \leftrightarrow u\in X \wedge u\in Y)$
Complemento	$\bigwedge X\bigvee Y\bigwedge u(u\in Y \leftrightarrow u\notin X)$
Dominio	$\bigwedge A\bigvee B\bigwedge x(x\in B \leftrightarrow \bigvee y\ \langle x,y\rangle \in A)$
Producto cartesiano	$\bigwedge A\bigvee B\bigwedge xy(\langle x,y\rangle \in B \leftrightarrow x\in A)$
Relación inversa	$\bigwedge A\bigvee B\bigwedge xy(\langle x,y\rangle \in B \leftrightarrow \langle y,x\rangle \in A)$
Identidad	$\bigvee A\bigwedge xy(\langle x,y\rangle \in A \leftrightarrow x = y)$
Permutación	$\bigwedge A\bigvee B\bigwedge xyz(\langle x,y,z\rangle \in B \leftrightarrow \langle y,z,x\rangle \in A)$
Permutación	$\bigwedge A\bigvee B\bigwedge xyz(\langle x,y,z\rangle \in B \leftrightarrow \langle x,z,y\rangle \in A)$
Cuádrupla	$\bigwedge A\bigvee B\bigwedge xyzw(\langle x,y,z,w\rangle \in B \leftrightarrow \langle y,z,w\rangle \in A)$

Es inmediato que todas estas sentencias son teoremas de ACA_0. Las dos primeras porque son traducciones de teoremas de AP, y las demás se siguen fácilmente del esquema de comprensión aritmética. Por ejemplo, para probar el axioma de la relación inversa basta tomar

$$B \equiv \{n \mid \bigvee xy(n = \langle x,y\rangle \wedge \langle y,x\rangle \in A)\}.$$

Ahora vamos a ver que con estos axiomas es posible demostrar el esquema de comprensión aritmética. En primer lugar observamos que los axiomas de la intersección y el complemento nos permiten definir $A\cap B$, $A\cup B$, $\mathbb{N}$, $\mathbb{N}\setminus A$ y $\varnothing$.

De los axiomas del par se puede deducir cualquier caso particular de los esquemas siguientes:

$$\bigwedge x_1\cdots x_n y_1\cdots y_n(\langle x_1,\dots,x_n\rangle = \langle y_1,\dots,y_n\rangle \to x_1 = y_1 \wedge\cdots\wedge x_n = y_n)$$

$$\bigwedge x_1\cdots x_{n+p}(\langle\langle x_1,\dots,x_n\rangle, x_{n+1},\dots,x_{n+p}\rangle = \langle x_1,\dots,x_{n+p}\rangle).$$

El axioma de extensionalidad nos da que el conjunto cuya existencia postula el axioma del dominio es único, por lo que podemos definir

$$\mathcal{D}A \equiv B|\bigwedge x(x\in B \leftrightarrow \bigvee y\ \langle x,y\rangle \in A).$$

Aplicando a $\mathcal{D}A$ el axioma de la relación inversa (junto con el axioma de extensionalidad) obtenemos que

$$\bigwedge A\overset{1}{\bigvee}B\bigwedge x(x\in B \leftrightarrow \bigvee y\ \langle y,x\rangle \in A),$$

lo que nos permite definir

$$\mathcal{R}A \equiv B|\bigwedge x(x \in B \leftrightarrow \bigvee y\ \langle y, x\rangle \in A).$$

Enumeramos a continuación algunos resultados más:

1) $\bigwedge A\bigvee B\bigwedge xy(\langle x, y\rangle \in B \leftrightarrow y \in A)$.
   Por el axioma de la relación inversa al axioma del producto cartesiano.

2a) $\bigwedge A\bigvee B\bigwedge xyz(\langle x, y, z\rangle \in B \leftrightarrow \langle x, y\rangle \in A)$

2b) $\bigwedge A\bigvee B\bigwedge xyz(\langle x, z, y\rangle \in B \leftrightarrow \langle x, y\rangle \in A)$

2c) $\bigwedge A\bigvee B\bigwedge xyz(\langle z, x, y\rangle \in B \leftrightarrow \langle x, y\rangle \in A)$
   Por el axioma del producto cartesiano $\bigwedge A\bigvee B\bigwedge wz(\langle w, z\rangle \in B \leftrightarrow w \in A)$. Haciendo $w = \langle x, y\rangle$ tenemos a), y aplicando los axiomas de permutación obtenemos b) y c).

3) $\bigwedge A\bigvee B\bigwedge x_1 \cdots x_n y(\langle x_1, \ldots, x_n, y\rangle \in B \leftrightarrow \langle x_1, \ldots, x_n\rangle \in A)$.
   Por el axioma del producto cartesiano, haciendo $x = \langle x_1, \ldots, x_n\rangle$.

4) $\bigwedge A\bigvee B\bigwedge x_1 \cdots x_n y_1 \cdots y_k(\langle x_1, \ldots, x_n, y_1, \ldots, y_k\rangle \in B \leftrightarrow \langle x_1, \ldots, x_n\rangle \in A)$
   Por 3) aplicado k veces.

5) $\bigwedge A\bigvee B\bigwedge x_1 \cdots x_n y(\langle x_1, \ldots, x_{n-1}, y, x_n\rangle \in B \leftrightarrow \langle x_1, \ldots, x_n\rangle \in A)$
   Por 2b)

6) $\bigwedge A\bigvee B\bigwedge y_1 \cdots y_k x_1 x_2(\langle y_1, \ldots, y_k, x_1, x_2\rangle \in B \leftrightarrow \langle x_1, x_2\rangle \in A)$
   Por 2c).

Veamos ahora que, si $t(x_1, \ldots, x_n)$ es un término aritmético de primer orden sin descriptores cuyas variables libres están entre las indicadas, podemos probar (por inducción sobre la longitud de t) que

$$\bigvee A\bigwedge x_1 \cdots x_n y(\langle x_1, \ldots, x_n, y\rangle \in A \leftrightarrow y = t(x_1, \ldots, x_n)).$$

En efecto, si $t \equiv x_i$ por el axioma de la identidad existe un B_1 tal que

$$\bigwedge x_i y(\langle x_i, y\rangle \in B_1 \leftrightarrow x_i = y).$$

Si $i > 1$ aplicamos 6) para concluir que existe un B_2 tal que

$$\bigwedge x_1 \cdots x_i y(\langle x_1, \ldots, x_i, y\rangle \in B_2 \leftrightarrow x_i = y).$$

Si $i < n$ aplicamos 5) varias veces para concluir que existe un A tal que

$$\bigwedge x_1 \cdots x_n y(\langle x_1, \ldots, x_n, y\rangle \in A \leftrightarrow x_i = y).$$

Si $t \equiv 0$, por el axioma del conjunto unitario $\bigvee B\bigwedge y(y \in B \leftrightarrow y = 0)$. Basta aplicar 1) con $x = \langle x_1, \ldots, x_n\rangle$.

Si $t \equiv t'_0$, por hipótesis de inducción existe un conjunto B tal que

$$\bigwedge x_1 \cdots x_n z(\langle x_1, \ldots, x_n, z\rangle \in B \leftrightarrow z = t_0(x_1, \ldots, x_n)).$$

Por otro lado, por el axioma del sucesor y de la relación inversa, existe un C tal que

$$\bigwedge yz(\langle y, z\rangle \in C \leftrightarrow y = z').$$

Por 5) existe B' tal que

$$\bigwedge x_1 \cdots x_n yz(\langle x_1, \ldots, x_n, y, z\rangle \in B' \leftrightarrow \langle x_1, \ldots, x_n, z\rangle \in B).$$

Por 6) existe C' tal que

$$\bigwedge x_1 \cdots x_n yz(\langle x_1, \ldots, x_n, y, z\rangle \in C' \leftrightarrow \langle y, z\rangle \in C).$$

Basta tomar $A \equiv \mathcal{D}(B' \cap C')$.

Supongamos ahora que $t \equiv t_1 + t_2$. Sean B_1 y B_2 tales que

$$\bigwedge x_1 \cdots x_n z(\langle x_1, \ldots, x_n, z\rangle \in B_1 \leftrightarrow z = t_1(x_1, \ldots, x_n)),$$

$$\bigwedge x_1 \cdots x_n z(\langle x_1, \ldots, x_n, z\rangle \in B_2 \leftrightarrow z = t_2(x_1, \ldots, x_n)).$$

Por el axioma de la suma existe un C tal que

$$\bigwedge x z_1 z_2(\langle y, z_1, z_2\rangle \in C \leftrightarrow y = z_1 + z_2).$$

Sean B'_1 y B'_2 tales que

$$\bigwedge x_1 \cdots x_n y z_1 z_2(\langle x_1, \ldots, x_n, y, z_1, z_2\rangle \in B'_1 \leftrightarrow \langle x_1, \ldots, x_n, z_1\rangle \in B_1)$$

$$\bigwedge x_1 \cdots x_n y z_1 z_2(\langle x_1, \ldots, x_n, y, z_1, z_2\rangle \in B'_2 \leftrightarrow \langle x_1, \ldots, x_n, z_2\rangle \in B_2)$$

La existencia de B'_1 se sigue del axioma del producto cartesiano aplicado al conjunto dado por 5). La existencia de B'_2 se sigue de aplicar 5) dos veces. Por el axioma de la cuádrupla existe C' tal que

$$\bigwedge x_1 \cdots x_n y z_1 z_2(\langle x_1, \ldots, x_n, y, z_1, z_2\rangle \in C' \leftrightarrow \langle y, z_1, z_2\rangle \in C).$$

Ahora basta tomar $A = \mathcal{D}\mathcal{D}(B'_1 \cap B'_2 \cap C')$.

El caso $t \equiv t_1 \cdot t_2$ es análogo. Veamos ahora que si $\phi(x_1, \ldots, x_n, X_1, \ldots, X_r)$ es una fórmula aritmética sin descriptores se puede probar

$$\bigwedge X_1 \cdots X_r \bigvee A \bigwedge x_1 \cdots x_n(\langle x_1 \ldots, x_n\rangle \in A \leftrightarrow \phi(x_1, \ldots, x_n, X_1, \ldots, X_n)).$$

En efecto, si $\phi \equiv t_1 = t_2$ hemos probado que existen conjuntos B_1 y B_2 tales que

$$\bigwedge x_1 \cdots x_n y(\langle x_1, \ldots, x_n, y\rangle \in B_1 \leftrightarrow y = t_1),$$

$$\bigwedge x_1 \cdots x_n y(\langle x_1, \ldots, x_n, y\rangle \in B_2 \leftrightarrow y = t_2).$$

Basta tomar $A = \mathcal{D}(B_1 \cap B_2)$.

Si $\phi \equiv t \in X$, sabemos que existe un B tal que

$$\bigwedge x_1 \cdots x_n y(\langle x_1, \ldots, x_n, y\rangle \in B \leftrightarrow y = t),$$

y por 1) existe un C tal que

$$\bigwedge x_1 \cdots x_n y(\langle x_1, \ldots, x_n, y\rangle \in C \leftrightarrow y \in X).$$

Basta tomar $A = \mathcal{D}(B \cap C)$.

Si $\phi \equiv X = Y$, basta tomar $A = \mathbb{N}$ o bien $A = \varnothing$.

Si el resultado vale para α y β y A_1 y A_2 son los conjuntos correspondientes, entonces $\mathbb{N} \setminus A_1$ cumple el resultado para $\neg\alpha$, mientras que $(\mathbb{N} \setminus A_1) \cup A_2$ lo cumple para $\alpha \to \beta$.

Supongamos finalmente que $\phi \equiv \bigwedge x\psi$. Por hipótesis de inducción existe un conjunto B tal que

$$\bigwedge x_1 \cdots x_n x(\langle x_1, \ldots, x_n, x\rangle \in B \leftrightarrow \psi(x_1, \ldots, x_n, x)).$$

Basta tomar $A = \mathbb{N} \setminus \mathcal{D}(\mathbb{N} \setminus B)$.

Finalmente observamos que, para toda fórmula aritmética sin descriptores $\phi(x, x_1, \ldots, x_s, X_1, \ldots, X_r)$, se cumple

$$\bigwedge X_1 \cdots X_r \bigwedge x_1 \cdots x_s \bigvee X \bigwedge x(x \in X \leftrightarrow \phi(x)).$$

En efecto, hemos probado que existe un B tal que

$$\bigwedge x_1 \cdots x_s x(\langle x_1, \ldots, x_s, x\rangle \in B \leftrightarrow \phi(x, x_1, \ldots, x_n)).$$

Por el axioma del conjunto unitario existe un C tal que

$$\bigwedge x(x \in C \leftrightarrow x = \langle x_1, \ldots, x_s\rangle).$$

Por el axioma de producto cartesiano existe un C' tal que

$$\bigwedge yx(\langle y, x\rangle \in C' \leftrightarrow y = \langle x_1, \ldots, x_n\rangle).$$

Basta tomar $X = \mathcal{R}(B \cap C')$, pues entonces

$$x \in X \leftrightarrow \bigvee y\ \langle y, x\rangle \in B \cap C' \leftrightarrow \langle x_1, \ldots, x_n, x\rangle \in B \leftrightarrow \phi(x, x_1, \ldots, x_n).$$

■

Sucesiones en $\mathbf{ACA_0}$ Una característica notable de ACA_0 es que permite definir sucesiones infinitas de conjuntos. En efecto, basta definir

$$X_n \equiv \{m \mid \langle n, m\rangle \in X\}$$

y así podemos ver cada conjunto X como una sucesión de conjuntos $\{X_n\}_{n\in\mathbb{N}}$. En particular podemos definir n-tuplas de conjuntos

$$(X_1, \ldots, X_n) \equiv X | \bigwedge u(u \in X \leftrightarrow$$
$$\bigvee y((y \in X_1 \wedge u = \langle 0^{(1)}, y\rangle) \vee \cdots \vee (y \in X_n \wedge u = \langle 0^{(n)}, y\rangle))).$$

La formalización de la lógica en $\mathbf{ACA}_0$ En ACA_0 tenemos definidos los conceptos de lenguaje formal, teoría axiomática etc., puesto que ya hemos visto que son definibles en AP, pero, como cabe esperar, la definición se simplifica en este contexto. Para empezar, no es necesario definir un lenguaje formal como cinco designadores y cuatro fórmulas, tal y como hacíamos en 8.1, sino que podemos definir una única fórmula que especifique cuándo un conjunto $\mathcal{L}$ es un lenguaje formal. Dicha fórmula debe expresar que un conjunto $\mathcal{L}$ es lenguaje formal si es una nónupla (en el sentido que acabamos de definir)

$$\mathcal{L} = (\{\ulcorner\neg\urcorner\}, \{\ulcorner\to\urcorner\}, \{\ulcorner\bigwedge\urcorner\}, \{\ulcorner|\urcorner\}, \{\ulcorner=\urcorner\}, V, R, F, \mathrm{Nar}),$$

donde los conjuntos $\{\ulcorner\neg\urcorner\}, \{\ulcorner\to\urcorner\}, \{\ulcorner\bigwedge\urcorner\}, \{\ulcorner|\urcorner\}, V, R, F$ son disjuntos dos a dos, $\ulcorner=\urcorner \in R$, $\mathrm{Nar} : R \cup F \longrightarrow \mathbb{N}$, etc.

Similarmente, para cada lenguaje formal podemos definir su conjunto de términos, fórmulas, etc.

Ahora estamos en condiciones de probar un teorema en AP_2 que no es demostrable en ACA_0, lo que nos muestra que ambas teorías no son equivalentes. Nos centramos en el lenguaje $\ulcorner\mathcal{L}_a\urcorner$.

Diremos que v es una *valoración* de una expresión $\theta \in \mathrm{Exp}(\ulcorner\mathcal{L}_a\urcorner)$ si v es una función definida sobre un conjunto (finito) de variables de $\ulcorner\mathcal{L}_a\urcorner$ que incluya a todas las variables libres en θ. Lo representaremos por $\mathrm{Val}(v, \theta)$. De acuerdo con lo visto en la sección 8.1, podemos ver a $\mathrm{Val}(v, \theta)$ como una fórmula de $\mathcal{L}_a$. Consideramos el término dado por:

$$v_x^a = (v \setminus \{(x, v(x))\}) \cup \{(x, a)\},$$

de modo que si $\mathrm{Val}(v, \theta)$ y x es una variable, entonces v_x^a es otra valoración de θ que coincide con v salvo a lo sumo en x, donde toma el valor a (aunque en principio no es preciso que v esté definida en x). Definimos el conjunto

$$D \equiv \{\langle \theta, v\rangle \mid \theta \in \mathrm{Exp}(\ulcorner\mathcal{L}_a\urcorner) \wedge \mathrm{Val}(v, \theta)\}.$$

Teorema 10.4 ($\mathbf{AP}_2$) *Existe una única función $F : D \longrightarrow \mathbb{N}$ que cumple las propiedades siguientes:*

1. $F(x, v) = v(x)$,
2. $F(\ulcorner 0\urcorner, v) = 0$,
3. $F(St, v) = F(t, v) + 1$,
4. $F(t_1 + t_2, v) = F(t_1, v) + F(t_2, v)$,
5. $F(t_1 \cdot t_2, v) = F(t_1, v) \cdot F(t_2, v)$,
6. $F(t_1 = t_2, v) = \begin{cases} 1 & \text{si } F(t_1, v) = F(t_2, v), \\ 0 & \text{en otro caso,} \end{cases}$
7. $F(\neg\alpha, v) = 1 - F(\alpha, v)$,
8. $F(\alpha \to \beta, v) = \begin{cases} 1 & \text{si } F(\alpha, v) = 0 \vee F(\beta, v) = 1, \\ 0 & \text{en otro caso,} \end{cases}$

9. $F(\bigwedge x\alpha, v) = \begin{cases} 1 & \textit{si } \bigwedge n\ F(\alpha, v^n_x) = 1, \\ 0 & \textit{en otro caso,} \end{cases}$

10. $F(x|\alpha, v) = \begin{cases} a & \textit{si } \bigwedge n(F(\alpha, v^n_x) = 1 \leftrightarrow n = a), \\ 0 & \textit{si no existe tal } a. \end{cases}$

DEMOSTRACIÓN: Llamamos $D_n \equiv \{\langle\theta, v\rangle \mid \langle\theta, v\rangle \in D \wedge \theta < n\}$ y sea

$$X \equiv \{n \mid \bigvee F(F : D_n \longrightarrow \mathbb{N} \wedge \cdots)\},$$

donde los puntos suspensivos son todas las condiciones del enunciado.[8] Es inmediato comprobar por inducción que $X = \mathbb{N}$: la condición $0 \in X$ es trivial y, supuesto $n \in X$, tomamos $F : D_n \longrightarrow \mathbb{N}$ según la definición y la extendemos a D_{n+1} de la única forma permitida por las propiedades. Ahora basta definir[9]

$$F \equiv \{\langle\langle\theta, v\rangle\, k\rangle \mid \langle\theta, v\rangle \in D$$

$$\wedge \bigvee nG\,(\langle\theta, v\rangle \in D_n \wedge G : D_n \longrightarrow \mathbb{N} \wedge G(\theta, n) = k \wedge \cdots)\},$$

donde los puntos suspensivos son nuevamente las condiciones del enunciado (sobre la función G). Es fácil comprobar que F cumple lo pedido. ∎

Definición 10.5 Llamaremos $\mathbb{N}(\ )[\ ]$ a la restricción de la función F dada por el teorema anterior al conjunto $D_t \equiv \{\langle t, v\rangle \mid t \in \mathrm{Term}(\ulcorner\mathcal{L}_a\urcorner) \wedge \langle t, v\rangle \in D\}$, mientras que $\mathbb{N} \vDash (\ )[\ ]$ será la relación

$$\{\langle\alpha, v\rangle \mid \langle\alpha, v\rangle \in D \wedge \alpha \in \mathrm{Form}(\ulcorner\mathcal{L}_a\urcorner) \wedge F(\alpha, v) = 1\}.$$

Se cumple entonces:

Teorema 10.6 ($\mathbf{AP}_2$) *Si v es una valoración definida sobre los términos o las fórmulas correspondientes a cada apartado, se cumple*

1. $\mathbb{N}(x)[v] = v(x)$,
2. $\mathbb{N}(\ulcorner 0\urcorner)[v] = 0$,
3. $\mathbb{N}(t')[v] = \mathbb{N}(t)[v] + 1$,
4. $\mathbb{N}(t_1 + t_2)[v] = N(t_1)[v] + \mathbb{N}(t_2)[v]$,
5. $\mathbb{N}(t_1 \cdot t_2)[v] = N(t_1)[v] \cdot \mathbb{N}(t_2)[v]$,
6. $\mathbb{N} \vDash (t_1 = t_2)[v] \leftrightarrow \mathbb{N}(t_1)[v] = \mathbb{N}(t_2)[v]$,
7. $\mathbb{N} \vDash \neg\alpha[v] \leftrightarrow \neg\mathbb{N} \vDash \alpha[v]$,

[8] Notemos que la existencia de X no se sigue del axioma de comprensión aritmética, sino que requiere el axioma de comprensión completo de AP_2, pues la fórmula que lo define tiene un cuantificador de segundo orden $\bigvee F$.

[9] Nuevamente la fórmula que define a G no es aritmética, por el cuantificador $\bigvee G$.

8. $\mathbb{N} \vDash (\alpha \to \beta)[v] \leftrightarrow (\neg \mathbb{N} \vDash \alpha[v] \vee \mathbb{N} \vDash \beta[v])$,

9. $\mathbb{N} \vDash \bigwedge x\alpha[v] \leftrightarrow \bigwedge n\ \mathbb{N} \vDash \alpha[v^n_x]$,

10. $\mathbb{N}(x|\alpha)[v] = \begin{cases} a & \textit{si } a \textit{ es el único que cumple } \mathbb{N} \vDash \alpha[v^a_x], \\ 0 & \textit{si no existe un único } a \textit{ en estas condiciones.} \end{cases}$

Escribiremos $\mathbb{N} \vDash \alpha \equiv \bigwedge v(\mathrm{Val}(v, \alpha) \to \mathbb{N} \vDash \alpha[v])$. A partir de aquí es sencillo formalizar la demostración del teorema de corrección en AP_2, y demostrar así que

$$\bigwedge \alpha \in \mathrm{Form}(\ulcorner \mathcal{L}_a \urcorner)(\underset{\ulcorner \mathrm{AP} \urcorner}{\vdash} \alpha \to \mathbb{N} \vDash \alpha).$$

Esto supone demostrar que todos los axiomas lógicos y todos los axiomas de AP cumplen $\mathbb{N} \vDash \alpha$ y que esto se conserva al aplicar las reglas de inferencia. Para ello hace falta formalizar la demostración del teorema 1.12. Nada de todo esto presenta dificultad alguna. Como consecuencia:

Teorema 10.7 *Se cumple*

$$\underset{\mathrm{AP}_2}{\vdash} \mathrm{Consis}\,\ulcorner \mathrm{AP} \urcorner.$$

Demostración: Razonando en AP_2, si suponemos que $\underset{\ulcorner \mathrm{AP} \urcorner}{\vdash} \ulcorner 0 \neq 0 \urcorner$, tendríamos que $\mathbb{N} \vDash \ulcorner 0 \neq 0 \urcorner$, lo cual equivale a que $0 \neq 0$, contradicción. Por lo tanto, $\neg \underset{\ulcorner \mathrm{AP} \urcorner}{\vdash} \ulcorner 0 \neq 0 \urcorner$, y esto es precisamente la sentencia $\mathrm{Consis}\,\ulcorner \mathrm{AP} \urcorner$. ∎

Notemos que $\mathrm{Consis}\,\ulcorner \mathrm{AP} \urcorner$ no puede demostrarse en ACA_0, pues entonces, por 10.2, sería demostrable en AP, y por el segundo teorema de incompletitud AP sería contradictoria. Así pues, tenemos un ejemplo concreto de teorema aritmético de AP_2 que no es un teorema de ACA_0.

10.2 La teoría de conjuntos ZF*

En 3.24 introdujimos la teoría (restringida) de Zermelo Z* y probamos que es una teoría aritmética, es decir, que sus axiomas bastan para definir los números naturales y demostrar los axiomas de Peano. Aquí vamos a considerar una teoría algo más potente, la teoría (restringida) de *Zermelo-Fraenkel* ZF*, que resulta de sustituir el esquema de especificación por el de reemplazo, de modo que sus axiomas son:

Extensionalidad	$\bigwedge xy(\bigwedge u(u \in x \leftrightarrow u \in y) \to x = y)$
Par	$\bigwedge xy \bigvee z \bigwedge u(u \in z \leftrightarrow u = x \vee u = y)$
Unión	$\bigwedge x \bigvee y \bigwedge u(u \in y \leftrightarrow \bigvee v(u \in v \wedge v \in x))$
Reemplazo	$\bigwedge xyz(\phi(x,y) \wedge \phi(x,z) \to y = z)$
	$\to \bigwedge a \bigvee b \bigwedge y(y \in b \leftrightarrow \bigvee x \in a\, \phi(x,y))$,

para toda fórmula $\phi(x, y)$, tal vez con más variables libres, distintas de b.

Ante todo demostramos que el esquema de reemplazo implica el de especificación, por lo que todos los teoremas de Z* son teoremas de ZF* y por consiguiente contamos con todos los teoremas demostrados en el capítulo III en las teorías B y Z*.

Teorema 10.8 (Especificación) *Para toda fórmula $\phi(x)$ (de $\mathcal{L}_{\mathrm{tc}}$), tal vez con más variables libres, se cumple*

$$\bigwedge a \bigvee b \bigwedge x(x \in b \leftrightarrow x \in a \wedge \phi(x)).$$

Demostración: Fijado un conjunto a, definimos $\psi(x,y) \equiv \phi(x) \wedge x = y$. Claramente cumple la hipótesis del esquema de reemplazo, luego existe un b tal que

$$\bigwedge y(y \in b \leftrightarrow \bigvee x \in a\, \psi(x,y)),$$

es decir,

$$\bigwedge y(y \in b \leftrightarrow \bigvee x \in a\, (\phi(x) \wedge x = y)),$$

y esto equivale claramente a $\bigwedge y(y \in b \leftrightarrow y \in a \wedge \phi(y))$. ∎

Nota Existen distintas variantes de fórmulas conocidas igualmente como "esquema de reemplazo". La característica esencial es la hipótesis de unicidad sobre la fórmula $\phi(x,y)$. Sin ella tendríamos esquemas de recolección, como el teorema 6.7 de KP (Σ_1-recolección), que en general son más fuertes que los esquemas de reemplazo. Algunas variantes del esquema de reemplazo son

$$\bigwedge x \overset{1}{\bigvee} y\, \phi(x,y) \to \bigwedge a \bigvee b \bigwedge y(y \in b \leftrightarrow \bigvee x \in a\, \phi(x,y)),$$

$$\bigwedge a(\bigwedge x \in a \overset{1}{\bigvee} y\, \phi(x,y) \to \bigvee b \bigwedge y(y \in b \leftrightarrow \bigvee x \in a\, \phi(x,y))).$$

Es claro que la forma que hemos adoptado como axioma de ZF* implica la segunda de las fórmulas anteriores, sin más que aplicar el axioma a la fórmula $\psi(x,y) \equiv x \in a \wedge \phi(x,y)$. A su vez, la segunda forma implica trivialmente la primera. Sin embargo, ninguna de estas dos formas de reemplazo implican la que hemos tomado como axioma. Si queremos tomar como axioma cualquiera de ellas, necesitamos añadir también como axioma adicional la existencia del conjunto vacío: $\bigvee y \bigwedge u\, u \notin y$.

Admitiendo esto, para probar el esquema de reemplazo suponemos

$$\bigwedge xyz(\phi(x,y) \wedge \phi(x,z) \to y = z)$$

y tomamos un conjunto a. Distinguimos dos casos: si $\bigwedge x \in a \neg \bigvee y\, \phi(x,y)$, basta tomar $b = \varnothing$. Por el contrario, si existe un $x_0 \in a$ y un y_0 tal que $\phi(x_0,y_0)$, consideramos la fórmula $\psi(x,y) \equiv \phi(x,y) \vee (\neg \bigvee z\, \phi(x,z) \wedge y = y_0)$, de modo que claramente $\bigwedge x \overset{1}{\bigvee} y \phi(x,y)$. Esto nos da un conjunto b tal que

$$\bigwedge y(y \in b \leftrightarrow \bigvee x \in a\, \psi(x,y)).$$

Ahora bien, esto mismo se cumple con ϕ en lugar de ψ, pues si $y \in b$, entonces existe un $x \in a$ tal que $\psi(x,y)$ y esto implica $\phi(x,y)$ o bien $y = y_0$, y en ambos casos $\bigvee x \in a\, \phi(x,y)$. Recíprocamente, si y cumple esto, también cumple $\bigvee x \in a\, \psi(x,y)$, luego $y \in b$.

Notemos que la existencia del conjunto vacío se sigue del axioma de reemplazo que hemos adoptado a través del esquema de especificación, tomando $\phi(x) \equiv x \neq x$. Para confirmar que las otras formas de reemplazo no implican la existencia del conjunto vacío basta considerar un modelo M de $\mathcal{L}_{\mathrm{tc}}$ cuyo universo tenga únicamente un objeto a, con la relación de pertenencia que cumple $M \vDash [a] \in [a]$. En este modelo existe un único conjunto a con la propiedad de que $a = \{a\}$. Es una simple rutina comprobar que cumple cualquiera de las dos formas alternativas de reemplazo que hemos considerado, así como los restantes axiomas de ZF^*, pero no el axioma del conjunto vacío, ni el esquema de especificación. ∎

Como ya hemos visto en Z^*, a partir del axioma del par podemos definir los pares desordenados $\{x, y\}$ y a su vez los pares ordenados $(x, y) \equiv \{\{x\}, \{x, y\}\}$, lo que a su vez nos permite hablar (formalmente) de relaciones y funciones. El teorema siguiente no puede probarse en Z^*, sino que requiere reemplazo.[10] La prueba es esencialmente la que hemos visto para KP (teorema 6.10):

Teorema 10.9 $\bigwedge xy \overset{1}{\bigvee} z \bigwedge w(w \in z \leftrightarrow \bigvee u \in x \bigvee v \in y\, z = (u, v))$

Demostración: Aplicamos el esquema de reemplazo a la fórmula

$$\phi(u, w) \equiv w = (u, v)$$

y al conjunto x, lo que nos da la existencia de un conjunto b tal que

$$\bigwedge w(w \in b \leftrightarrow \bigvee u \in x\, w = (u, v)),$$

es decir, $b = x \times \{v\}$. En segundo lugar aplicamos reemplazo a la fórmula

$$\phi(v, t) \equiv \bigwedge w(w \in t \leftrightarrow \bigvee u \in x\, w = (u, v)),$$

es decir, a $t = x \times \{v\}$, lo que nos da la existencia de un conjunto b tal que

$$\bigwedge t(t \in b \leftrightarrow \bigvee v \in y \bigwedge w(w \in t \leftrightarrow \bigvee u \in x\, w = (u, v))),$$

de modo que

$$b = \{t \mid \bigvee v \in y\, t = x \times \{v\}\}.$$

Resulta entonces que $\bigcup b$ es el conjunto buscado. La unicidad es inmediata por el axioma de extensionalidad. ∎

Esto nos permite definir el *producto cartesiano*

$$x \times y \equiv z | \bigwedge w(w \in z \leftrightarrow \bigvee u \in x \bigvee v \in y\ z = (u, v)).$$

Los conceptos de dominio, rango, etc. de un conjunto de definen por especificación como en KP.

Ahora podemos demostrar una variante del esquema de reemplazo similar al que demostramos para KP:

[10] Alternativamente, la existencia del producto cartesiano puede probarse en Z^* más el axioma de partes, que postula la existencia del conjunto $\mathcal{P}x$ para todo conjunto x, pues, por la definición de par ordenado, $x \times y$ puede especificarse como subconjunto de $\mathcal{P}\mathcal{P}(x \cup y)$.

Teorema 10.10 *Para toda fórmula $\phi(u,v)$, tal vez con más variables libres,*

$$\bigwedge x(\bigwedge u \in x \overset{1}{\bigvee} v\, \phi(u,v) \to \bigvee f y(f : x \longrightarrow y \text{ suprayectiva} \wedge \bigwedge u \in x\, \phi(u, f(u)))).$$

DEMOSTRACIÓN: Si b es el conjunto dado por el axioma de reemplazo a partir de x, podemos definir

$$f = \{(u,v) \in x \times b \mid \phi(u,v)\},$$

y es claro entonces que f cumple lo pedido con $y = \mathcal{R}f$. ∎

Éste es el contenido esencial del esquema de reemplazo: las fórmulas que satisfacen la hipótesis de unicidad sobre un conjunto definen funciones (cuyos rangos son, en particular, conjuntos). También podemos expresarlo en términos de clases:

Teorema 10.11 *Si unas clases cumplen $F : A \longrightarrow B$ suprayectiva y A es un conjunto, entonces F y B son conjuntos.*

DEMOSTRACIÓN: En principio, las hipótesis significan que tenemos tres fórmulas, $w \in F$, $u \in A$, $v \in B$, de modo que

$$\bigwedge w(w \in F \to \bigvee uv(u \in A \wedge v \in B \wedge w = (u,v))),$$

y además $\bigwedge u \in A \overset{1}{\bigvee} v \in B((u,v) \in F)$. Por otra parte, existe un x tal que

$$\bigwedge u(u \in F \leftrightarrow u \in x),$$

luego el teorema anterior aplicado a la fórmula $(u,v) \in F$ nos da un conjunto y y una aplicación $f : x \longrightarrow y$ suprayectiva tales que $\bigwedge u \in A\,(u, f(u)) \in F$. Por consiguiente, $\bigwedge w(w \in F \leftrightarrow w \in f)$, lo que significa que F es un conjunto, e igualmente $\bigwedge v(v \in B \leftrightarrow v \in y)$, luego B es un conjunto. ∎

Conjuntos cociente Una aplicación del esquema de reemplazo es la existencia de conjuntos cociente: si x es un conjunto y R es una relación de equivalencia en x, por especificación, para cada $u \in x$ podemos definir su clase de equivalencia $[u]_R^x \equiv \{v \in x \mid v\, R\, u\}$, y por el esquema de reemplazo aplicado a la fórmula $y = [u]_R^x$ existe el conjunto cociente[11]

$$x/R \equiv y | \bigwedge z(z \in y \leftrightarrow \bigvee u \in x\, z = [u[_R^x).$$

Equivalentemente, el conjunto cociente es el conjunto de todas las clases de equivalencia:

$$x/R = \{[u]_R^x \mid u \in x\}.$$

Más aún, el axioma de reemplazo nos da la aplicación canónica $p : x \longrightarrow x/R$ suprayectiva que cumple $p(u) = [u]_R^x$. ∎

[11] Como en el caso del producto cartesiano, la existencia de x/R puede justificarse también por especificación en $\mathcal{PP}x$, si admitimos el axioma de partes, y entonces no se requiere el axioma de reemplazo.

Para terminar destacamos que el teorema 5.56 muestra que todos los axiomas de ZF^* son demostrables en AP, de modo que AP interpreta a ZF^* y, en particular, la consistencia de AP implica la de ZF^*.

A decir verdad, los resultados de esta sección no muestran ninguna ventaja esencial de trabajar en ZF^* en lugar de en Z^*. Aquí nos hemos limitado a mostrar las consecuencias y equivalencias básicas del esquema de reemplazo, pero en el capítulo siguiente se verá lo que aporta realmente a la teoría de conjuntos.

10.3 Las teorías de conjuntos NBG^* y MK^*

Puesto que ZF^* extiende a Z^* y a B, tenemos que en ZF^* es posible considerar clases que no son conjuntos, como la clase V de todos los conjuntos, la clase Ω de todos los ordinales o la clase R de los conjuntos que no se pertenecen a sí mismos. También tenemos la clase ω de los números naturales, de la que no podemos demostrar que sea un conjunto, pero tampoco que no lo sea (precisamente esto es lo que afirma el axioma de infinitud, que consideraremos más adelante).

En esta sección vamos a construir dos teorías que formalicen la noción de clase propia exactamente del mismo modo que las teorías ACA_0 y AP_2 formalizan los conjuntos infinitos de números naturales definidos mediante fórmulas en AP. En realidad, si calcáramos el procedimiento que hemos empleado en la construcción de ACA_0, lo que obtendríamos sería esencialmente la llamada *teoría de von Neumann-Bernays.* Se trata de una teoría introducida por Paul Bernays en la que traducía otra teoría previa de John von Neumann que, más que una teoría de conjuntos, era una teoría de funciones, pues sus conceptos primitivos eran los de "función" y "argumento". Bernays vio que la teoría de von Neumann podía reformularse sustituyendo el concepto de "función" por el de "clase" y el de "argumento" por el de "conjunto", y el resultado es, como acabamos de indicar, una teoría totalmente análoga a ACA_0, en la que las variables se dividen en dos tipos (unas para clases y otras para conjuntos) y en la que hay también dos relatores de pertenencia (uno para clases y otro para conjuntos).

Ahora bien, mientras que en el caso de la aritmética de segundo orden podíamos considerar a los números naturales y a los conjuntos como objetos completamente distintos (un objeto, o bien es un número natural, o bien es un conjunto de números naturales, pero no ambas cosas) ahora nos encontramos con que tanto las clases como los conjuntos pretenden representar colecciones de objetos (colecciones de conjuntos, concretamente) y se plantea el problema de que una clase y un conjunto pueden tener la misma extensión (y, a pesar de ello, ser dos objetos distintos). Naturalmente, Bernays tuvo esto en cuenta, pero la lógica subyacente a su teoría era un tanto sofisticada. Sin embargo, Kurt Gödel simplificó drásticamente la situación al observar que en realidad no era necesario establecer diferencia alguna entre clases y conjuntos, sino que directamente se puede considerar que los conjuntos son clases (pero de forma que no toda clase es un conjunto). Así se obtiene la teoría de conjuntos de *von Neumann-Bernays-Gödel* (NBG) que vamos a presentar a continuación (salvo

algunos axiomas que omitiremos hasta el capítulo próximo), y cuya construcción sólo requiere una ligera variante respecto a la construcción que hemos dado de ACA_0.

Concretamente, sólo tenemos que observar que tanto los conjuntos de un modelo de ZF^* como las clases formadas por conjuntos de dicho modelo que satisfacen una determinada fórmula tienen en común que son colecciones de conjuntos, y que una característica que distingue a los unos de las otras es que todo conjunto x pertenece a otros conjuntos (como por ejemplo al conjunto $\{x\}$) y a otras clases (como a la clase V de todos los conjuntos), mientras que una clase propia no puede pertenecer a ningún conjunto ni a ninguna clase (pues unos y otras son colecciones de conjuntos).

Esto hace que, en lugar de añadir al lenguaje $\mathcal{L}_{\mathrm{tc}}$ de ZF^* un relator monádico que distinga las clases de los conjuntos (como el relator cto de $\mathcal{L}_a^2$ distingue los conjuntos de los números naturales), podemos mantener inalterado el lenguaje $\mathcal{L}_{\mathrm{tc}}$ y definir en él la fórmula

$$\mathrm{cto}\, X \equiv \bigvee Y\, X \in Y.$$

Esto significa que en NBG^* los conjuntos se definen como las clases que pertenecen a al menos otra clase.

A diferencia de lo que hemos hecho al definir ACA_0 y AP_2, no vamos a dividir las variables de $\mathcal{L}_{\mathrm{tc}}$ en dos tipos ni vamos a introducir un concepto de expresión estructurada, sino que únicamente adoptaremos (al menos provisionalmente y por mera comodidad) el convenio de nombrar las variables con letras mayúsculas y usar las minúsculas para indicar que representan conjuntos, es decir, convenimos en que

$$\bigwedge x\, \alpha \equiv \bigwedge X(\mathrm{cto}\, X \to \alpha), \quad \bigvee x\alpha \equiv \bigvee X(\mathrm{cto}\, X \wedge \alpha), \quad x|\alpha \equiv X|(\mathrm{cto}\, X \wedge \alpha).$$

Definición 10.12 La *teoría de conjuntos (restringida) de Morse-Kelley* es la teoría MK^* sobre $\mathcal{L}_{\mathrm{tc}}$ determinada por los axiomas siguientes:

Extensionalidad	$\bigwedge XY(\bigwedge u(u \in X \leftrightarrow u \in Y) \to X = Y)$
Comprensión	$\bigvee X \bigwedge u(u \in X \leftrightarrow \phi(u))$ $\quad (*)$
Vacío	$\bigvee x \bigwedge u\, u \notin x$
Par	$\bigwedge xy \bigvee z \bigwedge u(u \in z \leftrightarrow u = x \vee u = y)$
Unión	$\bigwedge x \bigvee y \bigwedge u(u \in y \leftrightarrow \bigvee v(u \in v \wedge v \in x))$
Reemplazo	$\bigwedge aF(\mathrm{Un}\, F \to \bigvee b \bigwedge v(v \in b \leftrightarrow \bigvee u \in a\, (u, v) \in F))$

$(*)$ para toda fórmula $\phi(u)$ tal vez con más variables libres (distintas de X).

En el enunciado del axioma de reemplazo hemos usado la notación usual para los pares ordenados en $\mathcal{L}_{\mathrm{tc}}$, es decir, $(u, v) \equiv \{\{u\}, \{u, v\}\}$, así como la abreviatura siguiente que define las clases *unívocas:*

$$\mathrm{Un}\, F \equiv \bigwedge uvw((u, v) \in F \wedge (u, w) \in F \to v = w).$$

Admitiremos además, como axioma (semi)lógico que $\bigwedge u\, u \notin X|X = X$, de donde se sigue que toda fórmula es equivalente a otra sin descriptores (sin más que suponer el axioma de extensionalidad), por lo que el esquema de comprensión puede restringirse sin pérdida de generalidad a fórmulas sin descriptores.

Se llama *teoría de conjuntos (restringida) de von Neumann-Bernays Gödel* (NBG*) a la subteoría de MK* que resulta de restringir el axioma de comprensión a fórmulas primitivas, donde una fórmula se dice *primitiva* si no tiene descriptores y todos sus cuantificadores están restringidos a conjuntos, es decir, son de la forma $\bigwedge x$.

Las formulas equivalentes (en NBG* o en cualquier extensión prefijada) a fórmulas primitivas se llaman *fórmulas normales* y es claro que el esquema de comprensión en NBG* (o en cualquier extensión) es válido para fórmulas normales cualesquiera (en la extensión correspondiente). También es inmediato que si α y β son normales también lo son[12]

$$\neg\alpha, \quad \alpha \to \beta, \quad \alpha \vee \beta, \quad \alpha \wedge \beta, \quad \alpha \leftrightarrow \beta, \quad \bigwedge x\alpha \quad \text{y} \quad \bigvee x\beta.$$

Vamos a ver que MK* es a ZF* como AP_2 es a AP, mientras que NBG* es el análogo de ACA_0. Las fórmulas primitivas son las análogas a las fórmulas aritméticas (en sentido estricto) y las fórmulas normales son el análogo a las fórmulas aritméticas en sentido amplio.

Resultados conjuntistas básicos Empezamos a estudiar NBG* extrayendo las primeras consecuencias elementales de sus axiomas. Observemos que el lenguaje formal de NBG* es el mismo que el de ZF*, pero, informalmente, ahora $\bigwedge X$, $\bigvee X$ ya no significa "para todo conjunto X" o "existe un conjunto X", sino "para toda clase X" y "existe una clase X", mientras que "conjunto" pasa a ser un término más técnico aún que en Z* o ZF*, pues ahora queda formalmente establecido (o, quedará, en cuanto demostremos que existen clases que no son conjuntos) que no toda colección de objetos es un conjunto.

Así, el axioma de extensionalidad afirma que si tenemos dos clases tales que todo conjunto que pertenezca a una pertenece también a la otra (y viceversa) entonces las dos clases son iguales.[13]

El axioma de comprensión afirma que existe una clase cuyos elementos son exactamente los conjuntos (no las clases) u que cumplen la propiedad $\phi(u)$. Dicha clase es única por el axioma de extensionalidad, luego podemos definir[14]

$$\{u \mid \phi(u)\} \equiv X|\bigwedge u(u \in X \leftrightarrow \phi(u)).$$

[12] Es fácil constatar que todos los conceptos conjuntistas usuales se expresan mediante términos y fórmulas normales (un término t es normal si lo es la fórmula $y = t$, para una variable y que no esté en t). Se trata simplemente de ir comprobando que ninguno requiere cuantificadores no acotados $\bigwedge X$ o $\bigvee X$. No haremos hincapié en ello.

[13] Notemos que en el axioma de extensionalidad (como en la definición de inclusión, un poco más abajo, y en otros contextos similares) da igual poner $\bigwedge u$ que $\bigwedge U$, pues las condiciones $U \in X$ y $U \in Y$ ya implican que U tiene que ser un conjunto, aunque no se exija.

[14] Insistimos en que la u minúscula indica que u es un conjunto, de modo que para pertenecer a esta clase no basta con cumplir $\phi(u)$ sino que además hay que ser un conjunto.

En NBG* esta descripción es propia cuando ϕ es una fórmula normal, mientras que en MK* siempre es una descripción propia.[15]

En particular de aquí deducimos la existencia de

$$\varnothing \equiv \{u \mid u \neq u\}, \quad V \equiv \{u \mid u = u\}, \quad R \equiv \{u \mid u \notin u\},$$

$$X \cap Y \equiv \{u \mid u \in X \wedge u \in Y\}, \quad X \cup Y \equiv \{u \mid u \in X \vee u \in Y\},$$

$$\bar{X} \equiv \{u \mid u \notin X\}, \quad X \setminus Y \equiv \{u \mid u \in X \wedge u \notin Y\}.$$

Así pues, tenemos definida la clase vacía, la clase universal, la intersección, unión y complemento de clases y es fácil demostrar sus propiedades básicas. Naturalmente, contamos también con el concepto de inclusión de clases:

$$X \subset Y \equiv \bigwedge u(u \in X \to u \in Y).$$

Otra cuestión es si esas clases son o no conjuntos. Por ejemplo, el axioma del conjunto vacío afirma que la clase $\varnothing$ es un conjunto. Por otra parte es inmediato que la clase de Russell R no es un conjunto, ya que si lo fuera, llegaríamos a la contradicción

$$R \in R \leftrightarrow R \notin R.$$

Esto prueba en NBG* que $\neg \operatorname{cto} R$, y no hay contradicción alguna, pues se cumple $R \notin R$ y esto no implica $R \in R$, porque el requisito para pertenecer a R es ser un conjunto y no pertenecerse a sí mismo, y falla la primera parte.

Por el axioma de comprensión tenemos que

$$\{X, Y\} \equiv Z | \bigwedge u(u \in Z \leftrightarrow u = X \vee u = Y)$$

es una descripción propia cuando X e Y son conjuntos, y por el axioma del par tenemos además que

$$\bigwedge xy \ \operatorname{cto}\{x, y\}.$$

Teniendo en cuenta la definición de par ordenado, ahora es inmediato que

$$\bigwedge xy \ \operatorname{cto}(x, y),$$

y es fácil probar el teorema fundamental sobre pares ordenados:

$$\bigwedge xyzw((x, y) = (z, w) \leftrightarrow x = z \wedge y = w).$$

El axioma de comprensión nos da también el producto cartesiano de clases:

$$X \times Y \equiv \{u \mid \bigvee vw(v \in X \wedge w \in Y \wedge u = (v, w))\}$$

A partir de aquí podemos hablar en NBG* de todos los conceptos conjuntistas relacionados con relaciones, funciones, dominios, rangos, etc. exactamente

[15] En cambio, en ZF* sólo podemos asegurar a priori que son descripciones propias las de la forma $\{u \in x \mid \phi(u)\}$, donde hay que dar un conjunto x dentro del cual especificamos los elementos que cumplen $\phi(u)$.

igual que en la teoría básica B, sólo que en B las clases son conceptos metamatemáticos, mientras que en NBG* son objetos de la teoría. En particular tenemos definido el concepto de aplicación $F : A \longrightarrow B$, en términos del cual el axioma de reemplazo puede reformularse diciendo que la imagen de un conjunto por una aplicación es un conjunto (como afirma 10.11):

Teorema 10.13 $\bigwedge FAB(F : A \longrightarrow B$ *suprayectiva* $\wedge \operatorname{cto} A \to \operatorname{cto} B)$.

Demostración: Si $F : A \longrightarrow B$, entonces $\operatorname{Un} F$, es decir, F está en las hipótesis del axioma de reemplazo, y lo que éste nos da es que

$$\bigvee y \bigwedge v(v \in y \leftrightarrow \bigvee u \in A\, F(u) = v),$$

pero por la suprayectividad esto es lo mismo que $\bigvee y \bigwedge v(v \in y \leftrightarrow v \in B)$, y por extensionalidad esto es lo mismo que afirmar que $B = y$ es un conjunto. ∎

De aquí se sigue a su vez que toda subclase de un conjunto es un conjunto

Teorema 10.14 $\bigwedge AB(A \subset B \wedge \operatorname{cto} B \to \operatorname{cto} A)$.

Demostración: O bien $A = \varnothing$, en cuyo caso ya sabemos que es un conjunto, o bien existe $a \in A$, con lo que podemos considerar la función $F : B \longrightarrow A$ suprayectiva dada por

$$F(u) = \begin{cases} u & \text{si } u \in A, \\ a & \text{si } u \notin A. \end{cases}$$

Como B es un conjunto, concluimos que A también es un conjunto por el teorema anterior. ∎

Observaciones Al contrario que en ZF*, en NBG* no podemos deducir el axioma del conjunto vacío del axioma de reemplazo. Podríamos haber evitado su uso en la prueba del teorema anterior observando que, trivialmente, la clase vacía cumple $\operatorname{Un}(\varnothing)$, y el conjunto b que se obtiene cuando se aplica el esquema de reemplazo al conjunto A y a $F = \varnothing$ es claramente $b = \varnothing$, luego $\varnothing$ es un conjunto. Sin embargo, es importante recalcar que con esto no hemos demostrado realmente $\operatorname{cto}\varnothing$, sino únicamente $\bigvee B \operatorname{cto} B \to \operatorname{cto}\varnothing$. Lo que se concluye de aquí es que el axioma del conjunto vacío se puede sustituir por un axioma más débil que afirme meramente la existencia de al menos un conjunto.

La razón por la que en ZF* no se requiere el axioma del conjunto vacío es que la existencia de un objeto es un teorema lógico, y en ZF* estamos hablando de la existencia de un conjunto. Eso basta para probar que existe el conjunto vacío partiendo del esquema de reemplazo o del de especificación. En cambio, en NBG* la existencia de un objeto significa únicamente que existe una clase, pero ninguno de los axiomas de NBG* (salvo el del conjunto vacío) implica la existencia de al menos un conjunto.

Explícitamente, el modelo M de $\mathcal{L}_{\mathrm{tc}}$ formado por un único objeto sin elementos es un modelo de NBG* menos el axioma del conjunto vacío. En ese modelo se cumple $V = \varnothing$, luego no puede probarse $V \neq \varnothing$ sin suponer el axioma del conjunto vacío o, equivalentemente, la existencia de un conjunto cualquiera.

Notemos también que el teorema 10.13 (que ya hemos visto que implica el teorema siguiente) implica a su vez el esquema de reemplazo, luego serviría como axioma equivalente. En efecto, si F es una clase unívoca y a es un conjunto, podemos considerar las clases

$$F' \equiv F \cap (A \times V), \qquad A = a \cap \mathcal{D}F, \qquad B = \{v \mid \bigvee u \in a\,(u, v) \in F\},$$

y es fácil ver que $F' : A \longrightarrow B$ suprayectiva, y se cumple que A es un conjunto, pues $A \subset a$. Por lo tanto B es un conjunto y cumple lo requerido por el esquema de reemplazo. ∎

El teorema anterior implica en particular que la intersección de conjuntos es un conjunto, mientras que para la unión necesitamos recurrir al axioma de la unión, que afirma que si x es un conjunto, entonces también lo es la clase

$$\bigcup x \equiv \{u \mid \bigvee v(u \in v \wedge v \in x)\}.$$

En particular, si x e y son conjuntos, también lo es $x \cup y = \bigcup\{x, y\}$.

Más en general, usaremos la notación

$$\bigcup_{u \in x} t(u) \equiv \bigcup\{v \mid \bigvee u \in x\, v = t(u)\},$$

y si el término $t(u)$ es normal y además $\bigwedge u \in x$ cto $t(u)$ (de modo que la clase $A = \{v \mid \bigvee u \in x\, v = t(u)\}$ está bien definida, y contiene ciertamente a todos los conjuntos $t(u)$), entonces la unión es un conjunto, pues la clase A lo es (por reemplazo aplicado a la función $F : x \longrightarrow A$ dada por $F = \{(u, t(u)) \mid u \in x\}$).

Otra consecuencia de 10.14 es que la clase universal V no es un conjunto, pues si lo fuera también lo sería la clase de Russel R.

Nota Observemos que en NBG* (o en cualquiera de sus posibles extensiones) es inmediato que la noción de "clase" dista mucho de capturar la noción metamatemática mal definida de "colección de objetos". Las clases formalizan algunas colecciones de objetos, pero no toda colección de objetos es una clase. Por ejemplo, fijado cualquier modelo de NBG*, las clases V, $V \setminus \{\varnothing\}$, $V \setminus \{\varnothing, \{\varnothing\}\}$ forman una colección de tres objetos que no son elementos de ninguna clase de dicho modelo. ∎

Teorema 10.15 $\bigwedge xy$ cto$(x \times y)$.

Demostración: En primer lugar observamos que si x e y son conjuntos y $v \in y$, entonces la clase $x \times \{v\}$ es un conjunto, por 10.13 aplicado a la función $F : x \longrightarrow x \times \{v\}$ dada por $F(u) = (u, v)$. En segundo lugar,

$$x \times y = \bigcup_{v \in y} x \times \{v\}$$

es un conjunto por la observación precedente sobre uniones. ∎

Nota Resulta interesante comparar esta demostración con su análoga en ZF*. En NBG* tenemos la existencia de la clase $x \times y$ y demostramos que ese objeto cuya existencia ya tenemos garantizada (por el esquema de comprensión) es un conjunto. En cambio, en ZF* podemos hablar de $x \times y$ como una fórmula, lo cual significa únicamente que podemos referirnos mediante una fórmula a los pares ordenados cuya primera componente está en x y su segunda componente está en y, pero tenemos que demostrar que existe un objeto (un conjunto) cuya extensión está formada precisamente por dichos pares. ∎

Otra aplicación del axioma de reemplazo es la siguiente:

Teorema 10.16 $\bigwedge FAB(F : A \longrightarrow B \to (\mathrm{cto}\, F \leftrightarrow \mathrm{cto}\, A))$.

Demostración: Si $\mathrm{cto}\, F$, basta tener en cuenta que $A \subset \bigcup\bigcup F$, mientras que si $\mathrm{cto}\, A$, entonces $\mathrm{cto}\, F[A]$ por reemplazo y $F \subset A \times F[A]$. ∎

La equivalencia entre ZF* y NBG* Ahora vamos a demostrar el teorema análogo a 10.2. Para ello necesitamos probar que ZF* se puede interpretar en NBG* en un sentido análogo a como AP se interpreta en ACA_0. En este caso, ZF* y NBG* tienen el mismo lenguaje formal $\mathcal{L}_{\mathrm{tc}}$, pero ahora, para cada expresión θ de $\mathcal{L}_{\mathrm{tc}}$ (pensado como el lenguaje de ZF*) definimos su *relativización a* V, que representaremos por θ^V, dada por las condiciones siguientes:

1. $X_i^V \equiv X_i$.
2. $(t_1 = t_2)^V \equiv t_1^V = t_2^V$.
3. $(t_1 \in t_2)^V \equiv t_1^V \in t_2^V$.
4. $(\neg\alpha)^V \equiv \neg\alpha^V$.
5. $(\alpha \to \beta)^V \equiv \alpha^V \to \beta^V$.
6. $(\bigwedge X_i\alpha)^V \equiv \bigwedge X_i(\mathrm{cto}\, X_i \to \alpha^V) \equiv \bigwedge x_i\, \alpha^V$.
7. $(X_i|\alpha)^V \equiv X_i|(\mathrm{cto}\, X_i \wedge \alpha^V) \equiv x_i|\alpha^V$.

Aquí hemos usado letras mayúsculas para nombrar las variables de $\mathcal{L}_{\mathrm{tc}}$, y letras minúsculas para representar las variables restringidas a conjuntos, pero si adoptamos el convenio de representar con letras minúsculas las variables de $\mathcal{L}_{\mathrm{tc}}$ cuando las usamos para representar fórmulas de ZF* y con mayúsculas cuando las usamos para representar fórmulas de NBG*, entonces la relativización de una expresión θ se expresa exactamente igual, entendiendo que, en la traducción, las minúsculas indican restricción a conjuntos. Por ejemplo, el axioma de extensionalidad de ZF* es

$$\bigwedge xy(\bigwedge u(u \in x \leftrightarrow u \in y) \leftrightarrow x = y),$$

donde las letras minúsculas representan variables arbitrarias (distintas entre sí), y su relativización es esta misma fórmula, pero entendiendo ahora que las letras

minúsculas representan variables restringidas a conjuntos, y entonces no es el axioma de extensionalidad de NBG*, que sería

$$\bigwedge XY(\bigwedge u(u \in x \leftrightarrow u \in y) \leftrightarrow X = Y),$$

pero es un caso particular.

Es claro que las fórmulas primitivas son precisamente las traducciones de fórmulas sin descriptores.

Observemos ahora que las relativizaciones de todos los axiomas de ZF* son teoremas de NBG*. El único caso que no es inmediato es el del reemplazo. Para probarlo sólo hemos de probar en NBG* la fórmula

$$\bigwedge x \overset{1}{\bigvee} y\, \phi^V(x,y) \to \bigwedge a \bigvee b \bigwedge y(y \in b \leftrightarrow \bigvee x \in a\, \phi^V(x,y)),$$

donde podemos suponer que ϕ no tiene descriptores. Basta definir

$$F \equiv \{(x,y) \mid \phi^V(x,y)\},$$

lo cual es correcto porque ϕ^V es primitiva, y entonces, si $\bigwedge x \overset{1}{\bigvee} y\, \phi^V(x,y)$, tenemos que F es unívoca (de hecho, $F : V \longrightarrow V$), luego por el axioma de reemplazo de NBG* tenemos que, dado un conjunto a,

$$\bigvee b \bigwedge v(v \in b \leftrightarrow \bigvee u \in a\, (u,v) \in F),$$

y esto es lo mismo que $\bigwedge v(v \in b \leftrightarrow \bigvee u \in a\, \phi^V(u,v))$, como había que probar.

Esto significa que NBG* interpreta a ZF* en un sentido análogo al definido en 3.28, pero para teorías sobre $\mathcal{L}_{\mathrm{tc}}$ en lugar de sobre $\mathcal{L}_a$. La única diferencia es que en las traducciones en el sentido de 3.28 acotábamos las variables con una fórmula $x \in \mathbb{N}$ y aquí las acotamos con $\mathrm{cto}\, x$. Esta diferencia no afecta en nada a la prueba del teorema 3.30, que vale con los mínimos cambios obvios en este contexto,[16] y nos da que (la relativización de) todo teorema de ZF* es demostrable en NBG*. En particular NBG* es una teoría aritmética en la que podemos definir los números naturales como ordinales.

Teorema 10.17 *Una sentencia α de $\mathcal{L}_{\mathrm{tc}}$ es un teorema de* ZF* *si y sólo si su relativización α^V es demostrable en* NBG*.

Demostración: Acabamos de probar una implicación. Fijemos un modelo[17] (que podemos suponer numerable) $M \vDash \mathrm{ZF}^*$. Como en 10.2, consideramos una enumeración $\phi_0, \phi_1, \ldots$ de las fórmulas de $\mathcal{L}_{\mathrm{tc}}$ que tienen a x_0 como variable libre y consideramos todos los pares $(\phi_i, \mathbf{a})$, donde $\mathbf{a}$ es una sucesión de elementos de M cuya longitud es igual al número de variables libres de ϕ_i distintas de x_0.

[16] En [LF 5.19] damos una definición de interpretación más general que incluye a 3.28 y al caso que consideramos ahora como casos particulars, y el teorema [LF 5.23] generaliza a 3.30 y es aplicable en este contexto.

[17] En [CS 9.27] damos una demostración constructiva de este mismo resultado, sin usar modelos.

Definimos la relación

$$(\phi_i, \mathbf{a}) \sim (\phi_j, \mathbf{b}) \quad \text{syss para todo } a \text{ de } M,\ M \vDash \phi_i[a, \mathbf{a}] \quad \text{syss} \quad M \vDash \phi_j[a, \mathbf{b}].$$

y consideramos las clases $[\phi_i, \mathbf{a}]$.

Observemos ahora que la condición "existe un c en M tal que para todo a de M se cumple

$$M \vDash \phi_i[a, \mathbf{a}] \quad \text{syss} \quad M \vDash [a] \in [c]"$$

se cumple para $(\phi_i, \mathbf{a})$ si y sólo si se cumple para cualquier otro $(\phi_j, \mathbf{b}) \sim (\phi_i, \mathbf{a})$. Por eso podemos considerar el conjunto C formado por todas las clases $[\phi_i, \mathbf{a}]$ que no cumplen esta condición, sin que importe el par $(\phi_i, \mathbf{a})$ considerado en la clase.

Definimos $\bar{M}$ como el conjunto formado por los elementos de M y los de C. Extendemos la relación de pertenencia definida en M del modo siguiente:

1. Si a, b están en M, entonces $\bar{M} \vDash [a] \in [b]$ si y sólo si $M \vDash [a] \in [b]$.
2. Si a está en M y $b \equiv [\phi_i, \mathbf{a}]$ está en C, entonces $\bar{M} \vDash [a] \in [b]$ syss $M \vDash \phi_i[a, \mathbf{a}]$.
3. En cualquier otro caso no se cumple $\bar{M} \vDash [a] \in [b]$.

Es claro entonces que, para todo a de $\bar{M}$, se cumple $\bar{M} \vDash \mathrm{cto}[a]$ syss a está en M, es decir, los elementos de $\bar{M}$ que pertenecen a otro elemento de $\bar{M}$ son exactamente los elementos de M. (Notemos que todo elemento de M pertenece a otro elemento de M, por ejemplo porque M satisface el axioma del par).

Como descripción impropia de $\bar{M}$ fijamos $M(\varnothing)$.

Una simple inducción prueba que, para todo término $t(x_1, \ldots, x_n)$ de $\mathcal{L}_{\mathrm{tc}}$ y todos los $a_1, \ldots, a_n$ en M, se cumple

$$M(t)[a_1, \ldots, a_n] \equiv \bar{M}(t^V)[a_1, \ldots, a_n],$$

y para toda fórmula $\alpha(x_1, \ldots, x_n)$ se cumple

$$M \vDash \alpha[a_1, \ldots, a_n] \quad \text{syss} \quad \bar{M} \vDash \alpha^V[a_1, \ldots, a_n].$$

En particular, si α es una sentencia tenemos que

$$M \vDash \alpha \quad \text{syss} \quad \bar{M} \vDash \alpha^V.$$

Veamos que $\bar{M}$ cumple el axioma de extensionalidad. Fijamos dos clases c y d de $\bar{M}$. Si ambas están en M tenemos que

$$M \vDash (\bigwedge u(u \in [c] \leftrightarrow u \in [d]) \to [c] = [d])$$

porque M verifica el axioma de extensionalidad, luego lo mismo vale en $\bar{M}$, por la observación precedente.

Si $c \equiv [\phi_i, \mathbf{a}]$ y $d \equiv [\phi_j, \mathbf{b}]$ y suponemos que

$$\bar{M} \vDash \bigwedge u(u \in [c] \leftrightarrow u \in [d]),$$

esto equivale a que para todo a en M se cumple

$$M \vDash \phi_i[a, \mathbf{a}] \quad \text{syss} \quad M \vDash \phi_j[a, \mathbf{b}],$$

luego $c \equiv [\phi_i, \mathbf{a}] \equiv [\phi_j, \mathbf{b}] \equiv d$, luego $\bar{M} \vDash [c] = [d]$.

Por último, si c esta en M y $d \equiv [\phi_i, \mathbf{a}]$, entonces no puede suceder que

$$\bar{M} \vDash \bigwedge u(u \in [c] \leftrightarrow u \in [d]),$$

pues esto significaría que para todo a en M se cumple $M \vDash [a] \in [c]$ syss $M \vDash \phi_i[a, \mathbf{a}]$, pero hemos excluido de $\bar{M}$ las clases $[\phi_i, \mathbf{a}]$ con esta propiedad (es decir, las clases cuya extensión sería la misma que la de un conjunto).

Concluimos que, en cualquier caso, $\bar{M}$ cumple el axioma de extensionalidad.

Veamos ahora que $\bar{M}$ cumple el esquema de comprensión (restringido a fórmulas primitivas). Toda fórmula primitiva es de la forma $\phi_i^V(x_0)$, para cierta fórmula $\phi_i(x_0)$, tal vez con más variables libres (pero no perdemos generalidad si suponemos que la variable de ϕ_i que aparece ligada en el axioma de reemplazo es precisamente la variable x_0 que hemos usado para definir los objetos de $\bar{M}$). Tenemos que encontrar un c en $\bar{M}$ tal que, para todo a en M, se cumpla

$$\bar{M} \vDash ([a] \in [c] \leftrightarrow \phi_i^V[a, \mathbf{a}]),$$

pero esto equivale a que $\bar{M} \vDash [a] \in [c]$ syss $M \vDash \phi_i[a, \mathbf{a}]$. Hay dos posibilidades: si existe un c en M en estas condiciones, entonces dicho c también está en M y cumple lo pedido. Si no existe tal c, entonces $[\phi_i, \mathbf{a}]$ está en M y cumple lo pedido.

Los axiomas del par, vacío y unión se cumplen en $\bar{M}$ porque son relativizaciones de axiomas (o de un teorema, en el caso del conjunto vacío) de ZF*, que son, por tanto, verdaderos en M. Sólo falta probar el axioma de reemplazo.

Tomamos un a en M y un d en $\bar{M}$ tal que $\bar{M} \vDash \mathrm{Un}([d])$, es decir,

$$\bar{M} \vDash \bigwedge xyz((x, y) \in [d] \wedge (x, z) \in [d] \to y = z).$$

Supongamos en primer lugar que $d \equiv [\phi_i, \mathbf{a}]$ y sea $\psi \equiv \bigvee x_0(x_0 = (x, y) \wedge \phi_i)$. Es fácil ver que

$$M \vDash \bigwedge xyz(\psi(x, y) \wedge \psi(x, z) \to y = z)[\mathbf{a}].$$

Por el esquema de reemplazo, que es verdadero en M, existe un b en M tal que

$$M \vDash \bigwedge y(y \in [b] \leftrightarrow \bigvee x \in [a]\, \psi(x, y))[\mathbf{a}],$$

pero esto equivale a

$$\bar{M} \vDash \bigwedge y(y \in [b] \leftrightarrow \bigvee x \in [a]\, (x, y) \in [d]),$$

como había que probar.

Si d también está en M basta considerar la fórmula $\psi(x,y,t) \equiv (x,y) \in t$, de modo que

$$M \vDash \bigwedge xyz(\psi(x,y,[d]) \wedge \psi(x,z,[d]) \to y = z),$$

y se concluye análogamente.

Finalmente concluimos como en 10.2: Si la relativización de una sentencia ϕ de $\mathcal{L}_{\mathrm{tc}}$ no es demostrable demostrable en ZF^*, entonces el teorema de completitud de Gödel (teorema 4.17) nos da que existe un modelo (que podemos suponer numerable) M de ZF^* tal que $M \vDash \neg\phi$, pero hemos visto que entonces $\bar{M} \vDash \neg\phi^V$ y, como $\bar{M}$ es un modelo de NBG^*, concluimos que la relativización ϕ^V no es un teorema de NBG^*. ∎

Nota Es fácil ver que el teorema anterior sigue siendo cierto si sustituimos ZF^* por Z^* y en NBG^* sustituimos el axioma de reemplazo por el teorema 10.14. ∎

En particular tenemos que NBG^* es consistente.[18]

La axiomatización finita de NBG* Hemos visto que el esquema de inducción de AP se sustituye por un único axioma en ACA_0, si bien en esta teoría tenemos el esquema de comprensión, pero luego hemos probado que dicho axioma puede sustituirse por un número finito de axiomas.

Igualmente, en ZF^* tenemos el esquema de reemplazo, que en NBG^* se sustituye por un único axioma, pero a cambio introducimos el esquema axiomático de comprensión. Ahora vamos a probar el teorema análogo a 10.3, que nos da que este esquema puede sustituirse por un número finito de casos particulares.

Teorema 10.18 NBG^* *es finitamente axiomatizable.*

Demostración: Consideramos la teoría cuyos axiomas son los de NBG^* menos el esquema de comprensión y además las sentencias siguientes:

Intersección	$\bigwedge XY\bigvee Z\bigwedge u(u \in Z \leftrightarrow u \in X \wedge u \in Y)$
Complemento	$\bigwedge X\bigvee Y\bigwedge u(u \in Y \leftrightarrow u \notin X)$
Pertenencia	$\bigvee A\bigwedge xy((x,y) \in A \leftrightarrow x \in y)$
Dominio	$\bigwedge A\bigvee B\bigwedge x(x \in B \leftrightarrow \bigvee y\,(x,y) \in A)$
Producto cartesiano	$\bigwedge A\bigvee B\bigwedge xy((x,y) \in B \leftrightarrow x \in A)$
Relación inversa	$\bigwedge A\bigvee B\bigwedge xy((x,y) \in B \leftrightarrow (y,x) \in A)$
Identidad	$\bigvee A\bigwedge xy((x,y) \in A \leftrightarrow x = y)$
Permutación 1	$\bigwedge A\bigvee B\bigwedge xyz((x,y,z) \in B \leftrightarrow (y,z,x) \in A)$
Permutación 2	$\bigwedge A\bigvee B\bigwedge xyz((x,y,z) \in B \leftrightarrow (x,z,y) \in A)$

[18] Como en el caso de 10.2, la prueba del teorema anterior no es constructiva, pero es posible dar una prueba constructiva basada en técnicas de la teoría de la demostración [CS A.6]. No obstante la prueba que hemos dado con modelos es la que permite entender realmente la situación.

En los axiomas de permutación hay que entender que las ternas ordenadas (y, más en general, las n-tuplas ordenadas) se definen recurrentemente mediante

$$(x) \equiv x, \qquad (x_1, \ldots, x_n) \equiv ((x_1, \ldots, x_{n-1}), x_n).$$

Concretamente, $(x, y, z) \equiv ((x, y), z)$.

Es claro que todas estas sentencias son casos particulares del esquema de comprensión, luego todos los teoremas de esta teoría son teoremas de NBG*. Sólo tenemos que probar el recíproco. Más concretamente, sólo tenemos que demostrar todos los casos particulares del esquema de comprensión restringido a fórmulas primitivas.

Es fácil ver que todas las propiedades demostradas en la prueba de 10.3 hasta las propiedades numeradas de 1) a 6) tienen su análogo exacto en nuestro contexto con la misma prueba, sin más que cambiar las n-tuplas ordenadas aritméticas definidas allí por las conjuntistas que acabamos de definir. Basta probar que, para toda fórmula primitiva $\phi(x_1, \ldots, x_n)$, tal vez con más variables libres,

$$\underset{\mathrm{NBG}^*}{\vdash} \bigvee A \bigwedge x_1 \cdots x_n((x_1, \ldots, x_n) \in A \leftrightarrow \phi(x_1, \ldots, x_n)),$$

pues haciendo $n = 1$ tenemos el esquema de comprensión.

Sean $Y_1, \ldots, Y_m$ las variables libres distintas de $x_1, \ldots, x_n$ que haya en ϕ. Podemos suponer que ϕ tiene libre alguna de las variables x_i, pues de lo contrario sirve $A = \mathrm{V}$ o $A = \varnothing$ (la segunda existe por el axioma del conjunto vacío y la primera por el axioma del complemento aplicado a $\varnothing$). Podemos suponer también que las variables Y_i no aparecen nunca como primer término de un relator de pertenencia, pues $Y_i \in t$ puede sustituirse por $\bigvee x(x = Y_i \wedge x = t)$ y a su vez esto se sustituye por la fórmula

$$\neg \bigwedge x \neg(\bigwedge u(u \in x \leftrightarrow u \in Y_i) \wedge x \in t),$$

que sigue siendo primitiva. Similarmente podemos exigir que no haya subfórmulas de tipo $X \in X$. Sea, pues, ϕ una fórmula primitiva en estas condiciones. Demostraremos el resultado por inducción sobre la longitud de ϕ.

1. Si $\phi \equiv x_r \in x_s$ entonces $r < s$ o bien $s < r$. Según el caso obtenemos

   $$\bigvee F \bigwedge x_r x_s((x_r, x_s) \in F \leftrightarrow x_r \in x_s) \qquad \text{(por el axioma de pertenencia)}$$
   $$\bigvee F \bigwedge x_r x_s((x_s, x_r) \in F \leftrightarrow x_r \in x_s) \qquad \text{(por pertenencia y rel. inversa)}$$

   En otros términos, si llamamos p al menor de r, s y q al mayor de ambos, se cumple

   $$\bigvee F \bigwedge x_p x_q((x_p, x_q) \in F \leftrightarrow \phi(x_1, \ldots, x_n)).$$

   Si $p \neq 1$ aplicamos[19] 6) para concluir

   $$\bigvee F_1 \bigwedge x_1 \cdots x_p\, x_q((x_1, \ldots, x_p, x_q) \in F_1 \leftrightarrow \phi(x_1, \ldots, x_n)).$$

[19] Esta referencia y las que siguen remiten a los teoremas del apartado 3 de esta misma sección.

Si $q \neq p+1$ aplicamos 5) para concluir

$$\bigvee F_2 \bigwedge x_1 \cdots x_q((x_1, \ldots, x_q) \in F_2 \leftrightarrow \phi(x_1, \ldots, x_n)).$$

Si $q \neq n$ aplicamos 3) y obtenemos

$$\bigvee A \bigwedge x_1 \cdots x_n((x_1, \ldots, x_n) \in A \leftrightarrow \phi(x_1, \ldots, x_n)).$$

2. Si $\phi \equiv x_r \in Y_k$ distinguimos el caso $n = r = 1$, que es trivial, pues $\bigwedge x_1((x_1) \in Y_k \leftrightarrow x_1 \in Y_k)$, luego $\bigvee A \bigwedge x_1((x_1) \in A \leftrightarrow x_1 \in Y_k)$.

   Si $n \neq 1$, o bien $r \neq n$, y entonces por el axioma del producto cartesiano

   $$\bigvee A \bigwedge x_r x_{r+1}((x_r, x_{r+1}) \in A \leftrightarrow x_r \in Y_k),$$

   o bien $r = n$, y entonces aplicamos 1):

   $$\bigvee A \bigwedge x_{r-1} x_r((x_{r-1}, x_r) \in A \leftrightarrow x_r \in Y_k).$$

   En ambos casos, aplicando 3) y 6) llegamos a

   $$\bigvee A \bigwedge x_1 \cdots x_n((x_1, \ldots, x_n) \in A \leftrightarrow \phi(x_1, \ldots, x_n)).$$

3. Si $\phi \equiv \neg\psi$, $\phi \equiv \chi \to \psi$, o bien $\phi \equiv \bigwedge x\, \theta$, en el último caso podemos suponer que $x \not\equiv x_i$ para todo i. Por hipótesis de inducción

   $$\bigvee B \bigwedge x_1 \cdots x_n((x_1, \ldots, x_n) \in B \leftrightarrow \psi(x_1, \ldots, x_n)),$$
   $$\bigvee C \bigwedge x_1 \cdots x_n((x_1, \ldots, x_n) \in C \leftrightarrow \chi(x_1, \ldots, x_n)),$$
   $$\bigvee D \bigwedge x_1 \cdots x_n\, x((x_1, \ldots, x_n, x) \in D \leftrightarrow \theta(x_1, \ldots, x_n, x)).$$

   Si $\phi \equiv \neg\psi$ tenemos $\bigvee A \bigwedge x_1 \cdots x_n((x_1, \ldots, x_n) \in A \leftrightarrow \phi(x_1, \ldots, x_n))$ sin más que tomar $A = V \setminus B$. Si $\phi \equiv \chi \to \psi$ llegamos a la misma conclusión tomando $A = (V \setminus C) \cup B$ y si $\phi \equiv \bigwedge x\, \theta$ damos varios pasos:

   $$\bigvee E \bigwedge x_1 \cdots x_n((x_1, \ldots, x_n) \in E \leftrightarrow \bigvee x\, (x_1, \ldots, x_n, x) \in V \setminus D)$$

   $$\bigvee A \bigwedge x_1 \cdots x_n((x_1, \ldots, x_n) \in A \leftrightarrow \neg \bigvee x\, (x_1, \ldots, x_n, x) \in V \setminus D).$$

   La primera fórmula se cumple con $E = \mathcal{D}(V \setminus D)$, y la segunda con $A = V \setminus E$.

   Así pues, $\bigvee A \bigwedge x_1 \cdots x_n((x_1, \ldots, x_n) \in A \leftrightarrow \bigwedge x\, \theta(x_1, \ldots, x_n, x))$, como queríamos probar. ∎

La formalización de la lógica en NBG* Todos los conceptos relacionados con lenguajes formales y teorías axiomáticas pueden ser formalizados en NBG* análogamente a como hemos visto en ACA_0 (teniendo en cuenta que los conjuntos de ACA_0 se corresponden con las clases de NBG*). Así, un lenguaje formal puede definirse como una nónupla ordenada de clases[20]

$$\mathcal{L} = (\{\ulcorner\neg\urcorner\}, \{\ulcorner\to\urcorner\}, \{\ulcorner\textstyle\bigwedge\urcorner\}, \{\ulcorner|\urcorner\}, \{\ulcorner=\urcorner\}, V, R, F, \mathrm{Nar}),$$

donde las n-tuplas ordenadas de clases se definen como en ACA_0:

$$(X_1, \ldots, X_n) \equiv X|\textstyle\bigwedge u(u \in X \leftrightarrow$$

$$\textstyle\bigvee y((y \in X_1 \wedge u = (0^{(1)}, y)) \vee \cdots \vee (y \in X_n \wedge u = (0^{(n)}, y))).$$

Como ahora no estamos imponiendo ninguna restricción sobre la complejidad de las expresiones empleadas, todos los conceptos lógicos metamatemáticos se formalizan de forma natural, sin dificultad alguna. Solamente nos encontramos un obstáculo insalvable en NBG* si tratamos de definir una fórmula $V \vDash \alpha[v]$. El análogo del teorema 10.4 es como sigue:

$$D \equiv \{\langle\theta, v\rangle \mid \theta \in \mathrm{Exp}(\ulcorner\mathcal{L}_a\urcorner) \wedge \mathrm{Val}(v, \theta)\}.$$

Teorema 10.19 (MK*) *Si* $D \equiv \{(\theta, v) \mid \theta \in \mathrm{Exp}(\ulcorner\mathcal{L}_{\mathrm{tc}}\urcorner) \wedge \mathrm{Val}(v, \theta)\}$, *existe una única función* $F : D \longrightarrow V$ *que cumple las propiedades siguientes:*

1. $F(x, v) = v(x)$,
2. $F(t_1 = t_2, v) = \begin{cases} 1 & \text{si } F(t_1, v) = F(t_2, v), \\ 0 & \text{en otro caso,} \end{cases}$
3. $F(t_1 \in t_2, v) = \begin{cases} 1 & \text{si } F(t_1, v) \in F(t_2, v), \\ 0 & \text{en otro caso,} \end{cases}$
4. $F(\neg\alpha, v) = 1 - F(\alpha, v)$,
5. $F(\alpha \to \beta, v) = \begin{cases} 1 & \text{si } F(\alpha, v) = 0 \vee F(\beta, v) = 1, \\ 0 & \text{en otro caso,} \end{cases}$
6. $F(\textstyle\bigwedge x\alpha, v) = \begin{cases} 1 & \text{si } \bigwedge u\ F(\alpha, v^u_x) = 1, \\ 0 & \text{en otro caso,} \end{cases}$
7. $F(x|\alpha, v) = \begin{cases} a & \text{si } \bigwedge u(F(\alpha, v^u_x) = 1 \leftrightarrow u = a), \\ \varnothing & \text{si no existe tal } a. \end{cases}$

[20]Notemos que, en ausencia del axioma de infinitud, no podemos exigir que las clases V y Nar sean conjuntos, pues deben contener infinitos elementos.

La demostración es totalmente análoga[21] a la de 10.4. A partir de aquí podemos definir el término $V(t)[v] \equiv F(t,v)$ y la fórmula $V \vDash \alpha[v] \equiv F(\alpha,v) = 1$ de modo que se cumple el teorema siguiente:

Teorema 10.20 (MK*) *Si v es una valoración definida sobre los términos o las fórmulas correspondientes a cada apartado, se cumple*

1. $V(x)[v] = v(x)$,
2. $V \vDash (t_1 = t_2)[v] \leftrightarrow V(t_1)[v] = V(t_2)[v]$,
3. $V \vDash (t_1 \in t_2)[v] \leftrightarrow V(t_1)[v] \in V(t_2)[v]$,
4. $V \vDash \neg\alpha[v] \leftrightarrow \neg V \vDash \alpha[v]$,
5. $V \vDash (\alpha \to \beta)[v] \leftrightarrow (\neg V \vDash \alpha[v] \vee V \vDash \beta[v])$,
6. $V \vDash \bigwedge x\alpha[v] \leftrightarrow \bigwedge u\ V \vDash \alpha[v^u_x]$,
7. $V(x|\alpha)[v] = \begin{cases} a & \text{si } a \text{ es el único que cumple } V \vDash \alpha[v^a_x], \\ \varnothing & \text{si no existe un único } a \text{ en estas condiciones.} \end{cases}$

La formalización del teorema de corrección no presenta ninguna dificultad (lo único que impide demostrarlo en NBG* es que no podemos definir la fórmula $V \vDash \alpha[v]$), e igualmente se formaliza la prueba de que todos los axiomas de ZF* son verdaderos en V, lo que nos permite concluir:

Teorema 10.21 $\underset{\text{MK}^*}{\vdash} \text{Consis}\,\ulcorner\text{ZF}^*\urcorner$.

Tenemos así un ejemplo concreto de teorema de MK* (que es equivalente a una afirmación sobre números naturales, e incluso a una sentencia que afirma simplemente que una cierta ecuación diofántica no tiene solución) que no es demostrable en NBG*, pues si fuera demostrable en NBG* también lo sería en ZF*, por el teorema 10.17, y por el teorema de incompletitud ZF* sería contradictoria.

Más aún, una simple inducción metamatemática demuestra que, para todo término $t(x_1,\ldots,x_n)$ y toda fórmula $\alpha(x_1,\ldots,x_n)$ de $\mathcal{L}_{\text{tc}}$, si llamamos

$$v \equiv \{(\ulcorner x_1\urcorner, x_1), \ldots, (\ulcorner x_n\urcorner, x_n)\},$$

entonces

$$\underset{\text{MK}^*}{\vdash} \bigwedge x_1 \cdots x_n (V(\ulcorner t\urcorner)[v] = t^V(x_1,\ldots,x_n)),$$

$$\underset{\text{MK}^*}{\vdash} \bigwedge x_1 \cdots x_n (V \vDash \ulcorner\alpha\urcorner[v] \leftrightarrow \alpha^V(x_1,\ldots,x_n)).$$

En particular, para toda sentencia α de $\mathcal{L}_{\text{tc}}$ se cumple que

$$\underset{\text{MK}^*}{\vdash} V \vDash \ulcorner\alpha\urcorner \leftrightarrow \alpha^V.$$

[21] Notemos que a la hora de definir $\text{Exp}(\ulcorner\mathcal{L}_{\text{tc}}\urcorner)$ podemos elegir si tomamos como expresiones números naturales o sucesiones finitas de números naturales. En el segundo caso, la definición de las clases D_n consideradas en la prueba debe cambiarse a

$$D_n \equiv \{(\theta,v) \mid (\theta,v) \in D \wedge \ell(\theta) < n\}.$$

Extensiones de ZF*, NBG* y MK* La teoría de conjuntos que utilizan habitualmente los matemáticos es ZFC, que resulta de añadir a ZF* un número finito de axiomas (que discutiremos en el capítulo siguiente). Todos los resultados que hemos presentado en esta sección se generalizan trivialmente a cualquier teoría T que resulte de añadir un número finito de axiomas a ZF* (lo que en la práctica equivale a añadir un único axioma γ, pues si queremos añadir varios podemos considerar la conjunción de todos ellos) y a las teorías T' y T'' que resultan de añadir a NBG* y MK* respectivamente el axioma γ^V.

En efecto, como NBG* es finitamente axiomatizable, es claro que T' también lo es. El teorema 10.17 se cumple porque, si α es una sentencia,

$$\underset{T}{\vdash} \alpha \quad \text{syss} \quad \underset{\text{ZF}^*}{\vdash} \gamma \to \alpha \quad \text{syss} \quad \underset{\text{NBG}^*}{\vdash} \gamma^V \to \alpha^V \quad \text{syss} \quad \underset{T'}{\vdash} \alpha^V.$$

Por otra parte, acabamos de probar que

$$\underset{\text{MK}^*}{\vdash} \alpha^V \to V \vDash \ulcorner \alpha \urcorner,$$

luego $\underset{T''}{\vdash} V \vDash \ulcorner T \urcorner$, luego $\underset{T''}{\vdash} \text{Consis}\, \ulcorner T \urcorner$. ∎

Capítulo XI

Los axiomas restantes de la teoría de conjuntos

En el capítulo anterior hemos presentado los axiomas básicos de la teoría de conjuntos en dos teorías equivalentes: ZF* y NBG*. La teoría axiomática con la que trabajan habitualmente los matemáticos cuenta con cuatro axiomas más: los axiomas de infinitud, regularidad, partes y elección. Ahora vamos a discutir cada uno de ellos para mostrar su impacto en la teoría. En este capítulo trabajamos indistintamente[1] en ZF* o NBG*, e indicaremos explícitamente cualquier uso de los axiomas adicionales que vamos a introducir. En cualquiera de las dos teorías podemos definir la clase Ω de todos los ordinales, que está bien ordenada por la inclusión (teorema 3.14) y la clase ω de todos los números naturales (pues las hemos definido, de hecho, en la teoría básica B) y podemos probar que la primera no es un conjunto. Enseguida veremos que la afirmación "ω es un conjunto" es una de las distintas formas equivalentes del axioma de infinitud.

11.1 El axioma de infinitud

Conviene observar que toda la teoría básica sobre conjuntos finitos que hemos desarrollado en la sección 6.4 en el marco de KP es válida literalmente en nuestro contexto actual (incluso con demostraciones más simples, pues ahora no tenemos que preocuparnos por determinar el nivel de complejidad de las expresiones que consideramos). Puesto que ahora no necesitamos mantener la ambigüedad sobre la definición de $\mathbb{N}$ que usábamos para trabajar simultáneamente en $\mathrm{I}\Sigma_1$ o en KP, sino que podemos considerar concretamente que los números naturales son los elementos de ω, la definición 6.18 se puede expresar ahora de este modo:

Definición 11.1 x es finito $\equiv \bigvee f \bigvee n \in \omega\ f : n \longrightarrow x$ biyectiva.

[1] Los teoremas que involucren clases propias se corresponden con esquemas teoremáticos en ZF*, en los que las clases tienen que interpretarse como fórmulas.

Como en AP pueden demostrarse todos los axiomas de ZF* + "todo conjunto es finito" (teoremas 5.56 y 6.22), podemos concluir que es imposible demostrar en ZF* (luego en NBG*) la existencia de un conjunto infinito. En particular, no podemos demostrar que la clase ω de los números naturales sea un conjunto, pues, si lo fuera, sería un conjunto infinito, por ejemplo por 6.28 2), pues la relación de orden usual en ω no tiene máximo. Más en general:

Teorema 11.2 *Los ordinales finitos son los números naturales.*

Demostración: Es inmediato, por la definición de finitud, que los números naturales son ordinales finitos. Si $\alpha \in \Omega$ es finito pero no es un número natural, entonces, para todo $n \in \omega$, no puede suceder que $\alpha \leq n$ (pues esto equivale a $\alpha \in n \vee \alpha = n$ y entonces α sería un número natural, porque ω es transitiva), luego $n \in \alpha$, luego $\omega \subset \alpha$, y esto implica que ω es un conjunto finito, pero acabamos de justificar que si ω es un conjunto, es un conjunto infinito. ∎

El axioma de infinitud afirma esencialmente que ω es un conjunto, pero en lugar de enunciarlo así consideraremos una sentencia equivalente sin descriptores:

Definición 11.3 Llamaremos *axioma de infinitud* a la sentencia de $\mathcal{L}_{\mathrm{tc}}$

$$\mathrm{AI} \equiv \bigvee x(\bigvee u \in x \bigwedge v\ v \notin u \wedge \bigwedge u \in x \bigvee v \in x \bigwedge w(w \in v \leftrightarrow w \in u \vee w = u)).$$

Si consideramos AI como axioma adicional de NBG* hay que entender que las letras minúsculas están relativizadas a conjuntos. Una versión con descriptores obviamente equivalente es

$$\bigvee x(\varnothing \in x \wedge \bigwedge u \in x\, u' \in x),$$

donde, recordemos, $u' \equiv u \cup \{u\}$ es la operación conjuntista que, sobre los ordinales, determina el ordinal siguiente. Veamos algunas equivalencias elementales:

Teorema 11.4 *Las sentencias siguientes son equivalentes:*

$$1)\quad \mathrm{AI}, \qquad 2)\quad \mathrm{cto}\,\omega, \qquad 3)\quad \omega \in \Omega, \qquad 4)\quad \omega \neq \Omega.$$

Demostración: Dado un conjunto x en las condiciones dadas por AI, podemos especificar el subconjunto

$$y \equiv \{n \in x \mid n \in \omega\} \subset x,$$

y entonces claramente

$$y \subset \omega \wedge 0 \in y \wedge \bigwedge u \in y\, u' \in \omega.$$

Por el teorema 3.23 concluimos que $y = \omega$, luego[2] $\mathrm{cto}\,\omega$.

[2]Más precisamente, tenemos $\bigvee y \bigwedge n(n \in y \leftrightarrow n \in \omega)$, y esto es precisamente lo que en ZF* se entiende por "la clase ω es un conjunto".

Por otra parte, si existe un conjunto ω cuyos elementos son exactamente los números naturales, es inmediato que cumple lo requerido por AI, luego tenemos que 1) $\leftrightarrow$ 2).

Trivialmente 2) $\leftrightarrow$ 3), pues Ω es la clase de todos los conjuntos que son ordinales. También es claro que 3) $\to$ 4), pues podemos probar $\neg\,\mathrm{cto}\,\Omega$. Por último, si $\omega \neq \Omega$, eso significa que existe un ordinal α que no es un número natural, pero una simple inducción demuestra entonces que $\omega \subset \alpha$, luego $\mathrm{cto}\,\omega$. ∎

Así pues, sin AI no podemos demostrar la existencia de más ordinales que los números naturales, mientras que bajo AI obtenemos ω, que es el menor ordinal infinito y también el menor ordinal límite en el sentido siguiente:

Definición 11.5 Un ordinal $\alpha \in \Omega$ es un *ordinal sucesor* si $\bigvee\beta < \alpha\ \alpha = \beta'$. Los ordinales distintos de 0 que no son sucesores se llaman *ordinales límite*. Por lo tanto, todo ordinal α se encuentra necesariamente en uno y sólo uno de los tres casos siguientes: $\alpha = 0$, α es un ordinal sucesor, α es un ordinal límite.

Sabemos que todo número natural es 0 o bien el sucesor de otro número natural, luego ningún número natural es un ordinal límite. Por otra parte, si ω es un conjunto, entonces es un ordinal límite, ya que no puede ser $\omega = n'$, pues entonces n sería un número natural y n' también, luego ω sería finito. Por lo tanto, la existencia de ordinales límite es otra forma equivalente del axioma de infinitud, pues equivale a que existan ordinales distintos de los números naturales.

Nota Usando el teorema 3.26, en Z*+ AI podemos definir $F : \omega \longrightarrow V$ mediante las condiciones $F(0) = \omega \wedge \bigwedge n \in \omega\, F(n') = F(n)'$. Observemos que F es una clase definida por una cierta fórmula. Si llamamos $\omega + n \equiv F(n)$, tenemos un término de $\mathcal{L}_{\mathrm{tc}}$ y una simple inducción demuestra que $\bigwedge n \in \omega\, \omega + n \in \Omega$. Explícitamente:

$$\omega + 1 = \{0, 1, 2, \ldots, \omega\}, \quad \omega + 2 = \{0, 1, 2, \ldots, \omega, \omega + 1\}, \ \ldots$$

De este modo, la sucesión de los números naturales se prolonga a la sucesión de ordinales

$$0 < 1 < 2 < \cdots < \omega < \omega + 1 < \omega + 2 < \cdots$$

Ahora bien, se necesita aplicar el esquema de reemplazo a la fórmula $y = \omega + n$ para concluir que existe el conjunto $x = \{\omega + n \mid n \in \omega\}$, a partir del cual podemos aplicar 3.20 3) para concluir que $\omega + \omega \equiv \bigcup x$ es un ordinal límite, y es fácil ver que es el menor ordinal límite mayor que ω.

Tenemos así un primer ejemplo no trivial del papel que representa el esquema de reemplazo en la teoría de conjuntos. Puede demostrarse que en Z*+AI (supuesto que sea una teoría consistente) no es posible demostrar que existan más ordinales aparte de los números naturales y los ordinales $\omega + n$, con $n \in \omega$. En particular, no es posible demostrar que existan más ordinales límite además

de ω. Por el contrario, en ZF^*+ AI, cambiando ω por un ordinal arbitrario en la construcción que hemos dado de $\omega+\omega$, se demuestra sin dificultad que por encima de todo ordinal α hay un ordinal límite $\alpha+\omega$, y así la sucesión de los ordinales se puede prolongar más aún:

$$0<1<\cdots\omega<\omega+1<\cdots\omega+\omega+\omega<\omega+\omega+\omega+1<\omega+\omega+\omega+2<\cdots$$

En la página 406 definimos en general la suma de ordinales. ∎

Hemos visto que lo que afirma AI es que existe un conjunto que contiene a ω, lo cual es tanto como afirmar que ω es un conjunto. Sin embargo, es natural preguntarse por qué tomamos como axioma de infinitud que ω es un conjunto en lugar de considerar simplemente la sentencia "existe un conjunto infinito". Para entender la situación conviene considerar una definición alternativa de conjuntos finitos e infinitos debida a Dedekind:

Definición 11.6 Un conjunto x es *D-infinito* si existe $f : x \longrightarrow x$ inyectiva no suprayectiva. En caso contrario se dice *D-finito.*

En estos términos, el teorema 6.30 afirma que todo conjunto finito es D-finito, luego todo conjunto D-infinito es infinito. Sin embargo, en ZF^* (o NBG^*) no es posible demostrar el recíproco. Y además sucede que AI no es equivalente a la existencia de un conjunto infinito:

Teorema 11.7 *El axioma* AI *equivale a la existencia de un conjunto D-infinito.*

Tenemos que AI implica que ω es un conjunto, y ciertamente es un conjunto D-infinito (basta considerar la función sucesor $s : \omega \longrightarrow \omega$, que es inyectiva y no suprayectiva). Probar el recíproco nos costará un poco más.

Definición 11.8 Diremos que una terna $(\mathbb{N}, s, 0)$ es un *sistema de números naturales* si $\mathbb{N}$ es un conjunto, $s : \mathbb{N} \longrightarrow \mathbb{N}$ y se cumple

1. $0 \in \mathbb{N}$,
2. $\bigwedge n \in \mathbb{N}\ s(n) \in \mathbb{N}$,
3. $\bigwedge n \in \mathbb{N}\ s(n) \neq 0$,
4. $\bigwedge mn \in \mathbb{N}\ (s(m) = s(n) \to m = n)$,
5. $\bigwedge y(y \subset \mathbb{N} \wedge 0 \in y \wedge \bigwedge n \in y\, s(n) \in y \to y = \mathbb{N})$.

Así, un sistema de números naturales viene a ser un modelo de los cinco axiomas de Peano originales.

Si x es un conjunto y $f : x \longrightarrow x$ es una aplicación inyectiva y no suprayectiva, podemos tomar $0 \in x \setminus f[x]$. Diremos que un conjunto es *inductivo* si cumple

$$\mathrm{Ind}(y) \equiv y \subset x \wedge 0 \in y \wedge \bigwedge u \in y\ f(u) \in y.$$

Definimos el conjunto

$$\mathbb{N} \equiv \{n \in x \mid \bigwedge y\,(\mathrm{Ind}(y) \to n \in y)\}.$$

En otras palabras, definimos $\mathbb{N}$ como el conjunto de todos los conjuntos que pertenecen a todos los conjuntos inductivos. Si llamamos $s = f|_{\mathbb{N}}$, resulta que $(\mathbb{N}, s, 0)$ es un sistema de números naturales.

En efecto, la propiedad 1) se cumple porque, por definición, 0 pertenece a todo conjunto inductivo. Similarmente, si $n \in \mathbb{N}$, entonces n está en todo conjunto inductivo, luego $s(n)$ también, por definición de conjunto inductivo, luego $s(n) \in \mathbb{N}$. Esto prueba 2). La propiedad 3) se cumple porque 0 no está en la imagen de f, luego tampoco en la de la restricción s, y 4) se cumple porque f es inyectiva, luego s también. La propiedad 5) se cumple porque sus hipótesis son que $y \subset \mathbb{N} \wedge \mathrm{Ind}(y)$, luego por definición de $\mathbb{N}$ tenemos que $\mathbb{N} \subset y$.

Hemos demostrado que si existe un conjunto D-infinito entonces existe un sistema de números naturales. El paso siguiente es demostrar un teorema de recursión:

Teorema 11.9 *Sea $(\mathbb{N}, s, 0)$ un sistema de números naturales, sea $G : X \longrightarrow X$ una aplicación cualquiera (no exigimos que X o G sean conjuntos) y sea $x \in X$. Entonces existe una única aplicación $f : \mathbb{N} \longrightarrow X$ de manera que $f(0) = x$ y $\bigwedge n \in \mathbb{N}\, f(s(n)) = G(f(n))$.*

Demostración: Diremos que h es una *aproximación* si $h : d \longrightarrow X$, donde $d \subset \mathbb{N}$ es un conjunto con la propiedad de que

$$0 \in d \wedge \bigwedge n \in \mathbb{N}(s(n) \in d \to n \in d)$$

y h cumple que $h(0) = x \wedge \bigwedge n \in \mathbb{N}(s(n) \in d \to h(s(n)) = G(h(n)))$.

Observamos ahora que si $h : d \longrightarrow X$ y $h' : d' \longrightarrow X$ son dos aproximaciones y $n \in d \cap d'$, entonces $h(n) = h'(n)$.

En efecto, basta ver que el conjunto

$$y = \{n \in \mathbb{N} \mid n \in d \cap d' \to h(n) = h'(n)\}$$

es inductivo. Obviamente $0 \in y$ y, si $n \in y$, entonces, si $s(n) \in d \cap d'$ se cumple que $n \in d \cap d'$, luego $h(n) = h'(n)$, luego

$$h(s(n)) = G(h(n)) = G(h'(n)) = h'(s(n)),$$

luego $s(n) \in y$. Por otra parte, también es inductivo el conjunto

$$y' = \{n \in \mathbb{N} \mid \bigvee dh(h : d \longrightarrow X \text{ es una aproximación} \wedge n \in d)\}.$$

En efecto, se cumple que $0 \in y'$ porque $h = \{(0, x)\}$ es una aproximación y, si $n \in y'$, entonces existe una aproximación $h : d \longrightarrow X$ tal que $n \in d$. Si se cumple $s(n) \in d$, entonces $s(n) \in y'$. En caso contrario, basta observar que

$$h' = h \cup \{(s(n), G(h(n)))\} : d \cup \{s(n)\} \longrightarrow X$$

es también una aproximación, luego $s(n) \in y'$ en cualquier caso.

Finalmente definimos[3]

$$f = \{(n,u) \in \mathbb{N} \times X \mid \bigvee dh(h : d \longrightarrow X \text{ es una aproximación}$$

$$\wedge\, n \in d \wedge u = h(n))\}.$$

Los resultados que hemos obtenidos implican claramente que $f : \mathbb{N} \longrightarrow X$ es una aproximación que cumple lo pedido.

La unicidad se prueba sin dificultad: si f_1 y f_2 cumplen lo pedido, consideramos el conjunto

$$y = \{n \in \mathbb{N} \mid f_1(n) = f_2(n)\}$$

y una simple inducción nos da que $y = \mathbb{N}$. ∎

En particular, aplicando el teorema anterior a cualquier sistema de números naturales $(\mathbb{N}, s, 0)$ obtenemos una aplicación $f : \mathbb{N} \longrightarrow \omega$ tal que $f(0) = 0$ y $\bigwedge n \in \mathbb{N}\ f(s(n)) = f(n)'$. Es inmediato que

$$0 \in f[\mathbb{N}] \wedge \bigwedge n \in \omega(n \in f[\mathbb{N}] \to n' \in f[\mathbb{N}]),$$

luego $f[\mathbb{N}] = \omega$ y así f es suprayectiva. Por el axioma de reemplazo concluimos que $\mathrm{cto}\,\omega$, y esto completa la prueba del teorema 11.7. ∎

Nota Puede probarse que en la demostración del teorema 11.7 resulta imprescindible el uso del axioma de reemplazo, sin el cual la equivalencia no es necesariamente cierta. Lo mismo sucede con otras equivalencias del axioma de infinitud. Por ejemplo, la versión que dio Zermelo al introducir la teoría que hoy lleva su nombre era ésta:

$$\bigvee x(\varnothing \in x \wedge \bigwedge u \in x\, \{u\} \in x).$$

No es difícil probar que a partir de este axioma es posible construir un sistema de números naturales, por lo que (usando el axioma de reemplazo) podemos probar AI. Recíprocamente, si $\mathrm{cto}\,\omega$, sabemos que existe $f : \omega \longrightarrow V$ tal que $f(0) = \varnothing$ y $\bigwedge n \in \omega\, f(n') = \{f(n)\}$, y es claro entonces que $f[\omega]$ (que es un conjunto por el axioma de reemplazo) cumple el axioma de infinitud de Zermelo. Sin embargo, puede probarse que dicho axioma no equivale a AI en la teoría de Zermelo. Vemos así nuevos ejemplos de la utilidad del axioma de reemplazo: sin él tendríamos un abanico de axiomas de infinitud no equivalentes entre sí. ∎

El teorema 11.9 implica que todos los sistemas de números naturales son equivalentes:

[3]Si X es una clase propia, en ZF* necesitamos el esquema de reemplazo, por ejemplo en su versión 10.10, para garantizar la existencia de f. En NBG* podemos definir f como clase, pero también necesitamos el axioma de reemplazo para probar que es un conjunto (por ejemplo, por 10.16).

Teorema 11.10 *Si* $(\mathbb{N}, s, 0)$ *y* $(\tilde{\mathbb{N}}, \tilde{s}, \tilde{0})$ *son sistemas de números naturales, existe una única aplicación* $f : \mathbb{N} \longrightarrow \tilde{\mathbb{N}}$ *biyectiva tal que*

$$f(0) = \tilde{0} \wedge \bigwedge n \in \mathbb{N}\ f(s(n)) = \tilde{s}(f(n)).$$

DEMOSTRACIÓN: La existencia y unicidad de f viene dada por 11.9. Para probar que f es inyectiva razonamos por inducción sobre el conjunto

$$y = \{n \in \mathbb{N} \mid \bigwedge m \in \mathbb{N}(m \neq n \to f(n) \neq f(m))\}.$$

Se cumple que $0 \in y$, pues si $f(0) = f(n)$ pero $n \neq 0$, entonces $n = s(u)$, para cierto $u \in \mathbb{N}$, luego $\tilde{0} = f(s(u)) = \tilde{s}(f(u))$, en contra de la definición de sistema de números naturales.

Si $n \in y$ pero $f(s(n)) = f(m)$, para cierto $m \in \mathbb{N}$, $m \neq s(n)$, ya hemos visto que no puede ser $m = 0$, luego existe un $u \in \mathbb{N}$ tal que $m = s(u)$, luego $\tilde{s}(f(n)) = f(s(n)) = f(s(u)) = \tilde{s}(f(u))$, luego $f(n) = f(u)$, luego $n = u$, por hipótesis de inducción, luego $m = s(u) = s(n)$, contradicción.

La suprayectividad de f se prueba trivialmente por inducción. ∎

Así pues, dados dos sistemas de números naturales, los elementos de uno se pueden poner en correspondencia biunívoca con los del otro de modo que el cero de uno se corresponde con el cero del otro y el sucesor de cada número natural de uno se corresponde con el sucesor de su correspondiente en el otro. Esto viene a decir que los dos sistemas son "esencialmente la misma estructura", por lo que es irrelevante trabajar con uno o con otro. Fijado uno cualquiera, el teorema 11.9 permite construir una suma y un producto $+, \cdot : \mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{N}$ determinados por las propiedades usuales:

$$\bigwedge n \in \mathbb{N}\ n + 0 = n, \quad \bigwedge mn \in \mathbb{N}(m + n' = (m + n)'),$$
$$\bigwedge n \in \mathbb{N}\, n \cdot 0 = 0, \quad \bigwedge mn \in \mathbb{N}(m \cdot n' = mn + m),$$

y a partir de ahí concluimos que cada sistema de números naturales determina una interpretación de AP en ZF^* o NBG^*, y podemos definir a través de ella todos los conceptos aritméticos definibles en AP.

Nota El teorema 11.10 lleva a muchos matemáticos a creer que los axiomas de Peano, en la forma en que aparecen en la definición 11.8, determinan por completo a los números naturales, lo cual es indiscutiblemente cierto en el sentido de que se cumple 11.10, pero hay quienes olvidan que este teorema sólo caracteriza "localmente" a los números naturales, en el seno de un modelo de $\mathrm{ZF}^* + \mathrm{AI}$ (o cualquier otra teoría T equivalente o más potente), en el sentido de que dos sistemas de números naturales en un mismo modelo de T son necesariamente "isomorfos", pero nada impide tener dos modelos de T de tal modo que los sistemas de números naturales de uno no sean isomorfos a los sistemas de números naturales del otro. De hecho, como consecuencia de los teoremas de incompletitud, hemos demostrado que toda teoría aritmética recursiva consistente admite infinitos modelos cuyos números naturales respectivos son "no isomorfos" dos a dos, en el sentido de que satisfacen propiedades distintas. ∎

Pero volvamos a la relación entre conjuntos infinitos y D-infinitos:

Teorema 11.11 *Un conjunto x es D-infinito si y sólo si existe una aplicación $f : \omega \longrightarrow x$ inyectiva.*

DEMOSTRACIÓN: Si x es un conjunto D-infinito, en la prueba del teorema 11.7 hemos visto en primer lugar que existe un sistema de números naturales $(\mathbb{N}, s, 0)$ tal que $\mathbb{N} \subset x$ y, en segundo lugar, que existe una biyección $f : \mathbb{N} \longrightarrow \omega$. Por lo tanto, $f^{-1} : \omega \longrightarrow x$ es la inyección buscada.

Recíprocamente, si $f : \omega \longrightarrow x$ inyectiva, se cumple que x es D-infinito, pues la aplicación $g : x \longrightarrow x$ dada por

$$g(u) = \begin{cases} f(n+1) & \text{si } u = f(n), \\ u & \text{si } u \notin f[\omega], \end{cases}$$

es claramente inyectiva y no suprayectiva. ∎

Así pues, de acuerdo con las definiciones que hemos dado, un conjunto x es infinito si sus elementos no pueden numerarse en la forma $a_0, \ldots, a_{n-1}$, para ningún $n \in \omega$, mientras que es D-infinito si entre sus elementos podemos encontrar una sucesión $a_0, a_1, a_2, \ldots$ sin repeticiones. Teniendo esto en cuenta, alguien podría "sentirse tentado" de demostrar así que todo conjunto infinito es D-infinito:

> Sea x un conjunto infinito. Entonces no es vacío, luego podemos tomar un $a_0 \in x$. No puede ser $x = \{a_0\}$, porque entonces x sería finito (tendría cardinal 1), luego existe un $a_1 \in x \setminus \{a_0\}$, pero no puede ser $x = \{a_0, a_1\}$, pues entonces x tendría cardinal 2 y sería finito. En general, si hemos encontrado elementos distintos $a_0, \ldots, a_n$ en x, no puede ocurrir que $x = \{a_0, \ldots, a_n\}$, porque entonces x sería finito, luego existe un $a_{n+1} \in x \setminus \{a_0, \ldots, a_n\}$, y de este modo construimos una sucesión $a_0, a_1, \ldots$ de elementos de x sin repeticiones, lo que prueba que x es D-infinito.

Aunque suena "convincente", lo cierto es que este argumento no puede ser formalizado en ZF*+ AI (luego tampoco en NBG*+ AI). Lo más parecido que podemos demostrar siguiendo el argumento anterior es que

$$\bigwedge n \in \omega \bigvee s(s : n \longrightarrow x \text{ inyectiva})$$

y, desde luego, cada sucesión s en esas condiciones no será suprayectiva por la infinitud de x, pero eso no implica que exista $s : \omega \longrightarrow x$ inyectiva. Si tratamos de probar esto imitando las numerosas definiciones recurrentes que hemos dado hasta aquí, tendríamos que definir algo así:

$$\phi(s, n) \equiv (n \in \omega \wedge s : n \longrightarrow x \wedge$$

$$\bigwedge i < n(s(i) = \text{"un elemento cualquiera de } x \setminus s[i]\text{"}).$$

Podemos demostrar que $\bigwedge n \in \omega \bigvee s\, \phi(s,n)$, pero no podemos probar (de hecho es obviamente falso) que $\bigwedge n \in \omega \overset{1}{\bigvee} s\, \phi(s,n)$, y esa unicidad que no tenemos es fundamental para acabar definiendo

$$f = \{(i,a) \in \omega \times x \mid \bigvee s(\phi(s,i+1) \wedge s(i) = a)\}.$$

Si tuviéramos unicidad podríamos asegurar que $f : \omega \longrightarrow x$ inyectiva, pero hay muchas formas de elegir "un elemento cualquiera de", y eso hace que cuando intentamos recopilar las aproximaciones finitas de f nos encontremos con muchas alternativas incompatibles, de modo que la f así definida no es ni siquiera una función. En todas las definiciones recurrentes que hemos dado hasta ahora siempre hemos contado con un criterio explícito para especificar cómo debe prolongarse una aproximación finita,[4] por lo que siempre hemos contado con la unicidad necesaria para unir todas las aproximaciones en una función final.

Las observaciones previas al teorema 11.16 más abajo prueban la existencia del conjunto $\mathcal{P}^{\mathrm{f}}x$ de todos los subconjuntos finitos de x. Aplicando el axioma de reemplazo a la fórmula $v = x \setminus u$, obtenemos la existencia del conjunto $\mathcal{P}^{\mathrm{cf}}x$ de todos los subconjuntos de x cuyo complementario es finito. Imaginemos que tuviéramos una función $e : \mathcal{P}^{\mathrm{cf}}x \longrightarrow V$ tal que

$$\bigwedge u \in \mathcal{P}^{\mathrm{cf}}x(u \neq \varnothing \to e(u) \in u),$$

es decir, una función que a cada subconjunto no vacío de x de complementario finito le asignara "uno cualquiera de sus elementos". Con dicha función podríamos llevar adelante la construcción recurrente, pues podríamos definir

$$\phi(s,n) \equiv (n \in \omega \wedge s : n \longrightarrow x \wedge \bigwedge i < n(s(i) = e(x \setminus s[i])),$$

y ahora, cada aproximación s no se prolonga con "un elemento cualquiera de $s \setminus \mathcal{R}s$", sino con "el elemento cualquiera concreto determinado por e", y ahora sí que podemos probar que $\bigwedge n \in \omega \overset{1}{\bigvee} s\, \phi(s,n)$, lo que a su vez permite probar que la extensión común f de todas las aproximaciones es una función que cumple todo lo requerido.

Así pues, lo único que está implícito en el argumento informal que hemos dado y que no puede demostrarse en ZF*+ AI (o NBG*+ AI) es la existencia de la función e. Las funciones que a cada elemento no vacío u de un conjunto x le asignan un elemento $e(u) \in u$ se llaman *funciones de elección,* y el problema es que no tenemos forma de probar la existencia de una función de elección sobre un conjunto arbitrario x (en nuestro caso la necesitábamos sobre $\mathcal{P}^{\mathrm{cf}}x$). Para demostrar que las dos definiciones de infinitud que tenemos son equivalentes necesitamos un axioma que postule la existencia de funciones de elección, y eso es precisamente lo que hace el axioma de elección. Volveremos sobre el problema de la existencia de funciones de elección en la sección que dedicamos más adelante a dicho axioma. ∎

[4] Por considerar el caso más reciente, en la prueba de 11.9 la condición es $h(s(n)) = G(h(n))$, de modo que la clase G es la que determina cómo debe prolongarse cada aproximación y eso nos permite probar que dos aproximaciones coinciden en su dominio común.

El axioma de infinitud nos permite formalizar el concepto de numerabilidad, que hasta ahora sólo hemos manejado metamatemáticamente:

Definición 11.12 Un conjunto x es *numerable* si es finito o existe $f : \omega \longrightarrow x$ biyectiva.

Equivalentemente, un conjunto es numerable si se puede biyectar con un elemento de $\omega' = \omega \cup \{\omega\}$. Los que se pueden biyectar con un número natural son los conjuntos finitos, y los que se pueden biyectar con ω son los infinitos numerables.

En estos términos, el teorema 11.11 afirma que un conjunto es D-infinito si y sólo si contiene un subconjunto infinito numerable. En particular, los conjuntos infinitos numerables son infinitos y D-infinitos.

Veamos algunos resultados útiles para demostrar que un conjunto dado es numerable:

Teorema 11.13 (AI) *Para todo $x \neq \varnothing$, las afirmaciones siguientes son equivalentes:*

1. *x es numerable,*
2. *Existe $f : x \longrightarrow \omega$ inyectiva,*
3. *Existe $g : \omega \longrightarrow x$ suprayectiva.*

Demostración: 1) $\Rightarrow$ 2) es inmediato: si x es finito, entonces existe un $n \in \omega$ y una biyección $f : x \longrightarrow n$, que también puede verse como aplicación $f : x \longrightarrow \omega$ inyectiva. Si x es infinito, existe una f biyectiva por definición de numerabilidad.

2) $\leftrightarrow$ 3) es consecuencia inmediata del teorema 6.29.

Para probar que 2) $\Rightarrow$ 1) basta ver que $f[x]$ es numerable, luego basta probar que todo subconjunto de ω es numerable.

Dado $a \subset \omega$, si existe un n tal que $a \subset n$, entonces a es finito, luego numerable. Supongamos que a no está acotado en ω, es decir, que se cumple $\bigwedge n \in \omega \bigvee m \in a\ m > n$. Definimos entonces $f : \omega \longrightarrow a$ mediante el teorema 3.26 con las condiciones

$$f(0) = \text{mín } a \wedge \bigwedge n \in \omega\ f(n+1) = \text{mín}\{m \in a \mid m > f(n)\}.$$

Claramente $\bigwedge n \in \omega\ f(n+1) > f(n)$, y una simple inducción prueba que $\bigwedge mn \in \omega(m < n \to f(m) < f(n))$. En particular f es inyectiva. En particular $f[\omega]$ es infinito, luego no puede estar acotado. Dado $m \in a$, tiene que existir un $n \in \omega$ tal que $f(n) \geq m$. Tomemos el mínimo posible. Si es $n = 0$, entonces $f(0)$ es el mínimo de a, luego $f(0) \leq m \leq f(0)$, luego $f(0) = m$. Si $n = k + 1$ entonces $f(k) < m$ por la minimalidad de n, luego $f(n) \leq m$ por la definición de f, luego igualmente $m = f(n)$. Esto prueba que $f[\omega] = a$, luego f es biyectiva. ∎

Ahora es inmediato que si $f : x \longrightarrow y$ es biyectiva y uno de los dos conjuntos es numerable, entonces el otro también lo es. Si f es inyectiva e y es numerable, x también lo es, y si f es suprayectiva y x es numerable, entonces y también lo es. En particular, todo subconjunto de un conjunto numerable es numerable.

Teorema 11.14 (AI) *Si x e y son conjuntos numerables, también lo son su unión $x \cup y$ y su producto cartesiano $x \times y$.*

DEMOSTRACIÓN: Los pares ordenados aritméticos $\langle m, n\rangle$ definen una biyección $\omega \times \omega \longrightarrow \omega$. Es claro que podemos construir una aplicación inyectiva $x \times y \longrightarrow \omega \times \omega$ que, compuesta con la biyección anterior, nos da una aplicación inyectiva $x \times y \longrightarrow \omega$ que prueba la numerabilidad del producto.

Si al menos uno de los dos conjuntos no es vacío, es fácil definir una aplicación suprayectiva $\omega \times 2 \longrightarrow x \cup y$, y $\omega \times 2$ es numerable por la parte ya probada, luego la unión también. ∎

En la sección 6.4 hemos visto que, para todo conjunto x y todo $n \in \omega$, es posible demostrar la existencia del conjunto x^n de aplicaciones $n \longrightarrow x$ y, más aún, podemos definir[5] el conjunto de sucesiones finitas en x:

$$x^{<\omega} \equiv \bigcup_{n\in\omega} x^n$$

Teorema 11.15 (AI) *Si x es numerable, también lo es $x^{<\omega}$.*

DEMOSTRACIÓN: A partir de una inyección $x \longrightarrow \omega$ es fácil definir una inyección $x^{<\omega} \longrightarrow \omega^{<\omega}$, luego basta probar que $\omega^{<\omega}$ es numerable, pero ya sabemos que toda sucesión finita de números naturales puede codificarse mediante un número natural. Una forma rápida de hacerlo es

$$s \mapsto \prod_{i<\ell(s)} p_i^{s_i},$$

donde $\{p_i\}_{i\in\omega}$ es la sucesión de los números primos. Tenemos así una aplicación $\omega^{<\omega} \longrightarrow \omega$ inyectiva. ∎

El axioma de reemplazo aplicado a la función $f : x^{<\omega} \longrightarrow V$ dada por $s \mapsto \mathcal{R}s$ nos da la existencia del conjunto de *partes finitas* de un conjunto x:

$$\mathcal{P}^{\mathrm{f}}x \equiv y|\bigwedge u(u \in y \leftrightarrow u \subset x \wedge u \text{ es finito}).$$

De hecho, $f : x^{<\omega} \longrightarrow \mathcal{P}^{\mathrm{f}}x$ suprayectiva, luego

Teorema 11.16 (AI) *Si x es numerable, también lo es $\mathcal{P}^{\mathrm{f}}x$.*

[5] Aquí tenemos un nuevo uso del axioma de reemplazo, que nos da el conjunto $\{x^n \mid n \in \omega\}$.

Nota La teoría de cardinales que hemos desarrollado en la sección 6.4 para conjuntos finitos se generaliza fácilmente a conjuntos numerables extendiendo así la definición de cardinal:

$$|x| \equiv c|(c \in \omega \cup \{\omega\} \wedge \bigvee f(f : c \longrightarrow x \text{ biyectiva}))$$

De este modo, si x es un conjunto finito su cardinal $|x|$ es el único número natural biyectable con él, mientras que si x es infinito numerable, entonces $|x| = \omega$ (y en cualquier caso x es biyectable con $|x|$). No obstante, cuando se considera a ω como cardinal, es tradición mantener la notación cantoriana y representarlo como $\aleph_0$ (álef cero).

La suma y el producto de cardinales pueden definirse como en el caso de los conjuntos finitos, y entonces el teorema 11.14 implica fácilmente que

$$\aleph_0 + \aleph_0 = \aleph_0 \cdot \aleph_0 = \aleph_0,$$

así como que $\bigwedge n \in \omega\, \aleph_0 + n = \aleph_0$ y $\bigwedge n \in \omega \setminus \{0\}\, \aleph_0 \cdot n = \aleph_0$. ∎

Conjuntos numéricos Las fórmulas $x \in \mathbb{Z}$ y $x \in \mathbb{Q}$ que en la sección 6.6 definimos en KP pueden definirse igualmente en ZF*, pero con AI es posible simplificar conceptualmente las construcciones.

Por ejemplo, los números enteros se definen a partir de la relación de equivalencia en $\mathbb{N} \times \mathbb{N}$ dada por[6]

$$(a, b) \sim (c, d) \leftrightarrow a + d = b + c.$$

La idea es que un número entero "debe ser" una clase de equivalencia de pares respecto de esta relación, pero en ZF* no podemos demostrar que las clases de equivalencia sean conjuntos (pues serían conjuntos infinitos) y así, para evitar que cada número entero fuera una clase propia, consideramos pares ordenados aritméticos (de modo que cada par es un número natural) y definimos los números enteros como los pares (que son números naturales) mínimos en sus respectivas clases de equivalencia. Si admitimos AI no necesitamos estos artificios. Podemos considerar pares ordenados conjuntistas, así como las clases de equivalencia usuales:

$$[a, b] \equiv \{(c, d) \in \mathbb{N} \times \mathbb{N} \mid (c, d) \sim (a, b)\},$$

que son subconjuntos de $\mathbb{N} \times \mathbb{N}$, y definimos el conjunto de los números enteros como el conjunto cociente $\mathbb{Z} \equiv (\mathbb{N} \times \mathbb{N})/\sim$, que es un conjunto por el axioma de reemplazo, según vimos en el capítulo anterior. A partir de aquí, todos los conceptos aritméticos sobre números enteros pueden definirse igualmente.

Ahora tenemos dos definiciones de números enteros, los definidos por la fórmula $x \in \mathbb{Z}$ de la sección 6.6, la cual determina una clase $\mathbb{Z}_0 \subset \mathbb{N}$, que por AI es un conjunto, y los que acabamos de definir. La relación entre ambos es que

[6] Aquí usamos $\mathbb{N}$ como sinónimo de ω.

tenemos una biyección $f : \mathbb{Z}_0 \longrightarrow \mathbb{Z}$ dada por $f(n) = [(n_0, n_1)]$. Esta biyección hace corresponder la suma y el producto que tenemos definidos en $\mathbb{Z}_0$ con las operaciones que podemos definir análogamente sobre $\mathbb{Z}$, de modo que resulta indistinto trabajar con una u otra definición.

Similarmente, los números racionales pueden definirse como los elementos del cociente $\mathbb{Q} \equiv (\mathbb{Z} \times (\mathbb{Z} \setminus \{0\}))/\sim$, donde ahora la relación de equivalencia es la dada por

$$(a, b) \sim (c, d) \leftrightarrow ad = bc.$$

Todas las propiedades aritméticas de los números naturales, enteros y racionales vistas en la sección 6.4 se formalizan en ZF*+ AI (o NBG*+ AI) con total naturalidad. ∎

11.2 El axioma de partes

Del mismo modo que en ZF* no puede probarse la existencia de conjuntos infinitos, puede probarse que, si ZF*+ AI es consistente, en esta teoría no puede demostrarse la existencia de conjuntos no numerables.[7] Tales conjuntos aparecen en la teoría como consecuencia del axioma de partes:

Definición 11.17 Llamaremos *axioma de partes:* a la sentencia

$$\mathrm{AP} \equiv \bigwedge x \bigvee y \bigwedge u(u \subset x \to u \in y).$$

Así, AP afirma la existencia de un conjunto y que contiene a todos los subconjuntos de un conjunto dado x. Por especificación y extensionalidad de aquí se sigue que

$$\bigwedge x \overset{1}{\bigvee} y \bigwedge u(u \in y \leftrightarrow u \subset x),$$

lo que nos permite definir el conjunto de partes

$$\mathcal{P}x \equiv y | \bigwedge u(u \in y \leftrightarrow u \subset x).$$

El teorema 6.36 prueba la existencia del conjunto de partes para todo conjunto finito, de modo que el axioma de partes es un teorema de ZF*+ "todo conjunto es finito".

Por supuesto, aun sin el axioma de partes, podemos definir la clase de partes de cualquier clase X:

$$\mathcal{P}X \equiv \{u \mid u \subset X\},$$

y lo que afirma AP es que la clase de las partes de un conjunto es también un conjunto.

Naturalmente, los conjuntos no numerables aparecen cuando el axioma AP se combina con AI en virtud del teorema de Cantor, cuya prueba anticipamos en la introducción y en la página 143:

[7] Lo probaremos en 12.19. Véanse las observaciones anteriores y posteriores.

Teorema 11.18 (Teorema de Cantor) *Si x es un conjunto, existe una aplicación inyectiva $f : x \longrightarrow \mathcal{P}x$, pero no existen aplicaciones suprayectivas (ni en particular biyectivas) entre ambos conjuntos.*

DEMOSTRACIÓN: Como aplicación inyectiva basta tomar la definida mediante $f(u) = \{u\}$. Supongamos ahora que existiera una aplicación f suprayectiva. Entonces podríamos definir el conjunto $a = \{u \in x \mid u \notin f(u)\} \in \mathcal{P}x$.

Como f es suprayectiva existe un $u \in x$ tal que $f(u) = a$, pero esto nos lleva a una contradicción, porque, o bien $u \in a$ o bien $u \notin a$, pero si $u \in a = f(u)$, debería ser $u \notin a$, por definición de a, mientras que si $u \notin a = f(u)$, debería ser $u \in a$. Concluimos que f, sea la aplicación que sea, no puede ser suprayectiva. ∎

Cuando aplicamos este teorema a un conjunto numerable, vemos que $\mathcal{P}x$ es claramente infinito, pero no puede ser numerable, porque entonces podría biyectarse con x.

La paradoja de Cantor La existencia de conjuntos no numerables es en cierto sentido el "residuo" que deja en la teoría de conjuntos la llamada paradoja de Cantor. Ésta surgía al tratar de aplicar el teorema anterior al "conjunto" V de todos los conjuntos. Puesto que los elementos de $\mathcal{P}V$ son conjuntos, tiene que cumplirse $\mathcal{P}V \subset V$ (de hecho, en NBG* se prueba trivialmente que $\mathcal{P}V = V$), pero entonces sí que existe una aplicación suprayectiva $V \longrightarrow \mathcal{P}V$, en contra de lo que afirma el teorema de Cantor.

Naturalmente, esto no da lugar a una contradicción en NBG* porque la clase V no es un conjunto. De hecho, este argumento es una prueba alternativa de que V no puede ser un conjunto. El argumento de Cantor requiere de forma esencial que la clase a la que se aplica sea un conjunto. Para ver por qué hace falta esa hipótesis basta rastrear la prueba en el caso concreto en que $f : V \longrightarrow \mathcal{P}V$ es la aplicación identidad. Entonces podemos definir la clase

$$A = \{u \in V \mid u \notin f(u)\},$$

que no es sino la clase de Russell $R = \{u \in V \mid u \notin u\}$. En lugar de llegar a una contradicción, llegamos a la conclusión de que A no es un conjunto, por lo que no es cierto que $A \in \mathcal{P}V$, por lo que no tiene ninguna antiimagen u que nos lleve a la contradicción.

Vemos así que la distinción entre clases y conjuntos nos libra de la paradoja de Cantor, al igual que nos libra de las demás paradojas conocidas de la teoría de conjuntos, y lo hace sin eliminar de la teoría el concepto sospechoso de "conjunto de partes", que es esencialmente el único concepto conjuntista sin un significado intuitivo preciso. A día de hoy no tenemos ninguna evidencia de que los conjuntos de partes lleven a contradicción alguna. El hecho de que los admitamos en la teoría de conjuntos no lleva a una contradicción, sino a un hecho que trasciende nuestra capacidad de razonamiento informal: la existencia de conjuntos no numerables. ∎

La teoría de Zermelo El interés principal del axioma de partes es que permite formalizar de forma natural cualquier construcción conjuntista que un matemático pueda necesitar en un momento dado. De hecho, cuando contamos con el axioma de partes, podemos prescindir en gran medida del axioma de reemplazo. En efecto, ahora podemos presentar la teoría completa de Zermelo:

Definición 11.19 La *teoría de conjuntos de Zermelo* Z es la teoría axiomática sobre $\mathcal{L}_{\mathrm{tc}}$ cuyos axiomas son los siguientes:[8]

Extensionalidad	$\bigwedge xy(\bigwedge u(u \in x \leftrightarrow u \in y) \to x = y)$
Par	$\bigwedge xy\bigvee z\bigwedge u(u \in z \leftrightarrow u = x \vee u = y)$
Unión	$\bigwedge x\bigvee y\bigwedge u(u \in y \leftrightarrow \bigvee v(u \in v \wedge v \in x))$
Especificación	$\bigwedge x\bigvee y\bigwedge u(u \in y \leftrightarrow u \in x \wedge \phi(u))$ (*)
Partes	$\bigwedge x\bigvee y\bigwedge u(u \subset x \to u \in y)$
Infinitud	$\bigvee x(\bigvee u \in x\bigwedge v\ v \notin u \wedge$ $\bigwedge u \in x\bigvee v \in x\bigwedge w(w \in v \leftrightarrow w \in u \vee w = u))$

(*) para toda fórmula $\phi(u)$ tal vez con más variables libres, distintas de y.

Tenemos que Z es una subteoría de ZF^*+ AI + AP y, según comentábamos, sus axiomas bastan para formalizar de forma natural la práctica totalidad de las construcciones conjuntistas que usan los matemáticos habitualmente, salvo las de los especialistas en teoría de conjuntos, cuyas exigencias van mucho más allá.[9]

En efecto, el axioma del par permite construir pares ordenados, y la existencia del producto cartesiano puede demostrarse por especificación, como subconjunto $x \times y \subset \mathcal{PP}(x \cup y)$, lo que resulta mucho más natural que apelar dos veces al axioma de reemplazo y una vez al de la unión. También hemos advertido que los conjuntos cociente pueden definirse por especificación, como subconjuntos $x/r \subset \mathcal{PP}x$. El axioma de infinitud permite construir el conjunto $\mathbb{N} \equiv \omega$ de los números naturales, y a partir de él $\mathbb{Z}$ y $\mathbb{Q}$ pueden definirse mediante cocientes de productos cartesianos, según hemos visto en la sección anterior.

Hasta aquí habíamos llegado sin el axioma de partes, pero ahora podemos ir más lejos y definir, por ejemplo, el conjunto

$$x^y \equiv z|\bigwedge f(f \in z \leftrightarrow f : y \longrightarrow x),$$

cuya existencia se demuestra por especificación, pues $x^y \subset \mathcal{PPP}(x \cup y)$. En particular tenemos la existencia de los conjuntos de sucesiones x^ω y también de sucesiones finitas $x^{<\omega} \subset \mathcal{PPP}(\omega \cup x)$.

[8] Ya hemos comentado que el axioma de infinitud que formuló Zermelo era otro distinto y no equivalente en su propia teoría al que estamos considerando, aunque cumple igualmente su función, que es garantizar la existencia de un sistema de números naturales y demostrar que dos cualesquiera de ellos son isomorfos.

[9] En realidad, para formalizar algunos teoremas importantes del álgebra, el análisis, etc. es necesario añadir a Z el axioma de elección, que discutiremos más adelante, pero esto no invalida nuestra afirmación, pues el axioma de elección no proporciona nuevas construcciones conjuntistas, sino que más bien postula o implica la existencia de determinados objetos que no pueden ser construidos explícitamente.

Más en general, para cada familia de conjuntos $\{x_i\}_{i\in a}$, podemos probar que son conjuntos los productos cartesianos generalizados:

$$\prod_{i\in a} x_i = \{s \in (\bigcup_{i\in a} x_i)^a \mid \bigwedge i \in a\, s_i \in x_i\},$$

etc.

También podemos definir en Z el conjunto $\mathbb{R}$ de los números reales. No vamos a entrar en detalles al respecto, pero señalaremos únicamente que las dos construcciones más habituales pueden formalizarse naturalmente en Z: la construcción de Dedekind introduce a $\mathbb{R}$ como un cierto subconjunto de $\mathcal{P}\mathbb{Q}$ (que se define sin problemas por especificación), mientras que la construcción de Cantor consiste en considerar el conjunto S de las sucesiones de Cauchy de números racionales (que se define de forma natural por especificación a partir del conjunto $\mathbb{Q}^{\mathbb{N}}$ de todas las sucesiones de números racionales) y $\mathbb{R}$ se introduce como un cociente de S respecto de una relación de equivalencia.

A partir de $\mathbb{R}$ se define sin dificultad el conjunto $\mathbb{C}$ de los números complejos y, por supuesto, junto con estos conjuntos numéricos es posible definir todas sus estructuras asociadas algebraicas y topológicas.

11.3 El axioma de regularidad I

El axioma de regularidad es totalmente prescindible en todas las áreas de la matemática excepto en la teoría de conjuntos propiamente dicha, pues lo único que hace es "regularizar" la clase universal y prohibir posibles patologías que, en caso de darse, no afectarían en nada al trabajo cotidiano de los matemáticos de otras áreas.

Por ejemplo, ninguno de los axiomas que hemos discutido hasta ahora impide que exista[10] un conjunto x con la propiedad de que $x = \{x\}$, o dos conjuntos x, y tales que $x = \{y\} \wedge y = \{x\}$, o un conjunto $x_0 = \{x_1\}$, donde $x_1 = \{x_2\}$, y a su vez $x_2 = \{x_3\}$ y así sucesivamente.

Cualquiera de estos conjuntos contradice la idea de que un conjunto "debería" resultar de agrupar unos objetos previamente existentes, de modo que conocemos un conjunto cuando conocemos sus elementos, pero si $x = \{x\}$ e intentamos "comprender" qué es x estudiando sus elementos, nos encontramos con que para entender qué es x tendríamos que entender antes qué es x. Similarmente, si tenemos una cadena decreciente de conjuntos $x_0 = \{x_1\}$, $x_1 = \{x_2\}$, etc., resulta que para saber qué es realmente x_0 tendríamos que saber qué es x_1, para lo cual tendríamos que saber qué es x_2, y así sucesivamente, de modo que nunca podemos acabar de "entender" qué es x_0.

Pese a todo, no hay nada de paradójico en estos conjuntos y, como decimos, ninguno de los axiomas que hasta ahora hemos considerado contradice la posibilidad de que existan.

[10] El teorema 12.23 y siguientes demuestran estos hechos.

Ahora bien, a un matemático que trabaja en álgebra, análisis, geometría, etc., aunque lo haga en el marco de una teoría de conjuntos, le importa poco que pueda haber o no conjuntos así. Sencillamente, si los hay, no les hará caso y ya está. La existencia de tales conjuntos patológicos sólo incomoda realmente a los matemáticos que estudian la teoría de conjuntos propiamente dicha, pues de que existan o no tales patologías depende la validez o no de ciertos principios generales.

Definición 11.20 El *axioma de regularidad* (o de *fundación*) es la sentencia

$$\bigwedge x(\bigvee u\, u \in x \to \bigvee u \in x \neg \bigvee v(v \in u \wedge v \in x)),$$

que, introduciendo descriptores, equivale claramente a

$$\bigwedge x(x \neq \varnothing \to \bigvee u(u \in x \wedge u \cap x = \varnothing)).$$

Ya nos habíamos encontrado antes este axioma y esta propiedad. En 3.9, a los conjuntos u tales que $u \in x \wedge u \cap x = \varnothing$ los llamamos $\in$-minimales de x, de modo que el axioma de regularidad afirma que todo conjunto no vacío tiene un $\in$-minimal. En 3.9 definimos también el concepto de conjunto bien fundado, que es un conjunto tal que todos sus subconjuntos no vacíos tienen un $\in$-minimal. Es obvio entonces que el axioma de regularidad equivale a la sentencia "todo conjunto está bien fundado".

Ahora bien, no podemos identificar el concepto de "conjunto mal fundado" con el de "conjunto patológico" en el sentido que explicábamos hace un momento, pues, por ejemplo, si dos conjuntos cumplen $x = \{y\} \wedge y = \{x\} \wedge x \neq y$, entonces x es un conjunto bien fundado, pero no por ello deja de ser patológico. De todos modos, ahora es claro que el axioma de regularidad impide que puedan existir conjuntos así, ya que entonces el conjunto $\{x, y\}$ no tendría $\in$-minimal.

Recordemos también que en la página 196 vimos que el axioma de regularidad puede demostrarse en KP a partir del esquema de Π_1-regularidad, pero nunca hemos llegado a extraer ninguna consecuencia de ello, salvo el hecho (de gran importancia a la hora de trabajar en KP) de que simplifica la definición de ordinal, puesto que, como implica que todo conjunto está bien fundado, podemos eliminar esta condición de la definición de ordinal 3.9.

Para empezar a entender lo que supone realmente el axioma de regularidad consideramos una consecuencia sencilla: no pueden existir conjuntos $x_1, \ldots, x_n$ sobre los que la pertenencia sea circular, es decir, tales que

$$x_1 \in x_n \in x_{n-1} \in \cdots \in x_2 \in x_1.$$

Si existieran, es claro que el conjunto $x = \{x_1, \ldots, x_n\}$ no tendría $\in$-minimal.

En particular (por el caso $n = 1$) tenemos que el axioma de regularidad implica que $\bigwedge x\ x \notin x$, es decir, que la clase de Russell (la clase de todos los conjuntos que no se pertenecen a sí mismos) coincide con la clase universal V.

Con el axioma de infinitud podemos demostrar una versión más general del resultado anterior:

Teorema 11.21 (AI) *Si se cumple el axioma de regularidad, no existe ninguna sucesión* $\{x_n\}_{n\in\omega}$ *que cumpla* $\bigwedge n \in \omega\ x_{n+1} \in x_n$.

DEMOSTRACIÓN: En efecto, si existiera tal sucesión, su rango, es decir, el conjunto $\{x_n \mid x \in \omega\}$ no tendría $\in$-minimal. ∎

Más gráficamente, no pueden existir sucesiones decrecientes respecto de la pertenencia:

$$\cdots x_3 \in x_2 \in x_1 \in x_0.$$

Notemos que no exigimos que los términos de la sucesión sean distintos entre sí, por lo que este teorema incluye el caso previo que hemos discutido sobre relaciones circulares de pertenencia. Si tenemos, por ejemplo $x \in y \in z \in x$, tenemos la sucesión

$$\cdots x \in y \in z \in x \in y \in z.$$

Nota Es posible que el lector se "sienta tentado" a demostrar así el recíproco del teorema anterior:

> Si no se cumple el axioma de regularidad, existe un conjunto $x_0 \neq \varnothing$ que no tiene $\in$-minimal. Como no es vacío, podemos tomar $x_1 \in x_0$. Como x_1 no puede ser $\in$-minimal en x_0, existe $x_2 \in x_1 \cap x_0$. Como x_2 no es $\in$-minimal en x_0, existe $x_3 \in x_2 \cap x_0$, y de este modo podemos prolongar indefinidamente la sucesión $\cdots x_4 \in x_3 \in x_2 \in x_1 \in x_0$, luego existe una sucesión en las condiciones del teorema anterior.

Por desgracia, estamos ante un nuevo caso de argumento "convincente" que no puede demostrarse sin el axioma de elección, exactamente por el mismo motivo que en el caso de la falsa demostración que analizamos tras el teorema 11.11: la falta de unicidad en la prolongación de las aproximaciones finitas de la sucesión que queremos construir hace que existan aproximaciones mutuamente contradictorias, por lo que, aunque podemos probar que existen aproximaciones de todas las longitudes, no podemos reunirlas en una misma sucesión final. ∎

Aunque la formulación que hemos presentado del axioma de regularidad es la más habitual, en ausencia del axioma de infinitud no es lo suficientemente potente. Ésa es la razón por la que en KP no aparece como axioma, sino que en su lugar tenemos el esquema de Π_1-regularidad, que es más potente. Para entender lo que sucede conviene considerar un nuevo axioma:

Definición 11.22 Llamaremos *axioma de la clausura transitiva* a la sentencia

$$\mathrm{CT} \equiv \bigwedge x \bigvee y (x \subset y \wedge \bigwedge u \in y\ u \subset y).$$

Recordemos (definición 3.9) que los conjuntos que cumplen $\bigwedge u \in y\ u \subset y$ se llaman conjuntos transitivos. El axioma de la clausura transitiva afirma que todo conjunto está contenido en un conjunto transitivo. Notemos que una forma más compacta de expresar que un conjunto es transitivo es $\bigcup y \subset y$. El nombre de "axioma de la clausura transitiva" se explica por el teorema siguiente, que afirma que todo conjunto está contenido en un mínimo conjunto transitivo:

Teorema 11.23 *El axioma* CT *equivale a*

$$\bigwedge x \overset{1}{\bigvee} y(x \subset y \wedge \bigcup y \subset y \wedge \bigwedge z(x \subset z \wedge \bigcup z \subset z \to y \subset z)).$$

DEMOSTRACIÓN: Es obvio que esta sentencia implica CT. Recíprocamente, si x es un conjunto cualquiera, por CT existe un conjunto transitivo w tal que $x \subset w$. Definimos

$$y = \{u \in w \mid \bigwedge z(x \subset z \wedge \bigcup z \subset z \to u \in z)\}.$$

Así, y es el conjunto de los conjuntos que pertenecen a todos los conjuntos transitivos que contienen a z. Obviamente $x \subset z$, y también se cumple que y es transitivo, pues si $u \in v \in y$, entonces, para todo conjunto transitivo z tal que $x \subset z$, tenemos que $u \in v \in z$, luego $u \in z$. En particular esto vale para $z = w$, luego $u \in w$, y así $u \in y$.

Esto prueba que y cumple lo exigido, y la unicidad es clara, pues si y_1 e y_2 cumplen ambos la condición, entonces ambos son transitivos, luego por la condición de minimalidad $y_1 \subset y_2 \wedge y_2 \subset y_1$, es decir, $y_1 = y_2$. ∎

Esto da pie a la definición siguiente:

Definición 11.24 Se define la *clausura transitiva* de un conjunto x como

$$\operatorname{ct} x \equiv y|(x \subset y \wedge \bigcup y \subset y \wedge \bigwedge z(x \subset z \wedge \bigcup z \subset z \to y \subset z)).$$

Acabamos de probar que CT equivale a que $\operatorname{ct} x$ esté definida para todo conjunto x. Más aún, de la prueba se desprende algo más fuerte: basta con que un conjunto x esté contenido en un conjunto transitivo w para que exista $\operatorname{ct} x$. A menudo es útil la relación siguiente:

Teorema 11.25 *Si x es un conjunto tal que cada $u \in x$ está contenido en un conjunto transitivo, entonces x está contenido en un conjunto transitivo, y además*

$$\operatorname{ct} x = x \cup \bigcup_{u \in x} \operatorname{ct} u.$$

DEMOSTRACIÓN: Sabemos que para todo $u \in x$ existe $\operatorname{ct} u$. Podemos definir $a = x \cup \bigcup_{u \in x} \operatorname{ct} u$, pues por reemplazo aplicado a la fórmula $y = \operatorname{ct} u$ se prueba la existencia del conjunto $b = \{\operatorname{ct} u \mid u \in x\}$, y entonces $a = x \cup \bigcup b$.

Vamos a probar que a es la clausura transitiva de x. Claramente $x \subset a$. Si $v \in u \in a$, entonces, o bien $u \in x$, en cuyo caso $v \in \operatorname{ct} u \subset a$, luego $v \in a$, o bien existe un $u' \in x$ tal que $u \in \operatorname{ct} u'$, en cuyo caso $v \in \operatorname{ct} u'$, por transitividad, luego $v \in a$, en cualquier caso. Esto prueba que a es transitivo.

Supongamos ahora que z es un conjunto transitivo tal que $x \subset z$. Si $u \in x$, entonces $u \in z$, luego $u \subset z$, luego $\operatorname{ct} u \subset z$. Es claro entonces que $a \subset z$, luego a cumple la definición de clausura transitiva. ∎

El teorema siguiente explica por qué CT no es un axioma muy conocido:

Teorema 11.26 AI → CT.

DEMOSTRACIÓN: Sea x un conjunto arbitrario. Aplicamos el teorema de recursión 3.26 para construir una función $\mathrm{ct}(x) : \omega \longrightarrow V$ determinada por las condiciones:

$$\mathrm{ct}_0(x) = x \wedge \bigwedge n \in \omega\ \mathrm{ct}_{n+1}(x) = \bigcup \mathrm{ct}_n(x).$$

Por el axioma de infinitud, $\mathrm{cto}\,\omega$ y por reemplazo también es un conjunto el rango de la aplicación, es decir, el conjunto $a = \{\mathrm{ct}_n(x) \mid n \in \omega\}$, luego también es un conjunto su unión, es decir,

$$\mathrm{ct}\,x \equiv \bigcup_{n\in\omega} \mathrm{ct}_n(x).$$

Claramente $x = \mathrm{ct}_0(x) \subset \mathrm{ct}\,x$. Si $u \in v \in \mathrm{ct}(x)$, entonces existe un $n \in \omega$ tal que $u \in v \in \mathrm{ct}_n(x)$, luego $u \in \bigcup \mathrm{ct}_n(x) = \mathrm{ct}_{n+1}(x) \subset \mathrm{ct}\,x$. Esto prueba que el conjunto $\mathrm{ct}\,x$ es transitivo.

Con esto ya estaría probado TC, pero vamos a ver que $\mathrm{ct}(x)$ es precisamente la clausura transitiva de x. Para ello tomamos un conjunto transitivo z tal que $x \subset z$.

Por definición $\mathrm{ct}_0(x) \subset z$. Si $\mathrm{ct}_n(x) \subset z$, entonces $\mathrm{ct}_{n+1}(x) \subset z$, pues si $u \in \mathrm{ct}_{n+1}(x) = \bigcup \mathrm{ct}_n(x)$, existe un v tal que $u \in v \in \mathrm{ct}_n(x) \subset z$, y por la transitividad de z tenemos que $u \in z$.

Concluimos por inducción que $\bigwedge n \in \omega\ \mathrm{ct}_n(x) \subset z$, de donde $\mathrm{ct}\,x \subset z$. ∎

Sin embargo, en 12.26 veremos que sin AI es imposible demostrar CT, y debemos tener esto en cuenta a la hora de definir un conjunto "no patológico":

Definición 11.27 Diremos que un conjunto x es *regular* si está contenido en un conjunto transitivo bien fundado. Llamaremos R a la clase de todos los conjuntos regulares.

Así, si x es regular, existe un conjunto z transitivo y bien fundado tal que $x \subset z$, luego existe $\mathrm{ct}\,x \subset z$ y, como todo subconjunto de un conjunto bien fundado está bien fundado, concluimos que la clausura $\mathrm{ct}\,x$ está bien fundada. El recíproco es trivialmente cierto, es decir, un conjunto es regular si y sólo si tiene clausura transitiva y ésta está bien fundada.

Veamos las propiedades básicas de los conjuntos regulares:

Teorema 11.28 *Se cumple:*

1. *R es una clase transitiva.*
2. *$\Omega \subset R$, luego R es una clase propia.*
3. *Toda subclase de R no vacía tiene $\in$-minimal.*
4. *$\mathcal{P}R = R$.*
5. *$\bigwedge A(R \cap \mathcal{P}A \subset A \to R \subset A)$.*
   En particular, $\bigwedge A(\mathcal{P}A \subset A \to R \subset A)$.

DEMOSTRACIÓN: 1) Se trata de probar que los elementos de los conjuntos regulares son regulares. Supongamos que $u \in v \in R$. Entonces $u \in \operatorname{ct} v$, luego $u \subset \operatorname{ct} v$, luego u está contenido en un conjunto transitivo y bien fundado, luego $u \in R$.

2) Todo ordinal es un conjunto transitivo y bien fundado, luego cumple la definición de conjunto regular.

3) Sea $A \subset R$ una clase no vacía y tomemos $y \in A$. Si $y \cap A = \varnothing$, entonces y es ya un $\in$-minimal de A. En caso contrario, sea $u \in y \cap A$. Como y es regular, tiene clausura transitiva $z = \operatorname{ct} y$ bien fundada. Definimos $x = z \cap A$, que no es vacío, pues $u \in x$. Como z está bien fundada, x tiene un $\in$-minimal u, que es también un $\in$-minimal de A, ya que $u \in x \subset A$ y si $v \in u \cap A$ entonces $v \in u \in x \subset z$, luego $v \in z$ por la transitividad de z, luego $v \in u \cap x = \varnothing$, contradicción. Por lo tanto, $u \cap A = \varnothing$.

4) La inclusión $R \subset \mathcal{P}R$ es equivalente a la transitividad de R. Si $x \subset R$, entonces para cada $u \in x$ existe $\operatorname{ct} u$ y está bien fundada, luego $\operatorname{ct} u \in R$, luego $\operatorname{ct} u \subset R$. Por 11.25 también x tiene clausura transitiva $\operatorname{ct} x \subset R$, y esta clausura está bien fundada por 3), luego $x \in R$.

5) Si R no está contenida en A, por c) existe un $\in$-minimal $u \in R \setminus A$, pero entonces $u \in (R \cap \mathcal{P}A) \setminus A$. ∎

Observemos que 5) es un principio de inducción: si queremos probar que todo conjunto regular tiene una propiedad (pertenecer a la clase A) podemos tomar como hipótesis de inducción que todos los elementos de un conjunto regular x tienen la propiedad y demostrar a partir de ahí que x también la tiene.

Obviamente esto es falso para conjuntos no regulares. Por ejemplo, si se cumple $x = \{x\}$, entonces es trivialmente cierto que si todos los elementos de x cumplen una propiedad, también la cumple x, pero de ahí no podemos concluir que x cumple todas las propiedades imaginables.

El axioma de regularidad se llama así porque pretende ser equivalente a que todo conjunto sea regular, pero dicha equivalencia sólo es válida en presencia del axioma de infinitud. Sin él se necesita CT:

Teorema 11.29 *Las afirmaciones siguientes son equivalentes:*

1. $V = R$,
2. *Toda clase no vacía tiene un $\in$-minimal,*
3. $\bigwedge A(\mathcal{P}A \subset A \to A = V)$,
4. *El axioma de regularidad más* CT.

DEMOSTRACIÓN: 1) $\Rightarrow$ 2) es inmediato por el apartado 3) del teorema anterior.

2) $\Rightarrow$ 3) Si $A \neq V$, podemos tomar un $\in$-minimal $x \in V \setminus A$, pero entonces, $x \in \mathcal{P}A \subset A$, contradicción.

3) $\Rightarrow$ 4) El teorema 11.25 implica que la clase A de todos los conjuntos que tienen clausura transitiva cumple $\mathcal{A} \subset A$, luego $A = V$, es decir, se cumple CT.

Por otra parte, por el teorema anterior, $\mathcal{P}R = R$, luego $V = R$, luego toda clase no vacía tiene un $\in$-minimal, luego todo conjunto no vacío tiene $\in$-minimal, y eso es lo que afirma el axioma de regularidad.

4) $\Rightarrow$ 1) Trivialmente, por CT todo conjunto tiene clausura transitiva y, por el axioma de regularidad ésta está bien fundada, luego todo conjunto es regular. ∎

Observaciones El axioma de regularidad implica que todo conjunto no vacío tiene $\in$-minimal, y lo que aporta CT es que extiende este hecho a las clases propias. En ZF*, esta versión para clases propias es un esquema:

Para toda fórmula $\phi(u)$, tal vez con más variables libres,

$$\bigvee u\, \phi(u) \to \bigvee u(\phi(u) \wedge \bigwedge v \in u \neg\phi(v)).$$

Notemos que este esquema restringido a fórmulas Π_1 es el esquema axiomático de Π_1-regularidad de KP.

Similarmente, en ZF*, el apartado 3) del teorema anterior tiene que enunciarse como un esquema:

Para toda fórmula $\phi(u)$, tal vez con más variables libres,

$$\bigwedge x(\bigwedge u \in x\, \phi(u) \to \phi(x)) \to \bigwedge x\, \phi(x),$$

que restringido a fórmula Σ_1 es un teorema de KP.

El teorema 11.28 explica por qué el axioma de regularidad es prácticamente irrelevante en la formalización de los razonamientos matemáticos: todos los conjuntos que manejan habitualmente los matemáticos se construyen a partir del conjunto ω de los números naturales, y del teorema se desprende (especialmente de la propiedad 4) que todas las construcciones matemáticas realizadas a partir de conjuntos regulares producen conjuntos regulares, por lo que los posibles conjuntos no regulares quedan simplemente fuera del "campo de trabajo" habitual de los matemáticos. La regularidad sólo es importante a la hora de probar teoremas generales sobre todos los conjuntos, cosa que prácticamente sólo hacen los especialistas en teoría de conjuntos. ∎

Todavía tenemos que extraer más consecuencias del axioma de regularidad, consecuencias que ayudan a acabar de entender qué estamos postulando exactamente con él, pues con lo visto hasta ahora todavía no puede decirse que esté totalmente claro. No obstante, muchas de ellas se enmarcan en un contexto más general que conviene desarrollar en una sección aparte:

11.4 Relaciones bien fundadas

Hasta ahora hemos demostrado distintas variantes de teoremas de inducción y de recursión en contextos diversos. Aquí vamos a ver que en ZF^* (o NBG^*) es posible enunciar teoremas generales de inducción y recursión de modo que cualquier otro es un caso particular de éstos. La idea fundamental es que en todo argumento de inducción o recursión subyace una relación bien fundada, en el sentido siguiente:

Definición 11.30 Una relación R está *bien fundada* en una clase A si

$$\bigwedge x(x \subset A \wedge x \neq \varnothing \to \bigvee y \in x \bigwedge z \in x\, \neg z\, R\, y).$$

En estas condiciones diremos que y es un *elemento R-minimal* de x.

Aquí entendemos que R es simplemente una clase $R \subset V \times V$ de pares ordenados. En ZF^* esta definición ha de entenderse como un esquema en el que $x \in A$ y $(x, y) \in R$ son dos fórmulas cualesquiera, tal vez con más variables libres.

En estos términos, el teorema 11.28 afirma que la relación de pertenencia $E = \{(x, y) \mid x \in y\}$ está bien fundada en la clase R de todos los conjuntos regulares. El axioma de regularidad equivale a que E esté bien fundada en la clase universal V. Más en general, E está bien fundada en una clase A si y sólo si A está bien fundada, en el sentido que ya teníamos definido.

Otro ejemplo importante de relación bien fundada es el orden estricto $<$ asociado a un buen orden $\leq$ en una clase A. En efecto, si $x \subset A$ es un conjunto no vacío, es claro que el mínimo de x es un minimal para la relación $<$ (no para el orden no estricto $\leq$).

Suponiendo AI, es fácil generalizar el teorema 11.21, de modo que si R está bien fundada en A no existen sucesiones $\{x_n\}_{n\in\omega}$ en A tales que

$$\bigwedge n \in \omega\ x_{n+1}\, R\, x_n,$$

y "casi" se puede demostrar el recíproco, pues para ello se necesita el axioma de elección, que presentaremos más adelante.

Notemos que en la definición de relación bien fundada en A sólo hemos exigido que los subconjuntos (no las subclases) de A tengan minimales, pues un cuantificador $\bigwedge X$ tendría sentido en NBG^*, pero no en ZF^*. Sin embargo, vamos a necesitar extender la propiedad a subclases arbitrarias, para lo cual tenemos que generalizar los resultados de la sección anterior y definir clausuras:

Definición 11.31 Una relación R es *conjuntista* en una clase A si para todo $x \in A$ la clase

$$A_x^R = \{y \in A \mid y\, R\, x\}$$

de los anteriores de x es un conjunto.

Diremos que una subclase $B \subset A$ es *R-A-transitiva* si

$$\bigwedge xy \in A(x\,R\,y \wedge y \in B \to x \in B).$$

Diremos que R es *clausurable* en A si para todo $x \in A$ existe un conjunto R-A-transitivo y tal que $A_x^R \subset y \subset A$.

Observaciones La relación de pertenencia E es conjuntista en toda clase A, pues $A_x^E = x \cap A$, que es un conjunto.

Por otra parte, toda clase A es R-A-transitiva, por lo que toda relación es clausurable en todo conjunto x (pues el propio x es un conjunto R-x-transitivo).

Las clases transitivas son precisamente las clases E-V-transitivas y el axioma CT afirma precisamente que E es clausurable en V. Toda relación clausurable en A es obviamente conjuntista en A. ∎

Ahora generalizamos el teorema 11.23:

Teorema 11.32 *Si R es una relación clausurable en una clase A, para todo $x \in A$, el conjunto A_x^R está contenido en un mínimo subconjunto R-A-transitivo de A (en el sentido de que está a su vez contenido en cualquier otro subconjunto R-A-transitivo de A que contenga a x).*

Demostración: Dado $x \in A$, sea $w \subset A$ un conjunto R-A-transitivo tal que $A_x^R \subset w$. Definimos

$$y = \{u \in w \mid \bigwedge z(A_x^R \subset z \subset A \wedge z \text{ es } R\text{-}A\text{-transitivo} \to u \in z)\}.$$

Veamos que y cumple lo pedido. Observemos que y es simplemente el conjunto de los elementos de A que están en todos los subconjuntos R-A-transitivos de A que contienen a A_x^R. Obviamente $A_x^R \subset y$.

Tomemos $v, u \in A$ tales que $v\,R\,u \wedge u \in y$. Si z cumple $A_x^R \subset z \subset A$ y es R-A-transitivo, entonces $u \in z$, luego $v \in z$ por la transitividad, luego $u \in y$. Esto prueba que y es R-A-transitivo, y la minimalidad es inmediata. ∎

Definición 11.33 Sea R una relación clausurable en una clase A. Para cada $x \in A$, definimos la *clausura* de un $x \in A$ como

$$\mathrm{cl}_A^R(x) \equiv y|(A_x^R \subset y \subset A \wedge \bigwedge z(A_x^R \subset z \subset A \wedge z \text{ es } R\text{-}A\text{-transitivo} \to y \subset z)).$$

También podemos generalizar el hecho de que el axioma de infinitud garantiza la existencia de clausuras:

Teorema 11.34 (AI) *Toda relación conjuntista en una clase A es clausurable en A.*

DEMOSTRACIÓN: Sea R una relación conjuntista en A. El teorema de recursión 3.26 nos da una aplicación $\mathrm{cl}_A^R(x)[\,] : \omega \longrightarrow \mathcal{P}A$ determinada por[11]

$$\mathrm{cl}_A^R(x)[0] = A_x^R \wedge \bigwedge n \in \omega\ \mathrm{cl}_A^R(x)[n+1] = \bigcup_{u \in \mathrm{cl}_A^R(x)[n]} A_u^R.$$

A su vez definimos $\mathrm{cl}_A^R(x) \equiv \bigcup_{n\in\omega} \mathrm{cl}_A^R(x)[n]$, que es un conjunto por el axioma de reemplazo (y AI, que nos asegura que ω es un conjunto). Se cumple claramente que $A_x^R \subset \mathrm{cl}_A^R(x) \subset A$. Se cumple que esta clausura es R-A-transitiva, pues si $v, u \in A$ cumplen $v\,R\,u \wedge u \in \mathrm{cl}_A^R(x)$, entonces existe un $n \in \omega$ tal que $u \in \mathrm{cl}_A^R(x)[n]$, luego $v \in A_u^R \subset \mathrm{cl}_A^R(x)[n+1] \subset \mathrm{cl}_A^R(x)$.

Ahora probamos que cumple algo más que lo que exige la definición de clausura: está contenida en toda clase R-A-transitiva $B \subset A$ (no necesariamente un conjunto) que contenga a A_x^R. En efecto, basta razonar por inducción sobre n que $\mathrm{cl}_A^R(x)[n] \subset B$. Para $n = 0$ es obvio y, si vale para n, entonces todo $v \in \mathrm{cl}_A^R(x)[n+1]$ cumple que existe un $u \in \mathrm{cl}_A^R(x)[n]$ tal que $v\,R\,u$. Por hipótesis de inducción $u \in B$ y por transitividad $v \in B$, luego $\mathrm{cl}_A^R(x)[n+1] \subset B$. ∎

El teorema siguiente recoge las propiedades básicas de las clausuras:

Teorema 11.35 *Sea R una relación clausurable en una clase A y sea $x \in A$. Entonces:*

1. $A_x^R \subset \mathrm{cl}_A^R(x) \subset A$.
2. $\mathrm{cl}_A^R(x)$ *es un conjunto R-A-transitivo.*
3. *Si $A_x^R \subset B \subset A$ es una clase R-A-transitiva, entonces* $\mathrm{cl}_A^R(x) \subset B$.
4. $\mathrm{cl}_A^R(x) = A_x^R \cup \bigcup_{y\in A_x^R} \mathrm{cl}_A^R(y)$.

DEMOSTRACIÓN: 1) y 2) son consecuencias inmediatas de la definición de clausura. Ésta sólo asegura que c) se cumple cuando B es un conjunto. Para probar 3) en general observamos que la construcción de la función $n \mapsto \mathrm{cl}_A^R(x)[n]$ definida en la prueba del teorema anterior no requiere el axioma de infinitud, y una simple inducción prueba que $\bigwedge n\ \mathrm{cl}_A^R(x)[n] \subset \mathrm{cl}_A^R(x)$. Ahora podemos definir

$$\mathrm{cl}_A^R(x)^* = \{u \in \mathrm{cl}_A^R(x) \mid \bigvee n \in \omega\ u \in \mathrm{cl}_A^R(x)[n]\} \subset \mathrm{cl}_A^R(x),$$

y el mismo argumento del teorema anterior prueba que $\mathrm{cl}_A^R(x)^*$ es un conjunto R-A-transitivo que contiene a A_x^R, luego por definición de clausura tiene que ser $\mathrm{cl}_A^R(x)^* = \mathrm{cl}_A^R(x)$. Ahora podemos emplear el argumento del teorema anterior para concluir que $\mathrm{cl}_A^R(x) \subset B$.

4) Si $y \in A_x^R$, entonces $A_y^R \subset \mathrm{cl}_A^R(x)[1] \subset \mathrm{cl}_A^R(x)$, luego por 2) y 3) obtenemos que $\mathrm{cl}_A^R(y) \subset \mathrm{cl}_A^R(x)$. Por consiguiente $a = A_x^R \cup \bigcup_{y\in A_x^R} \mathrm{cl}_A^R(y)$ está contenido en $\mathrm{cl}_A^R(x)$.

[11] Notemos que esta construcción requiere que la relación sea conjuntista para que podamos asegurar que cada término de la sucesión es un conjunto. Si no, la sucesión no estaría bien definida.

Para demostrar la otra inclusión basta probar a es transitivo y aplicar 3). Sean, pues, u, $v \in A$ tales que $u\,R\,v \wedge v \in a$. Si $v \in \mathrm{cl}_A^R(y)$ para un $y \in A_x^R$, entonces, por la transitividad de la clausura, $u \in \mathrm{cl}_A^R(y)$, luego $u \in a$. Si $v \in A_x^R$, entonces $u \in \mathrm{cl}_A^R(v) \subset a$. ∎

Ahora ya podemos extender a clases propias la condición que define las relaciones bien fundadas:

Teorema 11.36 *Si R es una relación clausurable y bien fundada en una clase A, entonces toda subclase de A no vacía tiene un elemento R-minimal.*

Demostración: Sea $B \subset A$ una clase no vacía. Tomamos un $x \in B$. Si ya es un R-minimal, no hay nada que probar. En caso contrario, $B \cap A_x^R \neq \varnothing$, luego también $B \cap \mathrm{cl}_A^R(x) \neq \varnothing$. Como esto es un subconjunto no vacío de A, tiene un R-minimal u, y resulta que u también es R-minimal de B, pues ciertamente $u \in B$ y si existiera $v \in B \cap A_u^R$, tendríamos que $v\,R\,u \in \mathrm{cl}_A^R(x)$, luego $v \in \mathrm{cl}_A^R(x)$, luego $v \in B \cap \mathrm{cl}_A^R(x)$, en contra de la minimalidad de u. ∎

Con esto ya podemos demostrar un principio general de inducción, aunque no es el más general que vamos a enunciar:

Teorema 11.37 (Teorema general de inducción transfinita) *Sea R una relación clausurable y bien fundada en una clase A y sea B una clase cualquiera. Entonces*

$$\bigwedge x \in A(A_x^R \subset B \to x \in B) \to A \subset B.$$

Demostración: Si no se cumple $A \subset B$, entonces $A \setminus B$ es una subclase no vacía de A, luego tiene un R-minimal x, de modo que $A_x^R \subset B$, pero $x \notin B$. ∎

Lo que afirma el teorema anterior es que para demostrar que todo elemento $x \in A$ tiene una propiedad (estar en B), podemos suponer como hipótesis de inducción que todos los elementos de A_x^R la tienen. Similarmente, el teorema siguiente afirma que para definir una función $F : A \longrightarrow B$, si en A tenemos definida una relación clausurable y bien fundada, podemos definir $F(x)$ suponiendo que F está ya definida sobre los elementos de A_x^R:

Teorema 11.38 (Teorema general de recursión transfinita) *Sea R una relación clausurable y bien fundada en una clase A y sea $G : V \longrightarrow B$ una aplicación arbitraria. Entonces existe una única función $F : A \longrightarrow B$ tal que*

$$\bigwedge x \in A\, F(x) = G(x, F|_{A_x^R}).$$

Demostración: Por abreviar, a lo largo de esta prueba, "transitivo" significará R-A-transitivo.

Si $d \subset A$ es un conjunto transitivo, diremos que $h : d \longrightarrow B$ es una *d-aproximación* si

$$\bigwedge x \in d\; h(x) = G(x, h|_{A_x^R}).$$

Para cada $x \in A$, definimos

$$\hat{x} = \{x\} \cup \mathrm{cl}_A^R(x).$$

Es claro que $\hat{x}$ es transitivo y $x \in \hat{x}$ (de hecho, es el menor conjunto transitivo que contiene a x). Dividimos la prueba en varios pasos:

1) *Si h es una d-aproximación y h' es una d'-aproximación, entonces se cumple $h|_{d\cap d'} = h'|_{d\cap d'}$. En particular, para cada conjunto transitivo $d \subset A$ existe a lo sumo una d-aproximación.*

Lo probamos por inducción en $d \cap d'$, es decir, vamos a probar que todo elemento de $d \cap d'$ está en $\{u \in d \cap d' \mid h(u) = h'(u)\}$. Para ello tomamos $x \in d \cap d'$ y suponemos que $h(u) = h'(u)$ siempre que $u \in (d \cap d')_x^R$. Ahora bien, es inmediato que $d \cap d'$ es transitivo, de donde se sigue que $(d \cap d')_x^R = A_x^R$. Por consiguiente tenemos que $h|_{A_x^R} = h'|_{A_x^R}$, luego

$$h(x) = G(x, h|_{A_x^R}) = G(x, h'|_{A_x^R}) = h'(x).$$

2) *Para todo $x \in A$ existe una $\hat{x}$-aproximación.*

Lo probamos por inducción sobre x, es decir, suponemos que para todo $u \in A_x^R$ existe una $\hat{u}$-aproximación. Por 1) es única, luego podemos definir $h_u \equiv h|h$ es una $\hat{u}$-aproximación. Definimos $h = \bigcup_{u\in A_x^R} h_u$. De nuevo por 1) tenemos que h es una función y su dominio es

$$\bigcup_{u\in A_x^R} \hat{u} = \bigcup_{u\in A_x^R} (\{u\} \cup \mathrm{cl}_A^R(u)) = A_x^R \cup \bigcup_{u\in A_x^R} \mathrm{cl}_A^R(u) = \mathrm{cl}_A^R(x),$$

donde hemos aplicado el teorema 11.35.

Si $v \in \mathrm{cl}_A^R(x)$, entonces $h(v) = h_u(v)$, para cierto $u \in A_x^R$ tal que $v \in \hat{u}$. Puesto que $h_u \subset h$ y $A_v^R \subset \hat{u}$ (por ser $\hat{u}$ transitivo) tenemos que $h_u|_{A_v^R} = h|_{A_v^R}$. Como h_u es una $\hat{u}$-aproximación,

$$h(v) = h_u(v) = G(v, h_u|_{A_v^R}) = G(v, h|_{A_v^R}),$$

con lo que h resulta ser una $\mathrm{cl}_A^R(x)$-aproximación.

Puede probarse que $x \notin \mathrm{cl}_A^R(x)$, pero no es necesario, en cualquier caso podemos definir

$$h' = h \cup \{(x, G(x, h|_{A_x^R}))\},$$

de modo que $h : \hat{x} \longrightarrow \mathrm{V}$ y es inmediato que para todo $v \in \hat{x}$ se cumple $h'|_{A_x^R} = h|_{A_x^R}$, de donde se sigue claramente que h' es una $\hat{x}$-aproximación.

3) *Definimos $F = \bigcup_{x\in A} h_x$, donde $h_x \equiv h|h$ es una $\hat{x}$-aproximación.*

La unicidad de 1) hace que $F : A \longrightarrow B$, y los mismos razonamientos que hemos aplicado a h en el paso anterior prueban que para todo $x \in A$ se cumple $F(x) = G(x, F|_{A_x^R})$.

4) *La unicidad de F se prueba igual que* 1) ∎

Como primera aplicación de este teorema, dada una clase con una relación clausurable y bien fundada, vamos a asociar a cada uno de sus elementos un ordinal que exprese su "altura" en la relación, entendiendo que un elemento es más alto cuantos más elementos tiene por debajo.

Definición 11.39 Sea R una relación clausurable y bien fundada en una clase A. Definimos $\mathrm{rang} : A \longrightarrow \Omega$ como la única aplicación que cumple

$$\bigwedge x \in A\ \mathrm{rang}_A^R(x) = \bigcup_{y \in A_x^R} (\mathrm{rang}_A^R(y) + 1),$$

donde estamos representando por $\alpha + 1$ el ordinal siguiente a α.

Observemos que hemos definido el rango de un elemento supuesto definido el rango de los elementos anteriores a él. Más concretamente, estamos aplicando el teorema anterior a la función $G : V \longrightarrow \Omega$ dada por

$$G(z) = \begin{cases} \bigcup\limits_{y \in A_x^R} (s(y) + 1) & \text{si } z = (x, s), \text{ con } x \in A \wedge s : A_x^R \longrightarrow \Omega, \\ 0 & \text{en otro caso.} \end{cases}$$

Recordemos que la unión de un conjunto de ordinales no es más que su supremo. Hemos de entender que el supremo del conjunto vacío es 0 (lo cual es cierto, pues 0 es la menor cota superior de $\varnothing$). De este modo, los minimales de A tienen todos rango 0 y, en general, el rango de un elemento es el mínimo ordinal estrictamente mayor que los rangos de todos sus anteriores.

Teorema 11.40 *Sea R una relación clausurable y bien fundada en una clase A. Sean x, $y \in A$. Si $x \in \mathrm{cl}_A^R(y)$, entonces* $\mathrm{rang}_A^R x < \mathrm{rang}_A^R y$.

Demostración: Por inducción sobre y, es decir, suponemos que el resultado es cierto para todo $u \in A_y^R$ y suponemos que $x \in \mathrm{cl}_A^R(y)$. Entonces hay dos posibilidades, o bien $x \in A_y^R$, en cuyo caso $\mathrm{rang}_A^R(x) < \mathrm{rang}_A^R(y)$ por definición de rango, o bien $x \in \mathrm{cl}_A^R(u)$, para cierto $u \in A_y^R$. Entonces aplicamos la hipótesis de inducción: $\mathrm{rang}_A^R(x) < \mathrm{rang}_A^R(u) < \mathrm{rang}_A^R(y)$. ∎

Con la ayuda del rango podemos demostrar teoremas de inducción y recursión aún más potentes. En el caso de la inducción, vamos a ver que podemos tomar como hipótesis de inducción, no ya que todos los elementos anteriores a uno dado cumplen lo que queremos probar, sino que todos los elementos de su clausura lo cumplen (o sea, los anteriores, y los anteriores de los anteriores, etc.).

Teorema 11.41 (Teorema general de inducción transfinita) *Sea R una relación clausurable y bien fundada sobre una clase A. Entonces*

$$\bigwedge x \in A(\mathrm{cl}_A^R(x) \subset B \to x \in B) \to A \subset B.$$

DEMOSTRACIÓN: Si no se da la inclusión podemos tomar un $x \in A \setminus B$ de rango mínimo. Si $u \in \mathrm{cl}_A^R(x)$, entonces $\mathrm{rang}_A^R u < \mathrm{rang}_A^R x$, luego por minimalidad $u \in B$. Pero entonces la hipótesis nos da que $x \in B$, lo cual es absurdo. ∎

Similarmente, para definir una función sobre x podemos suponer que está ya definida sobre $\mathrm{cl}_A^R(x)$:

Teorema 11.42 (Teorema general de recursión transfinita) *Sea R una relación clausurable y bien fundada en una clase A y sea $G : V \longrightarrow B$ una aplicación arbitraria. Entonces existe una única función $F : A \longrightarrow B$ tal que*

$$\bigwedge x \in A\, F(x) = G(x, F|_{\mathrm{cl}_A^R(x)}).$$

La prueba de este teorema es idéntica a la de 11.38, salvo que el paso 1) y la unicidad de F se demuestran usando la versión fuerte del teorema general de recursión transfinita en lugar de la débil.

Conviene probar un recíproco del teorema 11.40:

Teorema 11.43 *Sea R una relación clausurable y bien fundada en una clase A, sea $y \in A$ y sea $\alpha < \mathrm{rang}_A^R(y)$. Entonces existe un conjunto $x \in \mathrm{cl}_A^R(y)$ tal que $\mathrm{rang}_A^R(x) = \alpha$.*

DEMOSTRACIÓN: Supongamos que no existe tal x. Por definición del rango, existe un $x \in A_y^R$ tal que $\alpha < \mathrm{rang}_A^R(x) + 1$, luego $\alpha \leq \mathrm{rang}_A^R(x)$, pero como estamos suponiendo que α no es el rango de ningún $x \in \mathrm{cl}_A^R(y)$, tiene que ser $\alpha < \beta = \mathrm{rang}_A^R(x)$.

Tomemos el mínimo ordinal β tal que $\beta > \alpha$ y existe un $x \in \mathrm{cl}_A^R(y)$ de modo que $\beta = \mathrm{rang}_A^R(x) = \bigcup_{u \in A_x^R} (\mathrm{rang}_A^R(u) + 1)$.

Ahora bien, si $u \in A_x^R$, entonces $\mathrm{rang}_A^R(u) < \beta$, luego por la minimalidad de β tiene que ser $\mathrm{rang}_A^R(u) \leq \alpha$. Como $u \in \mathrm{cl}_A^R(y)$, la hipótesis sobre α implica que de hecho $\mathrm{rang}_A^R(u) < \alpha$. Así llegamos a que

$$\alpha < \beta = \bigcup_{u \in A_x^R} (\mathrm{rang}_A^R(u) + 1) \leq \alpha,$$

contradicción. ∎

Los teoremas que hemos probado se particularizan a teoremas de inducción y recursión sobre ordinales. Observemos que la relación de pertenencia E es clausurable y bien fundada en Ω. En efecto, como Ω es transitiva, las clases E-Ω-transitivas son simplemente las subclases transitivas de Ω y $\mathrm{cl}_\Omega^E(\alpha) = \mathrm{ct}\,\alpha = \alpha$. La buena fundación nos la da el teorema 3.14. Observemos que dicho teorema prueba que todo subconjunto de Ω no vacío tiene mínimo, pero el teorema 11.36 implica que lo mismo vale para clases arbitrarias (aunque es más simple incluso probarlo directamente).

En el caso de los teoremas de inducción, su prueba es tan directa que no merece la pena reducirla a los teoremas generales:

Teorema 11.44 (Inducción transfinita para ordinales) *Si $A \subset \Omega$ cumple una de las dos condiciones siguientes:*[12]

1. $\bigwedge\alpha(\alpha \subset A \to \alpha \in A)$, *o bien*
2. $0 \in A \wedge \bigwedge\alpha \in A\, \alpha + 1 \in A \wedge \bigwedge\lambda(\bigwedge\delta < \lambda\, \delta \in A \to \lambda \in A)$,

entonces $A = \Omega$.

Demostración: Si $A \neq \Omega$, sea α el mínimo de $\Omega \setminus A$. Entonces $\alpha \subset A$, lo que nos da una contradicción con a). Si suponemos b), distinguimos tres casos: no puede ser $\alpha = 0$, pues entonces $\alpha \in A$, no puede ser $\alpha = \beta + 1$, pues entonces $\beta \in A$ por la minimalidad de α, luego $\alpha \in A$ por b), ni α puede ser un límite, pues entonces $\bigwedge\delta < \alpha\ \delta \in A$, luego $\alpha \in A$, y tenemos también una contradicción. ∎

Esto quiere decir que para demostrar que todo ordinal tiene una propiedad podemos suponer como hipótesis de inducción que la tienen todos los ordinales menores que uno dado y demostrar que éste también la tiene, o bien distinguir tres casos: probar que 0 la tiene, probar que si α la tiene también la tiene $\alpha+1$, y probar que, bajo la hipótesis de inducción de que todo ordinal menor que un límite λ la tiene, entonces λ también la tiene.

Menos obvio es el principio de recursión transfinita para ordinales:

Teorema 11.45 (Recursión transfinita para ordinales) *Dada cualquier función $G : V \longrightarrow A$, existe una única función $F : \Omega \longrightarrow A$ tal que*

$$\bigwedge\alpha\ F(\alpha) = G(F|_\alpha).$$

Esto es consecuencia directa del teorema 11.38, teniendo en cuenta que en este caso $\Omega^E_\alpha = \alpha$. Esto significa que para definir una función sobre Ω podemos definir $F(\alpha)$ supuesto que F está definida sobre los ordinales menores que α.

Ejemplo: Suma de ordinales Como aplicación inmediata del teorema anterior, para cada ordinal α podemos definir una aplicación $\alpha+ : \Omega \longrightarrow \Omega$ como la única que cumple

$$\alpha + 0 = \alpha \wedge \bigwedge\beta\ (\alpha + \beta') = (\alpha + \beta)' \wedge \bigwedge\lambda\ \alpha + \lambda = \bigcup_{\delta<\lambda} (\alpha + \delta).$$

Se trata de una aplicación directa del teorema anterior, donde $G : V \longrightarrow \Omega$ es la aplicación dada por

$$G(s) = \begin{cases} s(\beta)' & \text{si } s : \beta' \longrightarrow \Omega, \text{ para un } \beta \in \Omega, \\ \bigcup \mathcal{R}s & \text{si } s : \lambda \longrightarrow \Omega, \text{ y } \lambda \text{ es un ordinal límite,} \\ \alpha & \text{en otro caso.} \end{cases}$$

[12] Adoptamos el convenio de que $\bigwedge\alpha$ significa $\bigwedge\alpha \in \Omega$ y que $\bigwedge\lambda$ significa "para todo ordinal límite λ".

Así está definida la suma $\alpha+\beta$ de dos ordinales cualesquiera, y es inmediato que sobre los números naturales coincide con la suma usual. Además tenemos que $\alpha+1=\alpha+0'=(\alpha+0)'=\alpha'$, por lo que habitualmente escribiremos $\alpha+1$ en lugar de α', como ya hemos venido haciendo en algunas ocasiones.

Más aún, es claro que $\alpha+2=\alpha+1'=(\alpha+1)'=\alpha''$, y así sucesivamente, por lo que los ordinales inmediatamente posteriores a un ordinal α cualquiera son

$$\alpha<\alpha+1<\alpha+2<\alpha+3<\cdots$$

Si suponemos AI, podemos calcular $\alpha+\omega=\bigcup_{n\in\omega}(\alpha+n)$, que es claramente un ordinal límite, pues si $\delta<\alpha+\omega$ entonces existe un $n<\omega$ tal que $\delta\in\alpha+n$, luego $\delta'<(\delta+n)'=\delta+n'<\alpha+\omega$. De hecho, $\alpha+\omega$ es claramente el menor ordinal límite mayor que α.

En particular esto implica (siempre suponiendo AI) que la clase Λ de los ordinales límite no es un conjunto, pues si lo fuera tendría por supremo a su unión $\alpha=\bigcup\Lambda\in\Omega$, y $\alpha+\omega$ sería un ordinal límite mayor que todos los ordinales límite, lo cual es absurdo. ∎

Análogamente se puede definir de forma natural el producto y la exponenciación de ordinales, pero no vamos a estudiar aquí la aritmética ordinal. Terminamos esta sección con una última aplicación:

Teorema 11.46 *Si $(x,\leq)$ es un conjunto bien ordenado, existe un ordinal α y una biyección $f:x\longrightarrow\alpha$ que conserva el orden.*

Demostración: La relación R determinada por el orden estricto $<$ es trivialmente clausurable en x (porque es un conjunto) y para todo $a\in x$ se cumple que $x_a^<=\mathrm{cl}_x^<(a)$ es el conjunto de los elementos menores que a. Podemos considerar la aplicación $\mathrm{rang}_x^<:x\longrightarrow\Omega$. Por 11.40 tenemos que si $a<b\to\mathrm{rang}_x^<(a)<\mathrm{rang}_x^<(b)$, por lo que el rango es una aplicación inyectiva que conserva el orden. Por 11.43 sabemos que su imagen es un subconjunto transitivo de Ω, luego es un cierto ordinal α. En definitiva: $\mathrm{rang}_x^<:(x,\leq)\longrightarrow\alpha$ es una biyección que conserva el orden. ∎

11.5 El axioma de regularidad II

Con los resultados de la sección anterior ya podemos formarnos una idea clara de lo que es la clase R de los conjuntos regulares. Puesto que la relación de pertenencia E es clausurable y bien fundada en R, tenemos definido un rango sobre R:

Definición 11.47 Llamaremos *rango* de un conjunto regular al ordinal que le asigna la aplicación $\mathrm{rang}:R\longrightarrow\Omega$ determinada por

$$\mathrm{rang}\,x=\bigcup_{y\in x}(\mathrm{rang}\,y+1).$$

Para cada ordinal $\alpha \in \Omega$, definimos la clase

$$R_\alpha = \{x \in R \mid \operatorname{rang} x < \alpha\}.$$

El teorema siguiente nos muestra finalmente qué es R:

Teorema 11.48 *Se cumple:*

$$R_0 = \varnothing \wedge \bigwedge\alpha\ R_{\alpha+1} = \mathcal{P}R_\alpha \wedge \bigwedge\lambda\ R_\lambda = \bigcup_{\delta<\lambda} R_\delta,$$

$$R = \bigcup_{\alpha\in\Omega} R_\alpha.$$

Demostración: La única igualdad que no es trivial es $R_{\alpha+1} = \mathcal{P}R_\alpha$. Si $x \in \mathcal{P}R_\alpha$, entonces $x \subset R_\alpha$, luego

$$\operatorname{rang} x = \bigcup_{y\in x} (\operatorname{rang} y + 1) \leq \alpha < \alpha + 1,$$

luego $x \in R_{\alpha+1}$. Recíprocamente, si $x \in R_{\alpha+1}$, entonces todo $y \in x$ cumple $\operatorname{rang} y + 1 \leq \operatorname{rang} x < \alpha + 1$, luego $\operatorname{rang} y < \alpha$, luego $y \in R_\alpha$. Así pues, $x \subset R_\alpha$. ∎

Así pues, los conjuntos regulares son los que se obtienen a partir del conjunto vacío por sucesivas aplicaciones del operador $\mathcal{P}$. Las clases R_α forman una jerarquía cuyos primeros niveles son:

$$R_0 = \varnothing, \quad R_1 = \{\varnothing\}, \quad R_2 = \{\varnothing, \{\varnothing\}\}, \quad R_3 = \{\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \{\varnothing, \{\varnothing\}\}\} \ldots$$

El teorema siguiente recoge sus propiedades básicas: forman una sucesión transfinita creciente de clases transitivas y en cada nivel aparecen nuevos conjuntos que no están en los anteriores.

Teorema 11.49 *Se cumple:*

1. *Si $\alpha \leq \beta$ son ordinales, entonces $R_\alpha \subset R_\beta$.*
2. *Para cada ordinal α, la clase R_α es transitiva.*
3. *Para cada ordinal α, se cumple* $\operatorname{rang}\alpha = \alpha$*, luego $R_\alpha \cap \Omega = \alpha$.*

Demostración: a) es trivial.

b) Si $y \in x \in R_\alpha$, entonces $\operatorname{rang} y < \operatorname{rang} x < \alpha$, luego $y \in R_\alpha$.

c) Por inducción sobre α: si suponemos que $\operatorname{rang}\beta = \beta$ para todo $\beta < \alpha$, entonces

$$\operatorname{rang}\alpha = \bigcup_{\beta<\alpha} (\operatorname{rang}\beta + 1) = \bigcup_{\beta<\alpha} (\beta + 1) = \alpha.$$

∎

Ahora podemos ver mas claramente que todos los conjuntos de interés para los matemáticos son regulares. Por ejemplo, un par ordenado (m, n) de números

naturales está en R_ω, un número entero z se define (suponiendo AI) como un conjunto de pares de números naturales, luego $z \subset \mathrm{R}_\omega$, con lo que $z \in \mathrm{R}_{\omega+1}$ y, por consiguiente, $\mathbb{Z} \in \mathrm{R}_{\omega+2}$. Similarmente se comprueba que $\mathbb{Q} \in \mathrm{R}_{\omega+5}$. Si (suponiendo AP) construimos $\mathbb{R}$ por el método de Dedekind (con lo que un número real es un subconjunto de $\mathbb{Q}$), tenemos $\mathbb{R} \in \mathrm{R}_{\omega+7}$, etc. Así podemos ir situando en la jerarquía regular todas las construcciones matemáticas usuales.

Teniendo en cuenta que la clase de partes de un conjunto finito es también un conjunto finito, una simple inducción prueba que, para todo $n \in \omega$, la clase R_n es un conjunto finito. En ZF* podemos definir $R_\omega = \bigcup_{n\in\omega} R_n$, pero sin AI no podemos demostrar que $R_\omega \neq R$. Suponiendo AI (usando además los axiomas de reemplazo y de la unión) concluimos que R_ω es un conjunto. Entonces podemos definir $R_{\omega+1}$, pero sin AP no podemos demostrar que sea un conjunto, pues si lo fuera sería no numerable, y no se puede demostrar la existencia de conjuntos no numerables sin AP. Suponiendo AP, una simple inducción transfinita demuestra que todas las clases R_α son conjuntos.

Nota Si suponemos AP, podemos tomar el teorema 11.48 como definición de la sucesión transfinita $\{R_\alpha\}_{\alpha\in\Omega}$, como aplicación del teorema 11.45, podemos definir los conjuntos regulares como los elementos de $R = \bigcup_\alpha R_\alpha$ y definir el rango de un conjunto regular x como el mínimo ordinal α tal que $x \subset R_\alpha$. ∎

Ejercicio: Probar que en ZF* + regularidad + CT + $\omega = \Omega$ se prueba que todo conjunto es finito. Por consiguiente, en ZF* + regularidad + CT el axioma de infinitud es equivalente a la existencia de un conjunto infinito.

11.6 El axioma de elección

Nos ocupamos ahora del último de los axiomas usuales de la teoría de conjuntos:

Definición 11.50 El *axioma de elección* es la sentencia

$$\mathrm{AE} \equiv \bigwedge x \bigvee f(f \text{ es una función} \wedge \mathcal{D}f = x \wedge \bigwedge u \in x(u \neq \varnothing \to f(u) \in u)).$$

Así, AE afirma que, dado cualquier conjunto x, existe una función que a cada elemento $u \in x$ no vacío le elige uno de sus elementos. A una función de estas características se la llama una *función de elección* sobre x.

Observaciones Psicológicamente, el axioma de elección suele suscitar dos reacciones opuestas: la de quienes lo consideran "evidente" y se sorprenden de que no pueda ser demostrado a partir de los demás axiomas, y la de quienes no lo consideran evidente en absoluto y, o bien no lo aceptan, o bien lo aceptan "por conveniencia", pero "sin convicción". Dejando de lado las motivaciones más pasionales que racionales, un hecho objetivo es que hay muchos matemáticos

profesionales que vacilan a la hora de decidir si una determinada prueba usa o no el axioma de elección. Los que lo consideran "evidente" tienden a creer que no se usa cuando en realidad se usa, y los que recelan de él tienden a creer que se usa cuando en realidad no se usa.

Por una parte, hay que tener presente que no todas las elecciones requieren el axioma de elección. Por ejemplo, admitiendo AI y AP, podemos demostrar que el conjunto $\mathcal{P}\omega$ tiene una función de elección, a saber, la función

$$f \equiv \{(u,n) \in \mathcal{P}\omega \times \omega \mid (u = \varnothing \wedge n = 0) \vee (u \neq \varnothing \wedge n = \text{mín } u)\}$$

que a cada subconjunto no vacío de ω le asigna su mínimo respecto a la relación de orden usual en ω (cuya existencia hemos demostrado sin necesidad del axioma de elección).

La construcción anterior es un caso particular de un resultado más general (que no requiere AE): si el conjunto $\bigcup x$ admite un buen orden, entonces x tiene una función de elección. La demostración es la misma: para cada $u \in x$ no vacío, podemos definir $f(u) = \text{mín}\, u$.

Así pues, el axioma de elección sólo es necesario cuando no tenemos ningún criterio concreto que nos permita definir una función de elección. El problema es que "$f(u) =$ un elemento cualquiera de u" no es una fórmula de $\mathcal{L}_{\text{tc}}$ que pueda usarse para especificar una función $f \subset x \times \bigcup x$.

Consideremos el caso más simple no trivial: tomemos como hipótesis (bajo AI) que $f : \omega \longrightarrow x$ biyectiva, de modo que, llamando $x_i = f(i)$, tenemos un conjunto numerable $x = \{x_i \mid i \in \omega\}$, y admitamos también que

$$\bigwedge i \in \omega \bigvee ab\ x_i = \{a,b\},$$

es decir, que cada conjunto x_i tiene a lo sumo dos elementos. ¿Podemos asegurar bajo estas hipótesis que el conjunto x tiene una función de elección sin recurrir a AE? La respuesta es negativa. Si ZF es consistente (es decir, la teoría que cuenta con todos los axiomas que hemos introducido hasta ahora menos AE), entonces ZF tiene modelos en los que existen familias numerables de conjuntos de dos elementos sin funciones de elección.

Si el lector es propenso a tener por "evidente" el axioma de elección, podrá argumentar que en cualquier modelo en el que consideremos una familia así podemos sin duda considerar colecciones de pares ordenados formados por un x_i para cada i y uno de los dos elementos de x_i. Si esto es cierto o no, es una cuestión filosófica en la que no vamos a entrar aquí, pero lo que es innegable es que, aunque admitamos que existen tales colecciones de pares ordenados, una cosa es que existan y otra que sean necesariamente la extensión de un conjunto. Aunque "creamos" que existen, lo que es objetivamente innegable es que no tienen por qué ser la extensión de un conjunto, porque ninguno de los axiomas de la teoría de conjuntos fuerza a que lo sean.

Es esta faceta del axioma de elección como axioma que postula la existencia de conjuntos que no pueden ser construidos explícitamente la que motiva los

recelos contra él. Ahora bien, tales recelos pueden conducir a errores de bulto. Uno de los más groseros consiste en creer que si en el curso de una demostración contamos, por ejemplo, con $x \neq \varnothing$ y decimos "sea u un elemento de x", en ese punto estamos usando el axioma de elección, porque "elegimos" un elemento de x sin dar ningún criterio sobre cómo lo hacemos.

En realidad ahí no hay ningún uso del axioma de elección. Simplemente tenemos que $\bigvee u\, u \in x$ y aplicamos la regla de inferencia de eliminación del particularizador, nada que tenga que ver con un axioma de la teoría de conjuntos. Si podemos demostrar que existen funciones de elección para un conjunto dado, no estamos usando el axioma de elección al tomar una en concreto. El problema se da cuando no podemos demostrar que existan tales funciones de elección. El axioma de elección no sirve para eliminar particularizadores (para eso ya está la regla EP), sino que proporciona fórmulas de tipo $\bigvee f(\cdots)$, a las que luego les podemos eliminar el particularizador. ∎

Ejercicio: Demostrar en ZF^* que todo conjunto finito tiene una función de elección.

Vamos a ver algunos usos "auténticos" del axioma de elección demostrando algunos resultados que de hecho teníamos pendientes:

Teorema 11.51 (AE) *Un conjunto es D-infinito si y sólo si es infinito.*

Demostración: Una implicación es el teorema 6.30 y no requiere AE. Falta probar que todo conjunto infinito x es D-infinito. Notemos que aunque partamos de que x es infinito, no por ello contamos con el axioma de infinitud, pues en principio éste equivale a la existencia de un conjunto infinito.

Si suponemos AP podemos tomar una función de elección $f : \mathcal{P}x \longrightarrow x$, aunque sin AP podemos probar la existencia del conjunto $\mathcal{P}^{\mathrm{f}}x$ de los subconjuntos finitos de x y, por reemplazo aplicado a la fórmula $y = x \setminus u$ obtenemos la existencia del conjunto $\mathcal{P}^{\mathrm{cf}}x$ de los subconjuntos de x de complementario finito. Es suficiente tomar una función de elección $f : \mathcal{P}^{\mathrm{cf}}x \longrightarrow x$. Ésta nos permite definir por recurrencia[13] la función $H : \omega \longrightarrow x$ que cumple

$$H(n) = f(x \setminus H[n]).$$

Notemos que $H[n]$ es un conjunto finito, por ser imagen de un conjunto finito, luego $H(n) \in x \setminus H[n]$, es decir, que $H(n)$ es un elemento de x distinto de todos $H(i)$ con $i < n$. Esto implica que H es inyectiva y el teorema 11.11 nos da que x es D-infinito. ∎

Teorema 11.52 (AI, AE) *Una relación R está bien fundada en una clase A si y sólo si no existe ninguna sucesión $\{x_n\}_{n\in\omega}$ de elementos de A tal que $\bigwedge n \in \omega\ x_{n+1}\, R\, x_n$.*

[13] Aplicamos el teorema 11.38 a la clase ω con la relación de pertenencia E y la función $G(n, s) = f(x \setminus \mathcal{R}s)$. Notemos que $\omega_n^< = n$, con lo que $G(n, H|n) = f(x \setminus H[n])$. Notemos que no estamos usando para nada el axioma de infinitud.

DEMOSTRACIÓN: Una implicación es inmediata y no requiere AE: si existe tal sucesión, entonces su rango es un conjunto sin minimal, luego R no está bien fundada. Supongamos ahora que R no está bien fundada y sea $y \subset A$ un conjunto no vacío sin R-minimal. Esto significa que $\bigwedge u \in y \bigvee v \in y\ v\,R\,u$. Consideremos el conjunto

$$a = \{A_u^R \cap y \mid u \in y\},$$

que existe por reemplazo aplicado a la fórmula $z = A_u^R \cap y$ y, según lo que estamos suponiendo, es una familia de conjuntos no vacíos. Sea $f : a \longrightarrow V$ una función de elección y sea $g : y \longrightarrow y$ la función dada por $g(u) = f(A_u^R \cap y)$. De este modo se cumple que $\bigwedge u \in y\,(g(u) \in y \wedge g(u)\,R\,u)$.

Ahora fijamos un $u_0 \in x$ y definimos por recurrencia una función $x : \omega \longrightarrow y$ mediante $x_0 = u_0 \wedge x_{i+1} = g(x_i)$. Es claro que la sucesión $\{x_i\}_{i\in\omega}$ cumple lo requerido. ∎

Teorema 11.53 (AE) *Sean x e y dos conjuntos no vacíos. Existe $f : x \longrightarrow y$ inyectiva si y sólo si existe $g : y \longrightarrow x$ suprayectiva.*

DEMOSTRACIÓN: Una implicación es 6.29 2) y no requiere el axioma de elección. La implicación contraria es 6.29 3), sólo que allí añadíamos la hipótesis de que y admita un buen orden precisamente para evitar el uso del axioma de elección. Ahora podemos considerar el conjunto

$$a = \{g^{-1}[u] \mid u \in x\},$$

que existe por reemplazo, tomar una función de elección $h : a \longrightarrow y$ y definir f mediante $f(u) = h(g^{-1}[u])$. Notemos que se sigue cumpliendo $f \circ g = i_x$. ∎

Una parte de las demostraciones anteriores ha sido construir el conjunto concreto sobre el que necesitábamos la función de elección. Normalmente esto requiere algunas manipulaciones conjuntistas que suelen omitirse por rutinarias. El teorema siguiente muestra algunas formas equivalentes del axioma de elección que sólo se diferencian superficialmente en la forma de presentar la familia de conjuntos sobre la que hay que realizar la elección y la forma de recopilar los conjuntos elegidos:

Teorema 11.54 *Las afirmaciones siguientes son equivalentes:*

1. AE
2. *Si $g : y \longrightarrow x$ es una aplicación suprayectiva, existe $f : x \longrightarrow y$ tal que $f \circ g = i_x$.*
3. *Para toda familia $\{x_i\}_{i\in a}$ de conjuntos no vacíos (donde a es un conjunto) existe otra familia $\{s_i\}_{i\in a}$ tal que $\bigwedge i \in a\ s_i \in x_i$.*
4. *Para todo conjunto x formado por conjuntos no vacíos disjuntos dos a dos, existe un conjunto $a \subset \bigcup x$ tal que $\bigwedge u \in x \bigvee v\ u \cap a = \{v\}$.*

DEMOSTRACIÓN: 1) ⇒ 2) se sigue de la demostración del teorema anterior.

2) ⇒ 3) Consideremos la aplicación $g : \bigcup_{i\in a}\{i\}\times x_i \longrightarrow a$ dada por $g(i,u) = i$.

Como los conjuntos x_i son no vacíos, tenemos que g es suprayectiva. Sea $f : a \longrightarrow \bigcup_{i\in a}\{i\} \times x_i$ según b), es decir, tal que, para cada $i \in a$, se cumple que $f(i) = (i, v)$, para cierto $v \in x_i$. Basta tomar $s = \mathcal{R}f$.

3) ⇒ 4) Podemos ver a x como una familia $\{i\}_{i\in x}$ de conjuntos no vacíos. Por c) existe $\{s_i\}_{i\in x}$ tal que $\bigwedge i \in x\ s_i \in i$. Basta tomar $a = \mathcal{R}s$.

4) ⇒ 1) Dado un conjunto x, no perdemos generalidad si suponemos que no contiene a $\varnothing$. El conjunto $x' = \{\{i\} \times i \mid i \in x\}$ (existe por reemplazo) está formado por conjuntos no vacíos disjuntos dos a dos. Por d) existe un conjunto f que contiene exactamente un elemento de cada uno de ellos. Claramente f es una función de elección sobre x. ∎

Nota La familia $\{s_i\}_{i\in a}$ no es más que un elemento del producto cartesiano

$$\prod_{i\in a} x_i \equiv \{s \mid s : a \longrightarrow \bigcup_{i\in a} x_i \wedge \bigwedge i \in a\ s_i \in x_i\},$$

por lo que AE resulta ser equivalente a que el producto cartesiano de una familia de conjuntos no vacíos es no vacío.[14] ∎

Veamos una última consecuencia sencilla del axioma de elección:

Teorema 11.55 (AI, AP, AE) *Toda unión numerable de conjuntos numerables es numerable.*

DEMOSTRACIÓN: El enunciado ha de entenderse como que tenemos una familia $\{x_i\}_{i\in\omega}$ numerable de conjuntos tales que

$$\bigwedge i \in \omega \bigvee f\ f : \omega \longrightarrow x_i \text{ biyectiva.}$$

El problema es que necesitamos elegir una biyección f_i para cada cada i. Para ello necesitamos AP para considerar los conjuntos x_i^ω de aplicaciones $\omega \longrightarrow x_i$, del cual podemos especificar el subconjunto s_i de aplicaciones suprayectivas $\omega \longrightarrow x_i$ y formar a su vez la familia $\{s_i\}_{i\in\omega}$. Estamos suponiendo que todos los conjuntos s_i son no vacíos, luego por AE existe $\{f_i\}_{i\in\omega}$ de modo que $f_i : \omega \longrightarrow x_i$ suprayectiva.

A partir de aquí es fácil concluir: definimos $g : \omega \times \omega \longrightarrow \bigcup_{i\in\omega} x_i$ mediante $g(i, j) = f_i(j)$. Claramente g es suprayectiva y, como $\omega \times \omega$ es numerable, también lo es la unión. ∎

El axioma de elección interviene de forma esencial en la demostración de numerosos teoremas importantes del álgebra, el análisis o la topología (para

[14] Sin el axioma de partes no podemos demostrar que el producto cartesiano de una familia de conjuntos sea un conjunto, pero aquí no necesitamos que lo sea.

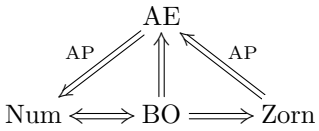
probar la existencia de base en todo espacio vectorial, la existencia de ideales maximales en anillos unitarios, la existencia de clausuras algebraicas, el teorema de Tychonoff, el teorema de Hann-Banach, etc.) En la prueba de estos resultados y otros muchos, es mucho más práctico utilizar una forma equivalente, un tanto técnica, conocida como lema de Zorn:

Recordemos que un conjunto parcialmente ordenado es un par $(x, \leq)$, donde $\leq$ es una relación reflexiva, antisimétrica y transitiva (pero no necesariamente conexa). Una cadena en x es un subconjunto $c \subset x$ para el que se cumpla $\bigwedge uv \in c(u \leq v \vee v \leq u)$. Una cota superior para un conjunto $c \subset x$ es un $u \in x$ tal que $\bigwedge v \in c\ v \leq u$. Un elemento maximal en x es un $m \in x$ tal que $\neg\bigvee u \in x\ m < u$. El teorema siguiente contiene el enunciado del lema de Zorn junto con otras afirmaciones equivalentes menos técnicas:

Teorema 11.56 (AP) *Las afirmaciones siguientes son equivalentes:*

1. **Axioma de elección** *Todo conjunto tiene una función de elección.*
2. **Principio de numerabilidad** *Todo conjunto puede biyectarse con un ordinal.*
3. **Principio de buena ordenación** *Todo conjunto puede ser bien ordenado.*
4. **Lema de Zorn** *Todo conjunto parcialmente ordenado no vacío en el que toda cadena tenga una cota superior tiene un elemento maximal.*
5. **Lema de Zorn (variante)** *En todo conjunto parcialmente ordenado no vacío en el que toda cadena tenga una cota superior, cada elemento está por debajo de un elemento maximal.*

Demostración: Vamos a probar las implicaciones siguientes:



de modo que sólo las señaladas requieren el axioma de partes. Vemos así que, en ausencia de AP, el principio de buena ordenación o el principio de numerabilidad son preferibles como axiomas al axioma de elección, pues implican éste y todas las otras consecuencias.

1) $\Rightarrow$ 2) **(AP)** Supongamos que un conjunto x no puede biyectarse con ningún ordinal. En particular tenemos que $x \neq \varnothing$. Fijemos una función de elección $f : \mathcal{P}x \longrightarrow x$ y consideremos la función $G : V \longrightarrow V$ definida por $G(s) = f(x \setminus \mathcal{R}s)$. El teorema 11.45 nos da una función $F : \Omega \longrightarrow x$ tal que

$$\bigwedge \alpha \in \Omega\ F(\alpha) = f(x \setminus F[\alpha]).$$

Veamos por inducción sobre α que $F|_\alpha : \alpha \longrightarrow x$ inyectiva.

Si $\alpha = 0$ es trivial. Si es cierto para α, entonces $F[\alpha] \neq x$, porque estamos suponiendo que x no puede biyectarse con un ordinal. Entonces, puesto que $x \setminus F[\alpha] \neq \varnothing$, tenemos que $F(\alpha) = f(x \setminus F[\alpha]) \in x \setminus F[\alpha]$, de donde se sigue claramente que $F|_{\alpha+1}$ es inyectiva.

Si λ es un ordinal límite y $F|_\alpha$ es inyectiva para todo $\alpha < \lambda$, entonces es claro que $F|_\lambda$ es inyectiva, pues si $\delta < \epsilon < \lambda$, también $\delta < \epsilon < \epsilon + 1 < \lambda$, y la inyectividad de $F|_{\epsilon+1}$ implica que $F(\delta) \neq F(\epsilon)$.

A su vez esto implica que $F : \Omega \longrightarrow x$ inyectiva, pero esto es imposible, pues entonces $F[\Omega] \subset x$ sería un conjunto y por 10.13 aplicado a F^{-1} también lo sería Ω.

2) $\Rightarrow$ 3) es inmediato: dado un conjunto x, tomamos un ordinal α y una biyección $f : x \longrightarrow \alpha$ y definimos la relación en x dada por $u \leq v \leftrightarrow f(u) \leq f(v)$. Es inmediato comprobar que se trata de un buen orden en x.

3) $\Rightarrow$ 2) Es consecuencia inmediata del teorema 11.46.

3) $\Rightarrow$ 1) Es inmediato, pues ya hemos observado que un buen orden en $\bigcup x$ permite definir una función de elección sobre x.

3) $\Rightarrow$ 5) Sea $(x, \leq)$ un conjunto en las hipótesis del lema de Zorn y fijemos un $u_0 \in x$. Hemos de encontrar un elemento maximal $m \in x$ tal que $u_0 \leq m$. Para ello suponemos que no existe tal elemento maximal, es decir, que si $u_0 \leq v$, siempre existe un $v' \in x$ tal que $v < v'$.

Como consecuencia, si $c \subset x$ es una cadena tal que $u_0 \in c$, existe un $v \in x$ tal que $\bigwedge u \in c\ u < v$. En efecto, estamos suponiendo que la cadena tiene cota superior, es decir, que existe un $v \in x$ tal que $\bigwedge u \in c\ u \leq v$. En particular, $u_0 \leq v$, luego, según acabamos de indicar, existe un $v' \in x$ tal que $v < v'$, y este v' cumple lo pedido.

De acuerdo con c), fijamos un buen orden $(x, \trianglelefteq)$ en el conjunto x. Consideramos la función $G : V \longrightarrow x$ dada por $G(s) = v$ si y sólo si $\mathcal{R}s$ es una cadena en x que contiene a u_0 y entonces v es el mínimo respecto de la relación $\trianglelefteq$ del conjunto $\{v \in x \mid \bigwedge u \in \mathcal{R}s\ u < v\}$, o bien $v = u_0$ en cualquier otro caso.[15]

El teorema 11.45 nos da una función $F : \Omega \longrightarrow x$ determinada por la condición $F(\alpha) = G(F|_\alpha)$. Como $\mathcal{R}F|_0 = \varnothing$ no contiene a u_0, la definición de G nos da que $F(0) = u_0$.

Veamos por inducción sobre α que $\bigwedge \delta < \alpha\ F(\delta) < F(\alpha)$.

Suponemos que el resultado es cierto para todo $\alpha < \beta$, y podemos suponer que $\beta > 0$, pues en caso contrario no hay nada que probar. Tenemos, pues, que si $\delta < \alpha < \beta$, entonces $F(\delta) < F(\alpha)$, lo que implica que $\mathcal{R}(F|_\beta) = F[\beta]$ es una cadena en x que contiene a $F(0) = u_0$. Por lo tanto, por definición de G, tenemos que $F(\beta)$ cumple $\bigwedge u \in F[\beta]\ u < F(\beta)$, pero esto es justo lo que teníamos que probar.

[15] Notemos que G es una clase (bien) definida por una fórmula concreta, que tiene libres las variables, s, v, x, $\leq$, $\trianglelefteq$, u_0.

Consecuentemente tenemos que $F : \Omega \longrightarrow x$ inyectiva, pero eso es imposible, porque entonces $F[\Omega] \subset x$ sería un conjunto y por reemplazo (por 10.13 aplicado a F^{-1}) Ω también.

5) $\Rightarrow$ 4) es trivial.

4) $\Rightarrow$ 1) **(AP)** Fijemos un conjunto x, que podemos suponer no vacío y tal que $\varnothing \notin x$, y sea

$$y = \{p \in \mathcal{P}(x \times \bigcup x) \mid \bigvee a(a \subset x \wedge p : a \longrightarrow \bigcup x \wedge \bigwedge u \in a(u \neq \varnothing \to p(u) \in u))\}$$

el conjunto de las funciones de elección sobre subconjuntos de x. Notemos que $\varnothing \in y$, luego $y \neq \varnothing$. Consideramos en y el orden parcial dado por la inclusión. Es claro que si $c \subset y$ es una cadena, entonces tiene por cota superior a $\bigcup c$, luego el lema de Zorn nos da un $f : a \longrightarrow \bigcup x$ en y maximal respecto de la inclusión. Basta probar que $a = x$, pues entonces f es una función de elección sobre x.

En caso contrario, tomamos $u \in x \backslash a$ y $v \in u$ (lo cual es posible, pues estamos suponiendo que $\varnothing \notin x$). Es claro entonces que $f \cup \{(u, v)\} \in y$ y contradice la maximalidad de f. Así pues, $a = x$. ∎

Observaciones El principio de numerabilidad afirma que, del mismo modo que los números naturales pueden usarse para contar los conjuntos finitos y numerables, los ordinales bastan para contar cualquier conjunto, es decir, que todo conjunto se puede poner en la forma $\{x_\alpha\}_{\alpha<\beta}$, para cierto ordinal β.

Ahora es inmediato que los axiomas "todo conjunto es finito" o "todo conjunto es numerable" implican el axioma de elección, pues implican el principio de numerabilidad.

Este principio nos da una demostración más sencilla del teorema 11.51: si un conjunto x es infinito, tomamos una biyección $f : \alpha \longrightarrow x$, para un cierto ordinal α, que no puede ser un número natural, luego $\omega \leq \alpha$, luego $f|_\omega : \omega \longrightarrow x$ es una aplicación inyectiva. ∎

Capítulo XII

Las teorías de conjuntos ZFC y NBG

Llegados a este punto ya conocemos todos los axiomas que los matemáticos aceptan habitualmente como base para sus razonamientos. Sólo nos falta dar nombre a las teorías correspondientes:

Definición 12.1 La *teoría de conjuntos de Zermelo–Fraenkel* (ZF) es la teoría axiomática sobre $\mathcal{L}_{\rm tc}$ cuyos axiomas son los siguientes:[1]

Extensionalidad	$\bigwedge xy(\bigwedge u(u \in x \leftrightarrow u \in y) \to x = y)$	
Par	$\bigwedge xy\bigvee z\bigwedge u(u \in z \leftrightarrow u = x \vee u = y)$	
Unión	$\bigwedge x\bigvee y\bigwedge u(u \in y \leftrightarrow \bigvee v(u \in v \wedge v \in x))$	
Reemplazo	$\bigwedge xyz(\phi(x,y) \wedge \phi(x,z) \to y = z)$	
	$\to \bigwedge a\bigvee b\bigwedge y(y \in b \leftrightarrow \bigvee x \in a\, \phi(x,y))$	$(*)$
Infinitud	$\bigvee x(\varnothing \in x \wedge \bigwedge u \in x\, u' \in x)$	
Partes	$\bigwedge x\bigvee y\bigwedge u(u \subset x \to u \in y)$	
Regularidad	$\bigwedge x(x \neq \varnothing \to \bigvee u(u \in x \wedge u \cap x = \varnothing))$	

$(*)$ para toda fórmula $\phi(x,y)$, tal vez con más variables libres, distintas de b.

La teoría que resulta de añadir a ZF el axioma de elección recibe el nombre de ZFC (por el inglés *choice*). Los cuatro primeros axiomas (que incluyen el esquema axiomático de reemplazo) constituyen la teoría básica que hemos llamado ZF*.

Llamaremos $\mathrm{ZF}_{\rm fin}$ a la teoría que resulta de sustituir el axioma de infinitud por el axioma "todo conjunto es finito" y sustituir el axioma de regularidad por $V = R$ o, equivalentemente, por el axioma TC que afirma que todo conjunto está contenido en un conjunto transitivo (y entonces AP y AE son redundantes).

[1] Como convenio para regular las descripciones impropias, en todas las teorías consideraremos como axioma adicional la sentencia $\bigwedge u\ u \notin x | x = x$.

Por otra parte tenemos la *teoría de conjuntos de von Neumann-Bernays-Gödel* (NBG), sobre el mismo lenguaje formal $\mathcal{L}_{\mathrm{tc}}$, cuyos axiomas son los siguientes más el axioma de elección:

Extensionalidad	$\bigwedge XY(\bigwedge u(u \in X \leftrightarrow u \in Y) \to X = Y)$
Comprensión	$\bigvee X\bigwedge u(u \in X \leftrightarrow \phi(u))$ $\quad(*)$
Vacío	$\bigvee x\bigwedge u\, u \notin x$
Par	$\bigwedge xy\bigvee z\bigwedge u(u \in z \leftrightarrow u = x \vee u = y)$
Unión	$\bigwedge x\bigvee y\bigwedge u(u \in y \leftrightarrow \bigvee v(u \in v \wedge v \in x))$
Reemplazo	$\bigwedge aF(\mathrm{Un}\, F \to \bigvee b\bigwedge v(v \in b \leftrightarrow \bigvee u \in a\,(u, v) \in F))$
Infinitud	$\bigvee x(\varnothing \in x \wedge \bigwedge u \in x\, u' \in x)$
Partes	$\bigwedge x\bigvee y\bigwedge u(u \subset x \to u \in y)$
Regularidad	$\bigwedge x(x \neq \varnothing \to \bigvee u(u \in x \wedge u \cap x = \varnothing))$

$(*)$ para toda fórmula primitiva $\phi(u)$ tal vez con más variables libres (distintas de X).

Recordemos que al trabajar en NBG adoptamos el convenio de que las letras mayúsculas representan clases arbitrarias y las minúsculas representan conjuntos. Los seis primeros axiomas (que incluyen el esquema de comprensión) constituyen la teoría básica que hemos llamado NBG*.

Hemos probado (teorema 10.17) que ZF* y NBG* son equivalentes, en el sentido de que una sentencia α es demostrable en ZF* si y sólo si la sentencia α^V que resulta de relativizar sus variables a conjuntos es demostrable en NBG*. Como los axiomas adicionales de NBG (infinitud, partes, regularidad y elección) son las relativizaciones de los axiomas correspondientes de ZFC, es inmediato que ZFC y NBG son equivalentes en el mismo sentido.

También hemos probado (teorema 10.18) que NBG* es finitamente axiomatizable y, por consiguiente, NBG también.

La *teoría de conjuntos de Morse-Kelley* (MK) es la que resulta de admitir en NBG el esquema de comprensión para fórmulas cualesquiera, no necesariamente primitivas. Hemos probado (teorema 10.21) que $\underset{\mathrm{MK}^*}{\vdash} \mathrm{Consis}\ulcorner\mathrm{ZF}^*\urcorner$ y, según la observación posterior, tenemos igualmente que

$$\underset{\mathrm{MK}}{\vdash} \mathrm{Consis}\ulcorner\mathrm{ZFC}\urcorner,$$

mientras que $\mathrm{Consis}\ulcorner\mathrm{ZFC}\urcorner$ no puede probarse en NBG si es consistente, pues si se pudiera probar, también podría probarse en ZFC, y ésta sería contradictoria por el segundo teorema de incompletitud, y con ella también NBG. Así pues, MK no es equivalente a ZFC y NBG, sino que es una teoría más potente.

Desde el momento en que tenemos $V = R$, es costumbre llamar V_α a las clases (que son conjuntos en ZF) que en el capítulo anterior llamábamos R_α, de

modo que en ZF la clase universal está estratificada en una jerarquía creciente de conjuntos transitivos $V = \bigcup_{\alpha\in\Omega} V_\alpha$, donde

$$V_0 = \varnothing \quad \wedge \quad \bigwedge\alpha\, V_{\alpha+1} = \mathcal{P}V_\alpha \quad \wedge \quad \bigwedge\lambda\, V_\lambda = \bigcup_{\delta<\lambda} V_\delta.$$

ZFC o NBG son las teorías de conjuntos que podríamos llamar "estándar", en el sentido de que actualmente los matemáticos consideran (salvo corrientes filosóficas minoritarias) que una demostración es válida si y sólo si es formalizable en ZFC (o, equivalentemente, en NBG), sin perjuicio de que algunos resultados requieran axiomas adicionales, pero la costumbre en tal caso es dejar constancia explícita de su uso.

En los capítulos anteriores ya hemos empezado a estudiar estas teorías, y aquí vamos a profundizar en este estudio. Empezaremos mostrando la relación que mantienen con otras teorías que hemos manejado.

12.1 Relación con KP

En el capítulo VI introdujimos la teoría de conjuntos de Kripke-Platek (KP). Recordemos sus axiomas:

Extensionalidad	$\bigwedge xy(\bigwedge u(u \in x \leftrightarrow u \in y) \to x = y)$
Par	$\bigwedge xy\bigvee z(x \in z \wedge y \in z)$
Unión	$\bigwedge x\bigvee y\bigwedge u \in x\bigwedge v \in u\ v \in y$
Δ_0-especificación	$\bigwedge x\bigvee y\bigwedge u(u \in y \leftrightarrow (u \in x \wedge \phi(u)))$ $(*)$
Δ_0-recolección	$\bigwedge u\bigvee v\ \phi(u,v) \to \bigwedge a\bigvee b\bigwedge u \in a\bigvee v \in b\ \phi(u,v)$ $(*)$
Π_1-regularidad	$\bigvee u\,\phi(u) \to \bigvee u\,(\phi(u) \wedge \bigwedge v \in u\,\neg\phi(u))$ $(**)$

$(*)$ Para toda fórmula ϕ (tal vez con más variables libres) de tipo Δ_0,
$(**)$ Para toda fórmula ϕ (tal vez con más variables libres) de tipo Π_1.

Llamaremos KPI a la teoría que resulta de añadir a KP el axioma de infinitud (el mismo axioma de infinitud de ZFC).

Nota Hemos adoptado esta versión para que los axiomas de KP sean demostrables en $\mathrm{I}\Sigma_1$ (teorema 6.16) y, más aún, para que $\mathrm{KP}_{\mathrm{fin}}$ sea esencialmente equivalente a $\mathrm{I}\Sigma_1$ (véase la nota de la página 229), pero es habitual considerar que la teoría de Kripke-Platek no tiene la restricción a fórmulas Π_1 en el axioma de regularidad. Nosotros llamaremos KP^+ a la teoría de Kripke-Platek con regularidad para fórmulas arbitrarias, lo cual equivale a contar con el principio de $\in$-inducción para fórmulas arbitrarias:

$$(\bigwedge u \in x\,\phi(u) \to \phi(x)) \to \bigwedge x\,\phi(x).$$

Similarmente, KPI^+ será la extensión correspondiente de KPI. ∎

Hay que señalar que KPI es una teoría bastante más potente que KP. Por ejemplo, en KP pueden probarse los mismos teoremas aritméticos que en $\mathrm{I}\Sigma_1$. En particular, no puede demostrarse el principio de inducción para fórmulas arbitrarias. En cambio:

Teorema 12.2 KPI *es una teoría aritmética (es decir, representa a* AP*).*

Demostración: Sabemos que KPI es una teoría aritmética con inducción Σ_1 porque lo es KP. Sólo falta probar que en KPI puede demostrarse el principio de inducción para fórmulas aritméticas arbitrarias. En efecto, si $\phi(x)$ es una fórmula aritmética (tal vez con más variables libres), llamamos $\phi^*(x,y)$ a la fórmula que resulta de cambiar cada $\bigwedge u \in \omega$ por $\bigwedge u \in y$ y cada $\bigvee u \in \omega$ por $\bigvee u \in y$, donde y es cualquier variable que no aparezca en ϕ. Como la suma y el producto son términos Δ_1, es claro que $\phi^*(x,y)$ es también una fórmula Δ_1, así como que

$$\phi(x) \leftrightarrow \bigvee y(y = \omega \wedge \phi^*(x,y)) \leftrightarrow \bigwedge y(y = \omega \to \phi^*(x,y)).$$

Ahora bien, la fórmula $y = \omega$ es Δ_0, ya que $\alpha \in \Omega$ es Δ_0 (pues equivale a que α sea transitiva y $\in$-conexa), $\alpha \in \omega$ equivale a

$$\alpha \in \Omega \wedge \bigwedge n \in \alpha(n = 0 \vee \bigvee m \in n\ n = m') \wedge (\alpha = 0 \vee \bigvee m \in \alpha\ \alpha = m')$$

y a su vez,

$$y = \omega \leftrightarrow \bigwedge u \in y\ u \in \omega \wedge 0 \in y \wedge \bigwedge n \in y\ n' \in y.$$

Concluimos que toda fórmula aritmética es Δ_1 en KPI, luego podemos definir el conjunto

$$a = \{x \in \omega \mid \phi(x)\}$$

por Δ_1-especificación, y el teorema 3.23 implica el principio de inducción para ϕ. ∎

Ahora podemos mostrar la relación entre estas teorías y ZF:

Teorema 12.3 *La teoría* KP^+ *es una subteoría de* ZF−AI+CT*, y por consiguiente* KPI^+ *es una subteoría de* ZF.

Demostración: Los axiomas de extensionalidad, par y unión son también axiomas de ZF-AI-AP, mientras que el esquema de Δ_0-especificación es parte del esquema de especificación, demostrable en ZF^* (teorema 10.8). Según el teorema 11.29 tenemos que en ZF−AI−AP+CT es demostrable $V = R$ y, según las observaciones posteriores, esto implica el esquema de regularidad de KP^+. Sólo falta probar que en ZF−AI+CT puede probarse el esquema de Δ_0-recolección. De hecho, demostraremos que se cumple el esquema de recolección sin restringirlo a fórmulas de tipo Δ_0.

Suponemos, pues, que $\bigwedge u \bigvee v\, \phi(u,v)$ (donde ϕ puede tener otros parámetros). Definimos $\psi(u,\alpha) \equiv \alpha$ es el mínimo ordinal tal que $\bigvee v \in V_\alpha\, \phi(u,v)$. Es claro entonces que $\bigwedge u \overset{1}{\bigvee} \alpha\, \psi(u,\alpha)$, luego por reemplazo, dado un conjunto a, existe

un conjunto c tal que $\bigwedge\alpha(\alpha \in c \leftrightarrow \bigvee u \in a\, \psi(u,\alpha))$. Así c es un conjunto de ordinales, luego $\beta = \bigcup c$ es su supremo, y es claro que $\bigwedge u \in a \bigvee v \in V_\beta\, \phi(u,v)$. Por lo tanto, el conjunto $b = V_\beta$ cumple lo exigido por el esquema de recolección. ∎

Así pues, todos los resultados que hemos demostrado en capítulos precedentes trabajando en KP son, de hecho, teoremas de ZF−AI+CT y, en particular, de ZF o de $\mathrm{ZF}_{\mathrm{fin}}$.

Veamos ahora que en KP es posible demostrar los resultados principales sobre regularidad que hemos obtenido para ZF. Empezamos por la existencia de clausura transitiva:

Teorema 12.4 (KP)

$$\bigwedge x \overset{1}{\bigvee} y\, (x \subset y \wedge \bigcup y \subset y \wedge \bigwedge z(x \subset z \wedge \bigcup z \subset z \to y \subset z)).$$

Demostración: Llamamos $\phi(x,y)$ a la fórmula del enunciado sin los dos primeros cuantificadores. Claramente $\phi(x,y) \wedge \phi(x,y') \to y = y'$, es decir, si existe un y que cumple $\phi(x,y)$, es único. Sea

$$\psi(x,y) \equiv x \subset y \wedge \bigcup y \subset y \wedge \bigwedge u \in y \bigvee f n(n \in \omega \wedge f : n+1 \longrightarrow y$$

$$\wedge\, f(0) = u \wedge f(n) \in x \wedge \bigwedge i \in n\ f(i) \in f(i+1)).$$

Es claro que ψ es Σ_1. Además $\psi(x,y) \to \phi(x,y)$, pues si $x \subset z$ y z es transitivo, se cumple que $y \subset z$. En efecto, dado $u \in y$, tomamos $f : n+1 \longrightarrow y$ de acuerdo con ψ y una simple inducción prueba que $\bigwedge i \leq n\ f(n-i) \in z$, y cuando $i = n$ obtenemos $u \in z$.

En particular $\psi(x,y) \wedge \psi(x,y') \to y = y'$, de modo que $\bigvee y\, \psi(x,y)$ es equivalente a $\overset{1}{\bigvee} y\, \phi(x,y)$. Veamos por Σ_1-inducción que $\bigwedge x \bigvee y\, \psi(x,y)$. Para ello suponemos que

$$\bigwedge u \in x \bigvee y\, \psi(u,y),$$

con lo que, de hecho, tenemos que $\bigwedge u \in x \overset{1}{\bigvee} y\, \psi(u,y)$. Por Σ_1-reemplazo existe $g : x \longrightarrow y$ suprayectiva tal que $\bigwedge u \in x\, \psi(u, g(u))$. Sea $z = x \cup \bigcup y$. Es fácil ver que z es un conjunto transitivo y $x \subset z$. Para probar $\psi(x,z)$ tomamos $u \in z$. Si $u \in x$ basta tomar $n = 1$, $f = \{(0,u)\}$ y se cumple lo requerido. En caso contrario existe $v \in y$ tal que $u \in v$ y existe un $u' \in x$ tal que $\psi(u',v)$. Esto implica a su vez que existe $h : n+1 \longrightarrow v$ tal que $h(0) = u$, $h(n) \in u'$, etc. Basta tomar $f = h \cup \{(n+1, u')\}$ y se cumple lo requerido.

Así pues, $\bigvee y\, \psi(x,y)$ y esto implica $\overset{1}{\bigvee} y\, \psi(x,y)$. Con esto queda probado que $\bigwedge x \overset{1}{\bigvee} y\, \psi(x,y)$ lo cual implica a su vez que $\bigwedge x \overset{1}{\bigvee} y\, \phi(x,y)$. Más aún, la unicidad implica que $\bigwedge xy(\phi(x,y) \leftrightarrow \psi(x,y))$. ∎

Así pues, la definición de clausura transitiva 11.24 es válida en KP. Teniendo en cuenta que las fórmulas $\neg\phi$ y ψ del teorema anterior son Σ_1, vemos que la

fórmula $y = \mathrm{ct}(x)$ es Δ_1. La prueba del teorema anterior (o, alternativamente, la de 11.25, que es trivialmente válida en KP) muestra también que

$$\mathrm{ct}(x) = x \cup \bigcup_{u\in x} \mathrm{ct}(u).$$

Ahora podemos probar un principio fuerte de Σ_1-inducción

Teorema 12.5 *Para toda fórmula $\phi(x)$ (resp. de clase Σ_1), la fórmula siguiente es un teorema de* KP^+ *(resp. de* KP*):*

$$\bigwedge x(\bigwedge u \in \mathrm{ct}(x)\, \phi(u) \to \phi(x)) \to \bigwedge x\, \phi(x).$$

Demostración: Veamos por Σ_1-inducción que $\bigwedge x \bigwedge u \in \mathrm{ct}(x)\, \phi(u)$. Notemos que si ϕ es de clase Σ_1, lo mismo sucede con la fórmula tras el $\bigwedge x$, pues

$$\bigwedge u \in \mathrm{ct}(x)\, \phi(u) \leftrightarrow \bigvee y(y = \mathrm{ct}(x) \wedge \bigwedge u \in y\ \phi(u)).$$

Suponemos que $\bigwedge v \in x \bigwedge u \in \mathrm{ct}(v)\, \phi(u)$, pero, por la hipótesis del teorema, esto implica $\bigwedge v \in x\, \phi(v)$, luego en total tenemos que $\phi(v)$ se cumple para todos los elementos de

$$x \cup \bigcup_{u\in x} \mathrm{ct}(u) = \mathrm{ct}(x).$$

Esto termina la inducción y, como $x \in \mathrm{ct}(\{x\})$, podemos concluir $\bigwedge x\, \phi(x)$. ∎

Teorema 12.6 (Σ_1-recursión) *Si ϕ es una fórmula Σ_1, existe otra fórmula ψ, también Σ_1, tal que la fórmula siguiente es un teorema de* KP:

$$\bigwedge x_1 \cdots x_n x f \overset{1}{\bigvee} y\ \phi(x_1, \ldots, x_n, x, f, y) \to$$

$$\bigwedge x_1 \cdots x_n x y(\psi(x_1, \ldots, x_n, x, y) \leftrightarrow \phi(x_1, \ldots, x_n, x, [\psi]_x, y)),$$

donde

$$[\psi]_x = \{(u, v) \mid u \in \mathrm{ct}(x) \wedge \psi(x_1, \ldots, x_n, u, v)\}.$$

Demostración: Por simplificar la notación, consideraremos un único parámetro x_1. Consideramos la fórmula Σ_1 siguiente:

$$\chi(x_1, x, y, f) \equiv f \text{ es una función} \wedge \mathcal{D}f = \mathrm{ct}(x) \wedge$$

$$\bigwedge u \in \mathcal{D}f\, \phi(x_1, u, f|_{\mathrm{ct}(u)}, f(u)) \wedge \phi(x_1, x, f, y).$$

Observemos que, en particular,

$$\chi(x_1, x, y, f) \to \phi(x_1, x, f, y), \tag{12.1}$$

así como que

$$\chi(x_1, x, y, f) \wedge u \in \mathrm{ct}(x) \to \chi(x_1, u, f(u), f|_{\mathrm{ct}(u)}). \tag{12.2}$$

Vamos a probar

$$\bigwedge x_1 x \overset{1}{\bigvee} y \bigvee f\ \chi(x_1, x, y, f). \tag{12.3}$$

En primer lugar demostramos la unicidad, es decir, que

$$\chi(x_1, x, y, f) \wedge \chi(x_1, x, y', f') \to y = y' \wedge f = f'. \tag{12.4}$$

Razonamos por inducción sobre $\mathrm{ct}(x)$. Por lo tanto, suponemos que para todo $u \in \mathrm{ct}(x)$ existen a lo sumo unos v y g tales que $\chi(x_1, u, v, g)$, así como que $\chi(x_1, x, y, f) \wedge \chi(x_1, x, y', f')$.

Entonces f y f' son funciones con dominio $\mathrm{ct}(x)$, y la hipótesis de inducción junto con (12.2) implica que para todo $u \in \mathrm{ct}(x)$ se cumple $f(u) = f'(u)$, luego $f = f'$. A su vez, por (12.1) tenemos $\phi(x_1, x, f, y) \wedge \phi(x_1, x, f', y')$, y la unicidad de ϕ implica que $y = y'$.

Seguidamente demostramos la existencia, también por inducción sobre $\mathrm{ct}(x)$. Ello significa suponer que $\bigwedge u \in \mathrm{ct}(x) \bigvee v g\, \chi(x_1, u, v, g)$ y demostrar lo mismo para x. Por (12.4) tenemos, de hecho, que

$$\bigwedge u \in \mathrm{ct}(x) \overset{1}{\bigvee} v g\, \chi(x_1, u, v, g).$$

Por Σ_1-reemplazo existe una función $f : \mathrm{ct}(x) \longrightarrow z$ tal que

$$\bigwedge u \in \mathrm{ct}(x) \overset{1}{\bigvee} g\ \chi(x_1, u, f(u), g).$$

Más concretamente, veamos que

$$\bigwedge u \in \mathrm{ct}(x)\, \chi(x_1, u, f(u), f|_{\mathrm{ct}(u)}).$$

En efecto, dado $u \in \mathrm{ct}(x)$, sea g tal que $\chi(x_1, x, f(u), g)$. Si $v \in \mathrm{ct}(u)$, por (12.2) tenemos $\chi(x_1, v, g(v), g|_{\mathrm{ct}(v)})$, pero, como $v \in \mathrm{ct}(x)$, también existe un g' tal que $\chi(x_1, v, f(v), g'))$ y (12.4) implica que $g(v) = f(v)$. Esto prueba que $g = f|_{\mathrm{ct}(u)}$.

En particular, por (12.1)

$$\bigwedge u \in \mathrm{ct}(x)\, \phi(x_1, u, f|_{\mathrm{ct}(u)}, f(u)).$$

Por la hipótesis sobre ϕ existe un único y tal que $\phi(x_1, x, f, y)$, y ahora es inmediato que $\chi(x_1, x, y, f)$.

Definimos $\psi(x_1, x, y) \equiv \bigvee f\, \chi(x_1, x, y, f)$, que claramente es una fórmula Σ_1. Por (12.3) tenemos que $\bigwedge u \in \mathrm{ct}(x) \overset{1}{\bigvee} v\, \psi(x_1, u, v)$, luego por Σ_1-reemplazo existe el conjunto $[\psi]_x$ indicado en el enunciado (y es una función de dominio $\mathrm{ct}(x)$).

Así, para cada x_1, x existe un único y tal que $\psi(x_1, x, y)$. Esto a su vez implica que existe un f tal que $\chi(x_1, x, y, f)$. Vamos a probar que, necesariamente, $f = [\psi]_x$. Ahora bien, si $u \in \mathrm{ct}(x)$, por (12.2) tenemos $\chi(x_1, u, f(u), f|_{\mathrm{ct}(u)})$, luego $\psi(x_1, u, f(u))$, luego $f(u) = [\psi]_x(u)$.

En definitiva, $\psi(x_1, x, y)$ equivale a que y es el único conjunto tal que $\chi(x_1, x, y, [\psi]_x)$, y por (12.1) es también el único y tal que $\phi(x_1, x, [\psi]_x, y)$. ∎

Si llamamos G a la clase definida por la fórmula ϕ del teorema anterior, la unicidad de su hipótesis hace que $G : V^{n+2} \longrightarrow V$, de modo que en lugar de $\phi(x_1, \ldots, x_n, x, f, y)$ podemos escribir $y = G(x_1, \ldots, x_n, f)$. Similarmente, la clase F definida por la fórmula ψ resulta ser una función $F : V^n \times V \longrightarrow V$ con la propiedad de que, para todo conjunto x, la restricción $F|_{\mathrm{ct}(x)}$ (con los parámetros $x_1, \ldots, x_n$ fijos) es un conjunto y la relación del enunciado equivale a

$$F(x_1, \ldots, x_n, x) = G(x_1, \ldots, x_n, x, F|_{\mathrm{ct}(x)}).$$

A menudo resulta práctico restringir los dominios de las aplicaciones, para descartar casos triviales. Concretamente, consideramos una clase $D \subset V^{n+1}$ de tipo Δ_1. Esto significa que existen fórmulas $\phi(x_1, \ldots, x_n, x)$ y $\psi(x_1, \ldots, x_n, x)$ de tipo Σ_1 y Π_1 respectivamente tales que

$$\bigwedge x_1 \cdots x_n x(\phi(x_1, \ldots, x_n, x) \leftrightarrow \psi(x_1, \ldots, x_n, x)),$$

y entonces escribimos $(x_1, \ldots, x_n, x) \in D$ como abreviatura por cualquiera de las dos fórmulas equivalentes anteriores. A partir de ellas podemos definir a su vez las fórmulas

$$\phi'(x_1, \ldots, x_n, x, f) \equiv \phi(x_1, \ldots, x_n, x) \wedge f : x \longrightarrow V,$$

$$\psi'(x_1, \ldots, x_n, x, f) \equiv \psi(x_1, \ldots, x_n, x) \wedge f : x \longrightarrow V,$$

que son también Σ_1 y Π_1 respectivamente, ya que la parte final de ambas es Δ_0. Por lo tanto, este par de fórmulas define una clase Δ_1 que llamaremos $E \subset V^{n+2}$. Supongamos ahora que $\chi(x_1, \ldots, x_n, x, f, y)$ es una fórmula Σ_1 tal que

$$\bigwedge x_1 \ldots x_n x f(\phi'(x_1, \ldots, x_n, x, f) \to \overset{1}{\bigvee} y\ \chi(x_1, \ldots, x_n, x, f, y)).$$

Esta condición puede expresarse diciendo que χ define una función $G : E \longrightarrow V$. Es en estos términos como han de entenderse las hipótesis del teorema siguiente:

Teorema 12.7 (Σ_1-recursión en KP) *Sea $D \subset V^{n+1}$ una clase de clase Δ_1, sea E la clase (también Δ_1) dada por*

$$(x_1, \ldots, x_n, x, f) \in E \leftrightarrow (x_1, \ldots, x_n, x) \in D \wedge f : x \longrightarrow V,$$

sea $G : E \longrightarrow V$ una función de clase Σ_1. Entonces existe $F : D \longrightarrow V$, definida por una fórmula Σ_1, tal que

$$\bigwedge x_1 \cdots x_n x \in D\ F(x_1, \ldots, x_n, x) = G(x_1, \ldots, x_n, x, F|_x).$$

Demostración: Manteniendo la notación previa al enunciado, consideramos la fórmula

$$\phi''(x_1, \ldots, x_n, x, f, y) \equiv \bigvee g(g = f|_x \wedge ((\phi'(x_1, \ldots, x_n, x, g) \wedge$$

$$\chi(x_1, \ldots, x_n, x, g, y)) \vee (\neg\psi'(x_1, \ldots, x_n, x, g) \wedge y = \varnothing))).$$

Teniendo en cuenta que $g = f|_x$ $[= f \cap (x \times V)]$ es Δ_0, es claro que ϕ'' es Σ_1 y cumple la hipótesis del teorema 12.6, el cual nos da una fórmula ψ'' de clase Σ_1 que define una función $F : V^{n+1} \longrightarrow V$. Si, concretamente, tomamos $(x_1, \ldots, x_n, x) \in D$, para calcular F hemos de considerar el conjunto

$$f = [\psi'']_x = \{(u, v) \mid u \in \mathrm{ct}(x) \wedge \psi''(x_1, \ldots, x_n, u, v)\},$$

que es la función $f : \mathrm{ct}(x) \longrightarrow V$ definida por Σ_1-reemplazo a partir de ψ'' y $\mathrm{ct}(x)$. Entonces $F(x_1, \ldots, x_n, x)$ es el único y que cumple $\psi''(x_1, \ldots, x_n, x, f)$. Ahora bien, como $g = f|_x : x \longrightarrow V$ es la función definida por Σ_1-reemplazo a partir de ψ'' y x (es decir, $g = F|_x$), tenemos $\phi'(x_1, \ldots, x_n, x, g)$, luego $F(x_1, \ldots, x_n, x)$ es el único y que cumple $\chi(x_1, \ldots, x_n, g, y)$. Equivalentemente:

$$F(x_1, \ldots, x_n, x) = G(x_1, \ldots, x_n, x, F|_x).$$

∎

Nota Observemos que la clase F dada por el teorema anterior es de hecho Δ_1, pues $y = F(x_1, \ldots, x_n, x)$ equivale a $\bigwedge z(\psi''(x_1, \ldots, x_n, x, z) \to z = y)$, que es una fórmula de tipo Π_1. ∎

Rango Ahora es claro que en KP podemos definir la función $\mathrm{rang} : V \longrightarrow \Omega$ mediante

$$\mathrm{rang}(x) = \bigcup_{u \in x} (\mathrm{rang}(u) + 1).$$

Explícitamente, consideramos la función G dada por

$$G(x, f) = \bigcup_{u \in x} (f(u) + 1).$$

Más explícitamente, dado cualquier conjunto f, tenemos que

$$\bigwedge v \in \mathcal{R} f \bigvee w \ (w = v \cup \{v\}),$$

luego por 6.7 (aplicado de hecho a una fórmula Δ_0), existe el conjunto

$$A = \{v \cup \{v\} \mid \bigvee u \ (u, v) \in f\}$$

y a su vez existe el conjunto $y = \bigcup A$, que, en el caso en que $f : x \longrightarrow \Omega$, es $y = \bigcup_{u \in x} (f(u) + 1)$. Así,

$$y = G(x, f) \leftrightarrow \bigvee A(y = \bigcup_{u \in A} u \wedge \bigwedge w \in A \bigvee u \in x \bigvee v(f(u) = v \wedge w = v \cup \{v\}))$$

$$\wedge \bigwedge u \in x \bigvee v \bigvee w \in A(f(u) = v \wedge w = v \cup \{v\})).$$

La última fórmula $\phi(x, f, y)$ es claramente Σ_1 y satisface la condición de unicidad sobre y (sobre la clase de pares (x, f) tales que $f : x \longrightarrow V$), por lo que define una función G en las condiciones del teorema de recursión y la función F dada por dicho teorema es la función rango.

Tenemos, pues, que la fórmula $\alpha = \mathrm{rang}\, x$ es Δ_1.

A partir de aquí se prueba por Σ_1-inducción que de hecho $\mathrm{rang}(x)$ es un ordinal. Ahora podemos definir las clases $V_\alpha = \{x \mid \mathrm{rang}(x) < \alpha\}$, que no son necesariamente conjuntos, y la prueba de los teoremas 11.48 y 11.49 es válida sin cambio alguno en este contexto.

12.2 La formalización de la lógica en ZF y KP

Ya hemos visto que todos los conceptos sintácticos (los relacionados con lenguajes formales, demostraciones, etc.) son formalizables en $\mathrm{I}\Sigma_1$ o, equivalentemente, en KP, y es claro que también pueden formalizarse (más fácilmente incluso) en ZF^*, puesto que es una teoría aritmética. Consideremos un lenguaje formal $\mathcal{L}$ en el sentido de la definición 8.1. Si, como es habitual, $\mathcal{L}$ tiene un número finito de constantes, relatores y funtores, también podemos definir en estas teorías los conceptos básicos de la teoría de modelos, estableciendo que un modelo de $\mathcal{L}$ es una terna (M, I, d), donde M es un conjunto, $d \in M$ es la descripción impropia del modelo, y el conjunto I es una aplicación

$$I : \mathrm{Const}(\mathcal{L}) \cup \mathrm{Rel}(\mathcal{L}) \cup \mathrm{Fn}(\mathcal{L}) \longrightarrow V$$

que asigna a cada constante de $\mathcal{L}$ un elemento de M, a cada relator n-ádico un subconjunto de M^n y a cada funtor n-ádico una función $M^n \longrightarrow M$.

Es claro que "(M, I, d) es un modelo de $\mathcal{L}$" es una fórmula Δ_1. Por brevedad escribiremos M en lugar de (M, i, d).

Fijado un modelo M y una expresión $\theta^* \in \mathrm{Exp}(\mathcal{L})$, sea s una sucesión de subexpresiones que defina a θ^*, sea Var el conjunto (finito) de las variables que aparecen en s y sea $\mathrm{Val} = M^{\mathrm{Var}}$, que es un conjunto cuya existencia se prueba en KP. Ahora consideramos la fórmula siguiente:

$$\phi(h, i, M, s, \mathrm{Val}) \equiv i \in \omega \wedge h : \mathcal{R}(s|_i) \times \mathrm{Val} \longrightarrow M \cup \{0, 1\} \wedge \bigwedge \theta \in \mathcal{R}(s|_i)(\cdots),$$

donde los puntos suspensivos representan la conjunción de las fórmulas:

1. Si $\theta \equiv x$ es una variable, entonces $h(x, v) = v(x)$,
2. Si $\theta \equiv c$ es una constante, entonces $h(c, v) = I(c)$,
3. Si $\theta \equiv f t_1 \cdots t_n$, entonces $h(\theta, v) = I(f)(h(t_1, v), \ldots, h(t_n, v))$,
4. Si $\theta \equiv R t_1 \cdots t_n$, entonces $h(\theta, v) = \begin{cases} 1 & \text{si } I(R)(h(t_1, v), \ldots, h(t_n, v)), \\ 0 & \text{en otro caso,} \end{cases}$
5. Si $\theta \equiv \neg\alpha$ entonces $h(\theta, v) = 1 - h(\alpha, v)$,
6. Si $\theta \equiv \alpha \to \beta$ entonces $h(\theta, v) = \begin{cases} 1 & \text{si } h(\alpha, v) = 0 \vee h(\beta, v) = 1, \\ 0 & \text{en otro caso,} \end{cases}$
7. Si $\theta \equiv \bigwedge x\alpha$, entonces $h(\theta, v) = \begin{cases} 1 & \text{si } \bigwedge u \in M\ h(\alpha, v_x^u) = 1, \\ 0 & \text{en otro caso,} \end{cases}$
8. Si $\theta \equiv x|\alpha$, entonces $h(\theta, v) = \begin{cases} a & \text{si } \bigwedge u \in M(h(\alpha, v_x^u) = 1 \leftrightarrow u = a), \\ d & \text{si no existe tal } a. \end{cases}$

Una comprobación rutinaria muestra que ϕ es Δ_1, por lo que podemos usar Σ_1-inducción para demostrar que

$$\bigwedge i \in \omega \bigvee h\, \phi(h, i, M, \mathrm{Val})$$

y la Π_1-inducción para probar la unicidad de h. Esto a su vez nos permite definir la fórmula Σ_1

$$\psi(\theta, v, y) \equiv \theta^* \in \mathrm{Exp}(\mathcal{L}) \wedge \bigvee sl\mathrm{VarVal}h(\cdots)$$

donde los puntos suspensivos hacen referencia a la fórmula que afirma que s es una sucesión de expresiones de longitud l que define a θ, que Var es el conjunto de las variables que aparecen en s, que $\mathrm{Val} = M^{\mathrm{Var}}$, que se cumple $\phi(h, l, M, \mathrm{Val})$ y que $h(\theta, v) = y$.

Finalmente, para cada término t y cada fórmula α de $\mathcal{L}$, definimos

$$M(t)[v] \equiv y \mid \psi(t, v, y), \qquad M \vDash \alpha[v] \equiv \psi(t, v, 1),$$

de modo que ambas expresiones son Δ_1 y se prueba que satisfacen las propiedades esperadas:

1. $M(x)[v] = v(x)$,
2. $M(c)[v] = I(c)$,
3. $M(ft_1 \cdots t_n)[v] = I(f)(M((t_1)[v], \ldots, M(t_n)[v])$,
4. Si $M \vDash Rt_1 \cdots t_n[v] \leftrightarrow I(R)(M(t_1)[v], \ldots, M(t_n)[v])$,
5. $M \vDash \neg\alpha[v] \leftrightarrow \neg M \vDash \alpha[v]$,
6. $M \vDash (\alpha \to \beta)[v] \leftrightarrow \neg M \vDash \alpha[v] \vee M \vDash \beta[v]$,
7. $M \vDash \bigwedge x\alpha[v] \leftrightarrow \bigwedge u \in M\ M \vDash \alpha[v^u_x]$,
8. $M(x|\alpha)[v] = \begin{cases} a & \text{si } \bigwedge u \in M(M \vDash \alpha[v^u_x] \leftrightarrow u = a), \\ d & \text{si no existe tal } a. \end{cases}$

En principio las definiciones exigen que v esté definida sobre todas las variables que aparecen en una sucesión de expresiones que defina a la expresión correspondiente, pero el argumento del teorema 1.9 prueba que en realidad sólo dependen de su restricción a las variables libres en la expresión correspondiente, por lo que podemos modificar la definición para que $M(t)[v]$ y $M \vDash \alpha[v]$ esté definido siempre que v es una valoración definida al menos sobre las variables libes en t o α.

A su vez podemos definir las fórmulas Δ_1:

$$M \vDash \alpha \equiv \bigwedge v \in M^{\mathrm{Vlib}(\alpha)}\ M \vDash \alpha[v],$$

$$M \vDash \Gamma \equiv M \text{ es un modelo de } \mathcal{L} \wedge \Gamma \subset \mathrm{Form}(\mathcal{L}) \wedge \bigwedge \gamma \in \Gamma\ M \vDash \gamma.$$

El teorema de corrección se formaliza sin dificultad: una inducción sobre la longitud de una demostración de γ nos da claramente que

$$\bigwedge M\Gamma\gamma(M \vDash \Gamma \wedge \Gamma \cup \{\gamma\} \subset \mathrm{Form}(\mathcal{L}) \wedge \Gamma \vdash \gamma \to M \vDash \gamma),$$

y a su vez esto implica que si T es una teoría axiomática (y definimos, naturalmente, $M \vDash T \equiv M \vDash \mathrm{Ax}(T)$) entonces

$$\bigvee M\ M \vDash T \to \mathrm{Consis}\, T,$$

pues si existiera una fórmula γ tal que $\underset{T}{\vdash} \gamma \wedge \underset{T}{\vdash} \neg\gamma$, entonces tendríamos la contradicción $M \vDash \gamma \wedge \neg M \vDash \gamma$.

El recíproco es el teorema de completitud, y es claro que no va a poder demostrarse en ZF^* o KP, pues muchas teorías axiomáticas (como AP, sin ir más lejos) sólo admiten modelos infinitos, y en ZF^* o KP no puede probarse la existencia de conjuntos infinitos.

No obstante, si admitimos el axioma de infinitud, es pura rutina comprobar que la demostración de 4.15 se formaliza literalmente en KPI o en ZF^*+AI, palabra por palabra (con las comprobaciones típicas sobre la complejidad de las fórmulas involucradas en el caso de KPI), de modo que en cualquiera de las dos teorías se demuestra, para toda teoría axiomática T:

$$\mathrm{Consis}\, T \leftrightarrow \bigvee M\ M \vDash T,$$

y además el modelo se puede tomar numerable.

Por ejemplo, es fácil ver que el modelo de $\ulcorner\mathcal{L}_a\urcorner$ de universo ω en el que los funtores de $\mathcal{L}_a$ se interpretan con las funciones sucesor, suma y producto cumple $\omega \vDash \ulcorner\mathrm{AP}\urcorner$, por lo que

$$\underset{\mathrm{KPI}}{\vdash} \mathrm{Consis}\, \mathrm{AP}, \qquad \underset{\mathrm{ZF}^*+\mathrm{AI}}{\vdash} \mathrm{Consis}\, \mathrm{AP}.$$

En cualquiera de estas dos teorías está definido el conjunto V_ω, y se puede comprobar[2] que el modelo de $\ulcorner\mathcal{L}_{\mathrm{tc}}\urcorner$ de universo V_ω que interpreta el relator $\ulcorner\in\urcorner$ como la relación $\in$ cumple

$$V_\omega \vDash \ulcorner\mathrm{ZF}_{\mathrm{fin}}\urcorner,$$

por lo que

$$\underset{\mathrm{KPI}}{\vdash} \mathrm{Consis}\, \mathrm{ZF}_{\mathrm{fin}}, \qquad \underset{\mathrm{ZF}^*+\mathrm{AI}}{\vdash} \mathrm{Consis}\, \mathrm{ZF}_{\mathrm{fin}}.$$

Nota Vemos así que, de entre todas las teorías que hemos estudiado, KPI es las más débil en la que es posible formalizar todos los resultados metamatemáticos sobre lógica que hemos demostrado en este libro. De hecho, el axioma de infinitud sólo es necesario para los relacionados con el teorema de completitud y la construcción de modelos en general.

[2]Posponemos la prueba hasta 12.17, pues más adelante contaremos con resultados generales que simplifican las comprobaciones.

Esto significa que, ante la pregunta de qué propiedades necesitamos postular de los objetos metamatemáticos (números, conjuntos, relaciones, funciones, sucesiones, etc.) para llevar adelante los razonamientos que sirven de fundamento a la lógica formal, la respuesta es que, para la parte puramente sintáctica, sólo necesitamos postular que los números naturales cumplen los axiomas de Peano (los de $\mathrm{I}\Sigma_1$ bastan en realidad), y para la parte semántica necesitamos como máximo la existencia objetiva de unos conjuntos que cumplan los axiomas de KPI. (En realidad, en el capítulo VI de [LF] se prueba que basta una teoría mucho más débil: AR_0.) ∎

Al admitir AI podemos simplificar las definiciones de lenguaje formal, teoría axiomática, modelo etc., pues todos estos conceptos pueden definirse como conjuntos en vez de clases. Por ejemplo, podemos imitar la definición que dimos en ACA_0 en la página 354, pero entendiendo ahora que los conjuntos son conjuntos en ZF^*+AI o en KPI, de modo que un lenguaje formal pasa a ser una nónupla ordenada de conjuntos

$$\mathcal{L} = (\ulcorner\neg\urcorner, \ulcorner\to\urcorner, \ulcorner\textstyle\bigwedge\urcorner, \ulcorner|\urcorner, \ulcorner=\urcorner, V, R, F, \mathrm{Nar}).$$

A partir de aquí, se aplica la definición 8.1 tomando como fórmula $x \in \mathrm{Var}(\mathcal{L})$ la fórmula $x \in \mathcal{L}_5$ (la quinta componente de la nónupla $\mathcal{L}$), etc. y tenemos, en particular, que todos los conceptos relacionados con $\mathcal{L}$ (ser un término, una fórmula, etc.) se expresan mediante fórmulas de tipo Δ_1. La diferencia es que ahora tenemos además que $\mathrm{Sig}(\mathcal{L})$, $\mathrm{Cad}(\mathcal{L})$, etc., son conjuntos.

Similarmente, bajo AI podemos definir un modelo de $\mathcal{L}$ incluso en el caso en que el número de constantes, relatores y funtores de $\mathcal{L}$ sea infinito.

Modelos de ZF Aunque la definición de modelo es más compleja, en ultima instancia un modelo de $\ulcorner\mathcal{L}_{\mathrm{tc}}\urcorner$ está completamente determinado por un par (M, R), donde M es un conjunto no vacío y $R \subset M \times M$ es la relación que interpreta al relator $\ulcorner\in\urcorner$. (En realidad hace falta especificar también una descripción impropia, pero vamos a considerar únicamente modelos de al menos los axiomas de $\ulcorner\mathrm{ZF}^*\urcorner$, y en tal caso tomaremos siempre como tal al conjunto vacío del modelo, por lo que no necesitamos explicitarla.) Así pues, escribiremos (M, R) para referirnos al modelo de $\ulcorner\mathcal{L}_{\mathrm{tc}}\urcorner$ determinado por M y R. Cuando omitamos R significará que estamos interpretando $\ulcorner\in\urcorner$ con la relación de pertenencia en M, es decir, $R = \{(x, y) \in M \times M \mid x \in y\}$, y entonces se dice que consideramos a M como *modelo natural* de $\mathcal{L}_{\mathrm{tc}}$.

Una consecuencia del segundo teorema de incompletitud es que si T es una teoría axiomática semirrecursiva consistente sobre $\mathcal{L}_{\mathrm{tc}}$ que extienda a ZF^* o a KP, no es posible demostrar en T la existencia de un modelo de $\ulcorner T\urcorner$, pues en tal caso tendríamos que

$$\underset{T}{\vdash} \mathrm{Consis}\,\ulcorner T\urcorner,$$

y el teorema de incompletitud implicaría que T es contradictoria.

En particular no puede probarse la existencia de modelos de $\ulcorner\mathrm{ZF}\urcorner$ en ZF, pero hasta cierto punto, podemos esquivar el problema, como vemos a continuación:

Modelos internos Si M es una clase (no necesariamente un conjunto) y $R \subset M \times M$, no podemos definir en general una relación $(M, R) \vDash \alpha[v]$ que tenga sentido para toda fórmula $\alpha \in \mathrm{Form}(\ulcorner\mathcal{L}_{\mathrm{tc}}\urcorner)$, pues si pudiéramos definir $V \vDash \alpha[v]$ en ZF, entonces podríamos demostrar $\mathrm{Consis}\,\ulcorner\mathrm{ZF}\urcorner$ en ZF como lo hemos probado en MK, y tendríamos una contradicción.

Sin embargo, sí que podemos definir $(M, R) \vDash \alpha$ para fórmulas metamatemáticas α. Sólo tenemos que adaptar el concepto que ya hemos manejado de interpretación de $\mathcal{L}_a$ en una teoría arbitraria (definición 3.28). En este caso necesitamos interpretar $\mathcal{L}_{\mathrm{tc}}$ en el propio $\mathcal{L}_{\mathrm{tc}}$:

Definición 12.8 Sean $x \in M$ y $x\,R\,y$ dos fórmulas del lenguaje $\mathcal{L}_{\mathrm{tc}}$ de la teoría de conjuntos cuyas variables libres sean a lo sumo $x, y, y_1, \ldots, y_r$ (con la variable y sólo en la segunda) y sea $\mathcal{L}'_{\mathrm{tc}}$ el lenguaje $\mathcal{L}_{\mathrm{tc}}$ sin las variables $y_1, \ldots, y_n$.

Para cada expresión θ de $\mathcal{L}'_{\mathrm{tc}}$, definimos la *relativización* a M, R de θ a la expresión θ^{MR} de $\mathcal{L}_{\mathrm{tc}}$ dada por las reglas siguientes:

1. $x^{MR} \equiv x$.
2. $(t_1 = t_2)^{MR} \equiv t_1^{MR} = t_2^{MR}$
3. $(t_1 \in t_2)^{MR} \equiv t_1^{MR}\,R\,t_2^{MR}$
4. $(\neg\alpha)^{MR} \equiv \neg\alpha^{MR}$,
5. $(\alpha \to \beta)^{MR} \equiv \alpha^{MR} \to \beta^{MR}$,
6. $(\bigwedge u\alpha)^{MR} \equiv \bigwedge u \in M\ \alpha^{MR} \equiv \bigwedge u(u \in M \to \alpha^{MR})$,
7. $(u|\alpha)^{MR} \equiv u \in M|\alpha^{MR} \equiv u|(u \in M \wedge \alpha^{MR})$.

Es claro que las variables libres de θ^{MR} son las de θ más a lo sumo $y_1, \ldots, y_r$.

Claramente, de aquí se sigue que

$$(\alpha \vee \beta)^{\mathrm{MR}} \equiv \alpha^{MR} \vee \beta^{MR}, \quad (\alpha \wedge \beta)^{\mathrm{MR}} \equiv \alpha^{MR} \wedge \beta^{MR},$$

$$(\alpha \leftrightarrow \beta)^{\mathrm{MR}} \equiv \alpha^{MR} \leftrightarrow \beta^{MR},$$

así como que $(\bigvee u\,\alpha)^{MR}$ es equivalente a

$$\bigvee u \in M\,\alpha^{MR} \equiv \bigvee u(u \in M \wedge \alpha^{MR}),$$

y que $(\overset{1}{\bigvee}\alpha)^{MR}$ es equivalente a $\overset{1}{\bigvee} u \in M\,\alpha^{MR} \equiv \overset{1}{\bigvee} u(u \in M \wedge \alpha^{MR})$.

A efectos prácticos, conviene destacar que

> *La relativización θ^{MR} es la expresión que resulta de sustituir cada subfórmula $x \in y$ de θ por $x\,R\,y$ y de acotar cada variable ligada en la forma $\bigwedge u \in M$, $\bigvee u \in M$, $u \in M|$.*

Más laxamente, θ^{MR} es lo que entendería por θ alguien que creyera que los únicos conjuntos que existen son los que cumplen $x \in M$ y que la relación de pertenencia es la dada por $x\,R\,y$.

Podemos escribir $(M,R) \vDash \alpha \equiv \bigwedge x_1 \cdots x_n \in M\ \alpha^{MR}$, donde $x_1,\ldots,x_n$ son las variables libres en α, pero es fundamental comprender que aquí α no es una variable de $\mathcal{L}_{\mathrm{tc}}$, como en $(M,R) \vDash \alpha[v]$, sino una metavariable, de modo que $(M,R) \vDash \alpha$ es una fórmula de $\mathcal{L}_{\mathrm{tc}}$ distinta para cada fórmula α de $\mathcal{L}'_{\mathrm{tc}}$.

Definición 12.9 Si S es una teoría axiomática sobre el lenguaje formal $\mathcal{L}'_{\mathrm{tc}}$, representaremos por $(M,R) \vDash S$ el conjunto de todas las fórmulas $(M,R) \vDash \alpha$, para cada axioma α de S (y diremos que (M,R) es un *modelo (interno)* de S.

Notemos que si S tiene infinitos axiomas, como es el caso de Z^* o ZF, entonces $(M,R) \vDash \alpha$ es un conjunto de infinitas fórmulas, por lo que las afirmaciones que involucren $(M,R) \vDash S$ deben interpretarse adecuadamente. Por ejemplo, si decimos que en una teoría T sobre $\mathcal{L}_{\mathrm{tc}}$ se prueba que (M,R) es un modelo de S, esto hay que entenderlo como un esquema teoremático, que afirma que en T se pueden demostrar las relativizaciones de todos los axiomas de S.

Nota En [LF 5.19] damos una definición general de interpretación de una teoría en otra. La definición que acabamos de dar no encaja exactamente en dicha definición porque estamos admitiendo la posibilidad de que las fórmulas $x \in M$ y $x\,R\,y$ tengan más variables libres, mientras que en [LF 5.19] pedimos que sus únicas variables libres sean x y x,y, respectivamente. Esto no supone ninguna diferencia esencial, por lo que los resultados básicos que vamos a probar se corresponden con los vistos en [LF] y también en la sección 3.4. La posible presencia de parámetros en las fórmulas que definen la relativización introduce una pequeña complicación en los argumentos, pero el hecho de que $\mathcal{L}_{\mathrm{tc}}$ no tenga constantes ni funtores simplifica las cosas. ∎

Los dos teoremas siguientes son los equivalentes a 3.29:

Teorema 12.10 *Si $t(x_1,\ldots,x_n)$ es un término de $\mathcal{L}'_{\mathrm{tc}}$ con a lo sumo las variables libres indicadas, entonces $x|(x=x) \in M \vdash \bigwedge x_1 \cdots x_n \in M\, t^{MR} \in M$.*

Demostración: Si $t \equiv x_i$ es obvio que $\vdash \bigwedge x_1 \cdots x_n \in M\ x_i \in M$. La alternativa es que $t \equiv u|\alpha$, y entonces, fijados $x_1,\ldots,x_n \in M$, o bien existe un único $u \in M$ tal que $u \in M \wedge \alpha^{MR}$, en cuyo caso

$$t^{MR} \equiv u|(u \in M \wedge \alpha^{MR}) \in M,$$

o, en caso contrario, $t^{MR} = x|(x=x)$, luego, si suponemos $x|(x=x) \in M$, tenemos igualmente que $t^{MR} \in M$. ∎

Teorema 12.11 *Si θ, x, t son, respectivamente una expresión, una variable y un término de $\mathcal{L}'_{\mathrm{tc}}$, entonces, si θ es un término,*

$$x|(x=x) \in M \vdash \bigwedge x_1 \cdots x_n \in M\ (\mathtt{S}^t_x\theta)^{MR} = \mathtt{S}^{t^{MR}}_x \theta^{MR})$$

y si θ es una fórmula

$$x|(x=x) \in M \vdash \bigwedge x_1 \cdots x_n \in M((\mathtt{S}^t_x\theta)^{MR} \leftrightarrow \mathtt{S}^{t^{MR}}_x \theta^{MR}),$$

donde $x_1,\ldots,x_n$ son todas las variables libres en las expresiones consideradas.

DEMOSTRACIÓN: Por inducción sobre la longitud de θ.

Si $\theta \equiv x$, entonces $\mathtt{S}^t_x\theta \equiv t$, y trivialmente $t^{MR} = \mathtt{S}^{t^{MR}}_x x$.

Si $\theta \equiv x_i \not\equiv x$, entonces $\mathtt{S}^t_x\theta \equiv x_i$, y lo que hay que probar es que $x_i = x_i$.

Si $\theta \equiv t_1 = t_2$, entonces, fijados $x_1, \ldots, x_n \in M$, por hipótesis de inducción $(\mathtt{S}^t_x t_i)^{MR} = \mathtt{S}^{t^{MR}}_x(t_i^{MR})$, y lo que hay que probar es

$$(\mathtt{S}^t_x(t_1 = t_2))^{MR} \equiv (\mathtt{S}^t_x t_1 = S^t_x t_2)^{MR} \equiv (\mathtt{S}^t_x t_1)^{MR} = (\mathtt{S}^t_x t_2)^{MR} \leftrightarrow$$

$$\mathtt{S}^{t^{MR}}_x t_1^{MR} = \mathtt{S}^{t^{MR}}_x t_2^{MR} \equiv S^{t^{MR}}_x(t_1^{MR} = t_2^{MR}) \equiv \mathtt{S}^{t^{MR}}_x(t_1 = t_2)^{MR}.$$

Si $\theta \equiv t_1 \in t_2$ el razonamiento es similar. Ahora tenemos que

$$(\mathtt{S}^t_x(t_1 \in t_2))^{MR} \equiv (\mathtt{S}^t_x t_1 \in S^t_x t_2)^{MR} \equiv (\mathtt{S}^t_x t_1)^{MR}\, R\, (\mathtt{S}^t_x t_2)^{MR} \leftrightarrow$$

$$\mathtt{S}^{t^{MR}}_x t_1^{MR}\, R\, \mathtt{S}^{t^{MR}}_x t_2^{MR} \equiv S^{t^{MR}}_x(t_1^{MR}\, R\, t_2^{MR}) \equiv \mathtt{S}^{t^{MR}}_x(t_1 \in t_2)^{MR},$$

donde, en la penúltima identidad, se usa que la fórmula $u\,R\,v$ no tiene libre la variable x de $\mathcal{L}'_{\mathrm{tc}}$.

Si $\theta \equiv \neg\alpha$, entonces, fijados $x_1, \ldots, x_n \in M$ y usando la hipótesis de inducción,

$$(\mathtt{S}^t_x\neg\alpha)^{MR} \equiv \neg(\mathtt{S}^t_x\alpha)^{MR} \leftrightarrow \neg\mathtt{S}^{t^{MR}}_x\alpha^{MR} \equiv \mathtt{S}^{t^{MR}}_x\neg\alpha^{MR}.$$

Si $\theta \equiv \alpha \to \beta$ el razonamiento es análogo.

Si $\theta \equiv \bigwedge u\alpha$ y x no está libre en θ, luego tampoco en θ^{MR}, tenemos

$$(\mathtt{S}^t_x\textstyle\bigwedge u\alpha)^{MR} \equiv (\bigwedge u\alpha)^{MR} \equiv \mathtt{S}^{t^{MR}}_x(\bigwedge u\alpha)^{MR}.$$

Si x está libre en θ y u no está libre en t (luego tampoco en t^{MR}), supuesto que $x_1, \ldots, x_n \in M$,

$$(\mathtt{S}^t_x\textstyle\bigwedge u\alpha)^{MR} \equiv (\bigwedge u\, \mathtt{S}^t_x\alpha)^{MR} \equiv \bigwedge u \in M\, (\mathtt{S}^t_x\alpha)^{MR}.$$

Por hipótesis de inducción, $x_1, \ldots, x_n, u \in M$ implica $(\mathtt{S}^t_x\alpha)^{MR} \leftrightarrow (\mathtt{S}^{t^{MR}}_x\alpha^{MR})$, luego, usando que x no está libre en $u \in M$,

$$(\mathtt{S}^t_x\textstyle\bigwedge u\alpha)^{MR} \leftrightarrow \bigwedge u(u \in M \to \mathtt{S}^{t^{MR}}_x\alpha^{MR}) \equiv \bigwedge u\, \mathtt{S}^{t^{MR}}_x(u \in M \to \alpha^{MR}) \leftrightarrow$$

$$\mathtt{S}^{t^{MR}}_x(\textstyle\bigwedge u \in M\alpha^{MR}) \equiv \mathtt{S}^t_x(\bigwedge u\alpha)^{MR}.$$

Si x está libre en θ, pero u está libre en t, tomamos la menor variable v no libre en θ ni en t, de modo que, supuesto que $x_1, \ldots, x_n \in M$,

$$(\mathtt{S}^t_x\textstyle\bigwedge u\alpha)^{MR} \equiv (\bigwedge v\, \mathtt{S}^t_x\mathtt{S}^v_u\alpha)^{MR} \equiv \bigwedge v \in M\, (\mathtt{S}^t_x\mathtt{S}^v_u\alpha)^{MR}.$$

Por la hipótesis de inducción aplicada primero a $\mathtt{S}^v_u\alpha$, tenemos que

$$(\mathtt{S}^t_x\textstyle\bigwedge u\alpha)^{MR} \leftrightarrow \bigwedge v \in M\, \mathtt{S}^{t^{MR}}_x(\mathtt{S}^v_u\alpha)^{MR},$$

pero, por la hipótesis de inducción aplicada a α, tenemos que

$$\bigwedge x_1 \cdots x_n, x, v \in M\, ((\mathtt{S}_u^v \alpha)^{MR} \leftrightarrow \mathtt{S}_u^v \alpha^{MR})$$

y, eliminando el $\bigwedge x$ (usando que $t^{MR} \in M$), queda que

$$\mathtt{S}_x^{t^{MR}} (\mathtt{S}_u^v \alpha)^{MR} \leftrightarrow \mathtt{S}_x^{t^{MR}} \mathtt{S}_u^v (\alpha^{MR}),$$

luego

$$(\mathtt{S}_x^t \bigwedge u \alpha)^{MR} \leftrightarrow \bigwedge v (v \in M \to \mathtt{S}_x^{t^{MR}} \mathtt{S}_u^v (\alpha^{MR})) \leftrightarrow \mathtt{S}_x^{t^{MR}} \bigwedge v (v \in M \to \mathtt{S}_u^v (\alpha^{MR})) \leftrightarrow$$

$$\mathtt{S}_x^{t^{MR}} (\bigwedge u \in M \alpha^{MR}) \equiv \mathtt{S}_x^{t^{MR}} (\bigwedge u \alpha)^{MR}.$$

Si $\theta \equiv u|\alpha$ y x no está libre en θ (luego tampoco en θ^{MR}), entonces

$$(\mathtt{S}_x^t (u|\alpha))^{MR} \equiv (u|\alpha)^{MR} \equiv \mathtt{S}_x^{t^{MR}} (u|\alpha)^{MR}.$$

Si x está libre en θ y u no está libre en t (luego tampoco en t^{MR}), supuesto que $x_1, \ldots, x_n \in M$,

$$(\mathtt{S}_x^t (u|\alpha))^{MR} \equiv (u|\mathtt{S}_x^t \alpha)^{MR} \equiv u|(u \in M \wedge (\mathtt{S}_x^t \alpha)^{MR}).$$

Si $x_1, \ldots, x_n, u \in M$, tenemos$(\mathtt{S}_x^t \alpha)^{MR} \leftrightarrow (\mathtt{S}_x^{t^{MR}} \alpha^{MR})$, luego, usando que x no está en $u \in M$,

$$(\mathtt{S}_x^t (u|\alpha))^{MR} = u|(u \in M \wedge \mathtt{S}_x^{t^{MR}} \alpha^{MR}) \equiv$$

$$\mathtt{S}_x^{t^{MR}} u|(u \in M \wedge \alpha^{MR}) \equiv \mathtt{S}_x^{t^{MR}} (u|\alpha)^{MR}.$$

Si x está libre en θ, pero u está libre en t, tomamos la menor variable v no libre en θ ni en t, de modo que, supuesto que $x_1, \ldots, x_n \in M$,

$$(\mathtt{S}_x^t u|\alpha)^{MR} \equiv (v|\mathtt{S}_x^t \mathtt{S}_u^v \alpha)^{MR} \equiv v|(v \in M \wedge (\mathtt{S}_x^t \mathtt{S}_u^v \alpha)^{MR}).$$

Aplicando la hipótesis de inducción como en el caso anterior,

$$(\mathtt{S}_x^t u|\alpha)^{MR} = v|(v \in M \wedge \mathtt{S}_x^{t^{MR}} (S_u^v \alpha)^{MR}) \equiv \mathtt{S}_x^{t^{MR}} (v|(v \in M \wedge \mathtt{S}_u^v (\alpha^{MR}))) =$$

$$\mathtt{S}_x^{t^{MR}} (u|(u \in M \wedge \alpha^{MR})) = \mathtt{S}_x^{t^{MR}} (u|\alpha)^{MR}.$$

∎

Con esto podemos probar el resultado fundamental sobre relativizaciones (compárese con 3.30):

Teorema 12.12 *Si $\alpha_1, \ldots, \alpha_m \vdash \alpha$, para ciertas fórmulas de $\mathcal{L}'_{\mathrm{tc}}$, entonces*

$$(\alpha_1^c)^{MR}, \ldots, (\alpha_m^c)^{MR},\ x|(x = x) \in M \vdash (\alpha^c)^{MR},$$

donde $(\alpha^c)^{MR} \equiv \bigwedge x_1 \cdots x_n \in M\, \alpha^{MR}$, siendo $x_1, \ldots, x_n$ las variables libres de α. En particular, si las fórmulas no tienen variables libres,

$$\alpha_1^{MR}, \ldots, \alpha_m^{MR}, x|(x = x) \in M \vdash \alpha^{MR}.$$

DEMOSTRACIÓN: Veamos en primer lugar que si γ es un axioma lógico, entonces $x|(x = x) \in M \vdash (\gamma^c)^{MR}$.

Por simplificar la notación vamos a suponer que las variables libres en γ son x, y, aunque todos los argumentos que veremos valen sin cambio alguno cualquiera que sea el número de variables libres (y se simplifican si no hay ninguna).

Si γ es uno de los axiomas K1, K2, K3, es inmediato que γ^{MR} es un axioma del mismo tipo, luego $\vdash \gamma^{MR}$, luego $\vdash (x \in M \wedge y \in M \to \gamma^{MR})$, luego, introduciendo generalizadores, $\vdash (\gamma^c)^{MR}$.

Si $\gamma \equiv \bigwedge u\alpha \to \mathtt{S}^t_u\alpha$, entonces, por el teorema anterior,

$$\gamma^{MR} \equiv \bigwedge u \in M\alpha^{MR} \to \mathtt{S}^{t^{MR}}_u \alpha^{MR},$$

y podemos razonar como sigue (la línea 2 es válida por 12.10) usando la premisa $x|(x = x) \in M$:

1)	$x \in M \wedge y \in M$	Hipótesis
2)	$t^{MR} \in M$	Consecuencia de 1)
3)	$\bigwedge u \in M\, \alpha^{MR}$	Hipótesis
4)	$t^{MR} \in M \to \mathtt{S}^{t^{MR}}_u \alpha|_M$	EG 3
5)	$\mathtt{S}^{t^{MR}}_u \alpha^{MR}$	MP 2, 4
6)	$\bigwedge u \in M\, \alpha|_M \to \mathtt{S}^{t^{MR}}_u \alpha^{MR}$	
7)	$x \in M \wedge y \in M \to \bigwedge u \in M\, \alpha^{MR} \to \mathtt{S}^{t^{MR}}_u \alpha^{MR}$	
8)	$\bigwedge xy \in M(\bigwedge u \in M\, \alpha^{MR} \to \mathtt{S}^{t^{MR}}_u \alpha^{MR})$	IG 7

Si $\gamma \equiv \bigwedge u(\alpha \to \beta) \to (\alpha \to \bigwedge u\beta)$, entonces

$$\gamma^{MR} \equiv \bigwedge u \in M\, (\alpha^{MR} \to \beta^{MR}) \to (\alpha^{MR} \to \bigwedge u \in M\, \beta^{MR}).$$

Razonamos como sigue:

1)	$\bigwedge u \in M(\alpha^{MR} \to \beta^{MR})$	Hipótesis
2)	α^{MR}	Hipótesis
3)	$u \in M$	Hipótesis
4)	$u \in M \to (\alpha^{MR} \to \beta^{MR})$	EG 1
5)	$\alpha^{MR} \to \beta^{MR}$	MP 3, 4
6)	β^{MR}	MP 2, 5
7)	$u \in M \to \beta^{MR}$	
8)	$\bigwedge u \in M\ \beta^{MR}$	IG 7
9)	$\alpha^{MR} \to \bigwedge u \in M\ \beta^{MR}$	
10)	$\bigwedge u \in M(\alpha^{MR} \to \beta^{MR}) \to (\alpha^{MR} \to \bigwedge u \in M\ \beta^{MR})$	

De aquí podemos pasar a $\bar{\gamma}^c$ de forma obvia.

Si $\gamma \equiv \bigwedge u(u = t \to \alpha) \leftrightarrow \mathtt{S}^t_u\alpha$, entonces, usando el teorema anterior, γ^{MR} equivale a

$$\bigwedge u \in M(u = t^{MR} \to \alpha^{MR}) \leftrightarrow \mathtt{S}^{t^{MR}}_u \alpha^{MR}.$$

1)	$x \in M \wedge y \in M$	Hipótesis
2)	$t^{MR} \in M$	Consecuencia de 1)
3)	$\bigwedge u \in M(u = t^{MR} \to \alpha^{MR})$	Hipótesis
4)	$t^{MR} \in M \to (t^{MR} = t^{MR} \to \mathtt{S}_u^{t^{MR}}\alpha^{MR})$	EG 3
5)	$\mathtt{S}_u^{t^{MR}}\alpha^{MR}$	MP 2, 4, I
6)	$\bigwedge u \in M(u = t^{MR} \to \alpha^{MR}) \to \mathtt{S}_u^{t^{MR}}\alpha^{MR}$	
7)	$\mathtt{S}_u^{t^{MR}}\alpha^{MR}$	Hipótesis
8)	$\bigwedge u(u = t^{MR} \to \alpha^{MR})$	II 7
9)	$u = t^{MR} \to \alpha^{MR}$	EG 8
10)	$u \in M \to (u = t^{MR} \to \alpha^{MR})$	Consecuencia de 9)
11)	$\bigwedge u \in M(u = t^{MR} \to \alpha^{MR})$	IG 10
12)	$\mathtt{S}_u^{t^{MR}}\alpha^{MR} \to \bigwedge u \in M(u = t^{MR} \to \alpha^{MR})$	
13)	$\bigwedge u \in M(u = t^{MR} \to \alpha^{MR}) \leftrightarrow \mathtt{S}_u^{t^{MR}}\alpha^{MR}$	IB 6, 12
14)	$x \in M \wedge y \in M \to \bigwedge u \in M\,(u = t^{MR} \to \alpha^{MR}) \leftrightarrow \mathtt{S}_u^{t^{MR}}\alpha^{MR}$	
15)	$\bigwedge xy \in M(\bigwedge u \in M(u = t^{MR} \to \alpha^{MR}) \leftrightarrow \mathtt{S}_u^{t^{MR}}\alpha^{MR})$	IG 14

Si $\gamma \equiv \overset{1}{\bigvee} u\alpha \to \mathtt{S}_u^{u|\alpha}\alpha$, entonces (bajo las hipótesis $x, y, x|(x = x) \in M$) es fácil ver que la fórmula γ^{MR} es equivalente a

$$\overset{1}{\bigvee} u(u \in M \wedge \alpha^{MR}) \to \mathtt{S}_u^{u|(u\in M\wedge\alpha^{MR})}\alpha^{MR},$$

y esto es un axioma lógico.

Si $\gamma \equiv \neg\overset{1}{\bigvee} u\alpha \to (u|\alpha) = x|(x = x)$, entonces γ^{MR} es equivalente a

$$\neg\overset{1}{\bigvee} u(u \in M \wedge \alpha|_M) \to u \in M|\alpha^{MR} = x \in M|x = x.$$

Basta observar que $(x \in M|x = x) = x|(x = x)$ para ver que $\gamma|_M$ es equivalente a un axioma lógico. Aquí hay que distinguir dos casos, según si $\overset{1}{\bigvee} y\ y \in M$ o bien $\neg\overset{1}{\bigvee} y \in M$. En el primer caso la igualdad se tiene por la unicidad, usando que $x|(x = x) \in M$. En el segundo las descripciones son impropias y llegamos a la misma igualdad mediante la regla de las descripciones impropias.

Tenemos así que las clausuras de los axiomas lógicos de $\mathcal{L}'_{\mathrm{tc}}$ son consecuencias lógicas de $x|(x = x) \in M$. Ahora supongamos que tenemos una demostración de α a partir de $\alpha_1, \ldots, \alpha_m$, digamos $\gamma_1, \ldots, \gamma_r$, y veamos por inducción sobre i que

$$(\alpha_1^c)^{MR}, \ldots, (\alpha_m^c)^{MR},\ x|(x = x) \in M \vdash (\gamma_i^c)^{MR}.$$

Si γ_i es un axioma lógico, tenemos probado que

$$x|(x = x) \in M \vdash (\gamma_i^c)^{MR}$$

y esto implica lo requerido. Si γ_i es una premisa, la conclusión es trivial.

Supongamos ahora que γ_i se deduce por MP de γ_j y $\gamma_j \to \gamma_i$. Entonces, por hipótesis de inducción

$$(\alpha_1^c)^{MR}, \dots, (\alpha_m^c)^{MR},\ x|(x = x) \in M \vdash (\gamma_j^c)^{MR}.$$

$$(\alpha_1^c)^{MR}, \dots, (\alpha_m^c)^{MR},\ x|(x = x) \in M \vdash ((\gamma_j \to \gamma_i)^c)^{MR},$$

Si $\gamma_i \equiv \beta$ se deduce por MP de $\gamma_j \equiv \alpha$ y $\gamma_k \equiv \alpha \to \beta$, supongamos, por concretar, que $\alpha \equiv \alpha(x, y, z)$ y $\beta \equiv \beta(x, w)$ (es decir, que las variables libres son las indicadas) y razonamos así, donde, por abreviar, llamamos $c = x|(x = x)$:

1)	$\bigwedge xyzw \in M\ (\alpha^{MR} \to \beta^{MR})$	Premisa
2)	$\bigwedge xyz \in M\, \alpha^{MR}$	Premisa
3)	$x \in M \wedge c \in M \wedge c \in M \wedge w \in M \to (\alpha^{MR}(x, c, c) \to \beta^{MR})$	EG 1
4)	$x \in M \wedge c \in M \wedge c \in M \to \alpha^{MR}(x, c, c)$	EG 2
5)	$x \in M \wedge w \in M$	Hipótesis
6)	$c \in M$	Premisa
7)	$x \in M \wedge c \in M \wedge c \in M \wedge w \in M$	IC 5, 6
8)	$\alpha^{MR}(x, c, c) \to \beta^{MR}(x, w)$	MP 3, 7
9)	$\alpha^{MR}(x, c, c)$	MP 4, 7
10)	β^{MR}	MP 8, 9
11)	$x \in M \wedge w \in M \to \beta^{MR}$	
12)	$\bigwedge xw \in M\, \beta^{MR}(x, w)$	IG 11

Si $\gamma_i \equiv \bigwedge u\gamma_j$ se deduce por generalización, por hipótesis de inducción tenemos que

$$(\alpha_1^c)^{MR}, \dots, (\alpha_m^c)^{MR},\ x|(x = x) \in M \vdash (\gamma_j^c)^{MR},$$

pero $(\gamma_j^c)^{MR}$ es equivalente a $(\gamma_i^c)^{MR}$. ∎

En otras palabras, el teorema anterior afirma que si en una teoría T se puede demostrar $(M, R) \vDash S$ y se cumple $\underset{S}{\vdash} \alpha$, entonces $\underset{T}{\vdash}(M, R) \vDash \alpha$, es decir, que en un modelo de una teoría se cumplen todos los teoremas de la teoría.

En la práctica, como ya venimos haciendo, en las teorías de conjuntos tomaremos como axioma que $x|(x = x) = \varnothing$, de modo que para aplicar el teorema anterior se requiere la premisa $\varnothing \in M$.

El caso más simple de relativización se da cuando M y R son dos variables (las dos únicas variables de $\mathcal{L}_{\mathrm{tc}}$ que no están en $\mathcal{L}'_{\mathrm{tc}}$) y las fórmulas $x \in M$, $x\,R\,y \equiv (x, y) \in R$ son las usuales. En tal caso tenemos dos definiciones de satisfacción de una fórmula en un modelo y el teorema siguiente nos da la relación entre ambas:

Teorema 12.13 (ZF*) *Sea M un conjunto tal que $\varnothing \in M$ y $R \subset M \times M$, entonces, para cada fórmula (metamatemática) $\phi(x_1, \dots, x_n)$, si*

$$v \equiv \{(\ulcorner x_1 \urcorner, x_1), \dots, (\ulcorner x_n \urcorner, x_n)\},$$

se cumple

$$\bigwedge x_1 \cdots x_n \in M((M,R) \vDash \ulcorner\phi\urcorner[v] \leftrightarrow \phi^{MR}(x_1,\ldots,x_n)).$$

En particular, si ϕ es una sentencia $(M,R) \vDash \ulcorner\phi\urcorner \leftrightarrow \phi^{MR}$.

DEMOSTRACIÓN: Se cumple también que si t es un término, entonces

$$\bigwedge x_1 \cdots x_n \in M\ (M,R)(t)[v] = t^{MR}(x_1,\ldots,x_n),$$

y ambos hechos se demuestran trivialmente por inducción sobre la longitud de una expresión. ∎

Nota Es importante entender que los dos miembros de la equivalencia dada por el teorema anterior corresponden a definiciones muy distintas. En el miembro izquierdo tenemos la fórmula $(M,R) \vDash \alpha[v]$, que está definida para toda fórmula $\alpha \in \mathrm{Form}(\ulcorner\mathcal{L}_{\mathrm{tc}}\urcorner)$, pero, a cambio, requiere que M sea un conjunto y no una clase propia. (Enseguida veremos por qué es imposible extender la definición al caso de clases propias.)

Por otro lado, el miembro derecho está definido para cualquier clase M (notemos que una fórmula arbitraria $x \in M$ —incluso con más parámetros— determina una clase), pero, a cambio, sólo está definida para fórmulas metamatemáticas. ∎

Definición 12.14 Cuando consideremos como modelo una clase M (sin descartar que sea un conjunto), sin especificar la relación de pertenencia R, se entenderá que ésta es la relación de pertenencia, es decir, la clase

$$E \equiv \{(x,y) \in M \times M \mid x \in y\},$$

y en tal caso se dice que M es un *modelo natural,* y escribiremos θ^M y $M \vDash \alpha$ en lugar de θ^{MR} y $(M,R) \vDash \alpha$. Si además M es una clase transitiva, entonces se dice que M es un *modelo transitivo.*

Observemos que la clase V se define mediante la fórmula $x = x$, por lo que relativizar a V no es más que exigir a cada variable ligada que cumpla $x = x$, luego $\alpha^V \leftrightarrow \alpha$. Así pues, podemos afirmar trivialmente que en ZFC se demuestra $V \vDash \mathrm{ZFC}$, pues lo único que estamos diciendo aquí es que si α es un axioma de ZFC, entonces α^V (que es lo mismo que α) es un teorema de ZFC.

En el teorema 10.19 vimos cómo definir en MK^* una fórmula $V \vDash \alpha[v]$, para cada $\alpha \in \mathrm{Form}(\ulcorner\mathcal{L}_{\mathrm{tc}}\urcorner)$, con la cual en 10.21 demostramos que $\underset{\mathrm{MK}^*}{\vdash} \mathrm{Consis}\,\ulcorner\mathrm{ZF}^*\urcorner$ o, más en general (véanse las observaciones siguientes a 10.21) $\underset{\mathrm{MK}}{\vdash} \mathrm{Consis}\,\ulcorner\mathrm{ZFC}\urcorner$.

Si fuera posible definir en ZFC una fórmula $(M,R) \vDash \alpha[v]$ cuando M y R son clases propias y que satisfaga las propiedades que definen la satisfacción de una fórmula, el mismo argumento nos daría que $\underset{\mathrm{ZFC}}{\vdash} \mathrm{Consis}\,\ulcorner\mathrm{ZFC}\urcorner$, y el segundo teorema de incompletitud de Gödel nos daría que ZFC es contradictorio.

Más explícitamente, el argumento es que si suponemos que existe una sucesión de fórmulas $(\alpha_1, \ldots, \alpha_n) \in \mathrm{Form}(\ulcorner\mathcal{L}_{\mathrm{tc}}\urcorner)^{<\omega}$ que constituye una demostración en $\ulcorner\mathrm{ZFC}\urcorner$ de $\alpha_n = \ulcorner 0 \neq 0\urcorner$, fijada una valoración v, podemos razonar por inducción sobre $i \leq n$ que $V \vDash \alpha_i[v]$, y así, para $i = n$, llegamos a la contradicción $V \vDash \ulcorner 0 \neq 0\urcorner[v]$, que es lo mismo que $0 \neq 0$.

Ahora bien, no podemos razonar así a partir del mero hecho de que en ZFC se demuestre que V es un modelo interno de $\ulcorner\mathrm{ZFC}\urcorner$. Esto sólo significa que, para cada axioma α de ZFC, en ZFC podemos probar α^V, pero, si suponemos que $\alpha_1, \ldots, \alpha_n$ es una sucesión de fórmulas que demuestran en $\ulcorner\mathrm{ZFC}\urcorner$ la contradicción $\alpha_n = \ulcorner 0 \neq 0\urcorner$, no tiene sentido plantear que, para cada i, se cumple α_i^V, pues, metamatemáticamente, α_i no es más que un término con las variables α, i, libres, y en ese contexto α_i^V no es nada.

Así pues, al considerar modelos internos no estamos contradiciendo en nada las limitaciones que impone el segundo teorema de incompletitud.

Modelos transitivos Por razones que enseguida se comprenderán, los modelos transitivos son los más cómodos de manejar, pero antes de introducir las técnicas necesarias para tratar con ellos conviene que nos familiaricemos con la situación a través de algunas consideraciones particulares. Supongamos que una clase M es un modelo transitivo de ZF.

Consideremos por ejemplo una propiedad elemental: la inclusión. En ZF se demuestra que

$$\bigwedge xy(x \subset y \leftrightarrow \bigwedge u(u \in x \to u \in x)).$$

Por lo tanto, si en ZF se demuestra que $M \vDash \mathrm{ZF}$, en ZF se tiene que demostrar también la relativización a la clase M de la sentencia anterior, es decir,

$$\bigwedge xy \in M(x \subset^M y \leftrightarrow \bigwedge u \in M(u \in x \to u \in x)).$$

Y ahora es el momento en que la transitividad de M representa su papel: si $u \in x \in M$, necesariamente $u \in M$, luego es redundante exigirlo y la sentencia anterior puede simplificarse a

$$\bigwedge xy \in M(x \subset^M y \leftrightarrow \bigwedge u(u \in x \to u \in x)).$$

Y la simplificación es fundamental, pues ahora podemos pasar a

$$\bigwedge xy \in M(x \subset^M y \leftrightarrow x \subset y).$$

Tenemos así que dos propiedades que en principio son distintas, a saber, "ser un subconjunto de" y "ser un subconjunto^M de", en realidad son la misma: dados dos conjuntos de M, se cumple en M que uno está contenido en el otro si y sólo si realmente uno está contenido en el otro. Esto se expresa diciendo que la inclusión es absoluta para modelos transitivos de ZF (de hecho, en este caso, para clases transitivas cualesquiera, pues el "teorema" de ZF que hemos usado es en realidad la definición de inclusión y no requiere que M satisfaga ningún axioma de ZF).

Tomemos ahora $x, y \in M$. Un teorema de ZF (un axioma, de hecho) afirma que, dados x, y, existe el conjunto $\{x, y\}$ que los contiene a ambos. Esto tiene que cumplirse en M, y el objeto que en M cumple la definición de $\{x, y\}$ es el que llamamos $\{x, y\}^M$, es decir, el objeto definido como $\{x, y\}$, pero restringiendo todos los cuantificadores a M. En lugar de analizar la descripción correspondiente, vamos a analizar su definición: sabemos que en ZF se demuestra

$$\bigwedge xy \bigwedge u(u \in \{x, y\} \leftrightarrow u = x \vee u = y),$$

luego en ZF se tiene que demostrar la relativización a la clase M de la fórmula anterior, es decir,

$$\bigwedge xy \in M \bigwedge u \in M(u \in \{x, y\}^M \leftrightarrow u = x \vee u = y).$$

De momento no sabemos qué conjunto es $\{x, y\}^M$, pero podemos asegurar que $\{x, y\}^M \in M$ (porque es el objeto que en M hace el papel de par desordenado[3] de x e y). Usando de nuevo la transitividad de M vemos que la fórmula anterior equivale a

$$\bigwedge xy \in M \bigwedge u(u \in \{x, y\}^M \leftrightarrow u = x \vee u = y),$$

pues tanto $u \in \{x, y\}^M \in M$ como $u = x \vee u = y$ implican ya $u \in M$ (en el primer caso por transitividad), luego es redundante exigirlo. Y así podemos concluir que

$$\bigwedge xy \in M\ \{x, y\}^M = \{x, y\}.$$

En definitiva, el objeto que en M representa el papel de par desordenado $\{x, y\}$ es el propio $\{x, y\}$. En particular tenemos que $\bigwedge xy \in M\ \{x, y\} \in M$. Expresaremos esto diciendo que el concepto de par ordenado es absoluto para modelos transitivos de ZF (en realidad de ZF*, pues el único teorema en que nos hemos apoyado es el axioma del par y se cumple el todo modelo de ZF*).

Lo mismo vale para los pares ordenados: como en ZF se demuestra

$$\bigwedge xy \bigwedge u(u \in (x, y) \leftrightarrow u = \{x\} \vee u = \{x, y\}),$$

en M debe cumplirse la relativización de este teorema, que, teniendo en cuenta que ya sabemos que la relativización de $\{x, y\}$ es trivial, se reduce a

$$\bigwedge xy \in M \bigwedge u \in M(u \in (x, y)^M \leftrightarrow u = \{x\} \vee u = \{x, y\}).$$

Nuevamente, puesto que $(x, y)^M \in M$, la transitividad de M nos permite eliminar la cota de $\bigwedge u \in M$, y así concluimos que

$$\bigwedge xy \in M\ (x, y)^M = (x, y).$$

Una vez más, el objeto que en M hace el papel de par ordenado de x e y es el propio (x, y), y en particular tenemos que $\bigwedge xy \in M\ (x, y) \in M$.

[3] Técnicamente, por el teorema 12.10.

Consideremos ahora el conjunto de partes. En ZF se demuestra

$$\bigwedge x \bigwedge u(u \in \mathcal{P}x \leftrightarrow u \subset x).$$

Por lo tanto, también se tiene que cumplir la relativización de esta sentencia:

$$\bigwedge x \in M \bigwedge u \in M(u \in \mathcal{P}^M x \leftrightarrow u \subset x).$$

Observemos que no hemos relativizado la inclusión, pues ya hemos visto que es absoluta. Ahora la transitividad de M no puede aplicarse por completo pues, aunque $u \in \mathcal{P}^M x$ implica ciertamente que $u \in M$, no podemos deducir lo mismo que $u \subset x$, por lo que sólo podemos simplificar a medias y escribir

$$\bigwedge x \in M \bigwedge u(u \in \mathcal{P}^M x \leftrightarrow u \in M \wedge u \subset x).$$

De aquí concluimos que $\bigwedge x \in M\ \mathcal{P}^M x = M \cap \mathcal{P}x$, de modo que no podemos afirmar en general que $\mathcal{P}x$ sea un concepto absoluto. Puede ocurrir (y más adelante veremos que ocurre ciertamente en algunos casos) que el objeto $\mathcal{P}^M x$ que representa en M el papel de conjunto de partes de x sea menor que "el auténtico" $\mathcal{P}x$.

Vamos a ver enseguida que los conceptos conjuntistas básicos son absolutos para modelos transitivos de ZF (y esto es precisamente lo que hace más útiles a estos modelos) ∎

Definición 12.15 Dadas dos clases $M \subset N$, una expresión $\theta(x_1, \ldots, x_n)$ de $\mathcal{L}_{\mathrm{tc}}$ es *absoluta* para $M - N$ si

$$\bigwedge x_1 \cdots x_n \in M(\theta^M(x_1, \ldots, x_n) = \theta^N(x_1, \ldots, x_n))$$

si θ es un término o

$$\bigwedge x_1 \cdots x_n \in M(\theta^M(x_1, \ldots, x_n) \leftrightarrow \theta^N(x_1, \ldots, x_n))$$

si θ es una fórmula.

Diremos que θ es *absoluta* para M cuando lo es para $M - V$.

Ya hemos tenido ocasión de comprobar que la mayoría de los conceptos conjuntistas básicos son Δ_1 en la jerarquía de Lévy (véase la sección A.3), y el teorema siguiente muestra que todos ellos son absolutos:

Teorema 12.16 *Si M es un modelo transitivo de una teoría T y $\varnothing \in M$, entonces toda fórmula de tipo Δ_1 en T es absoluta para M.*

Demostración: Lo probamos primero para fórmulas α de tipo Δ_0 por inducción sobre su longitud. Si $\alpha \equiv x = y$ o $\alpha \equiv x \in y$, su relativización es $\alpha^M \equiv \alpha$, luego α es trivialmente absoluta. Si $\alpha \equiv \neg\beta$ y β es absoluta para M, esto significa que

$$\bigwedge x_1 \cdots x_n \in M(\alpha^M(x_1, \ldots, x_n) \leftrightarrow \alpha(x_1, \ldots, x_n)),$$

y es obvio que lo mismo vale para $\neg\alpha$. Igualmente se razona el caso $\alpha \equiv \gamma \to \delta$.

Si $\alpha \equiv \bigwedge x \in x_i\ \beta(x, x_1, \ldots, x_n)$, entonces por hipótesis de inducción

$$\bigwedge x x_1 \cdots x_n \in M(\beta^M(x, x_1, \ldots, x_n) \leftrightarrow \beta(x, x_1, \ldots, x_n)),$$

de donde

$$\bigwedge x x_1 \cdots x_n \in M((x \in x_i \to \beta^M(x, x_1, \ldots, x_n)) \leftrightarrow (x \in x_i \to \beta(x, x_1, \ldots, x_n))).$$

En este punto usamos la transitividad de M, en virtud de la cual $x \in x_i$ ya implica $x \in M$, por lo que podemos escribir

$$\bigwedge x_1 \cdots x_n \in M \bigwedge x((x \in x_i \to \beta^M(x, x_1, \ldots, x_n)) \leftrightarrow (x \in x_i \to \beta(x, x_1, \ldots, x_n))).$$

de donde

$$\bigwedge x_1 \cdots x_n \in M(\bigwedge x(x \in x_i \to \beta^M(x, x_1, \ldots, x_n)) \leftrightarrow \bigwedge x \in x_i\ \beta(x, x_1, \ldots, x_n))).$$

De nuevo por la transitividad de M, esto equivale a

$$\bigwedge x_1 \cdots x_n \in M(\bigwedge x \in M(x \in x_i \to \beta^M(x, x_1, \ldots, x_n)) \leftrightarrow \bigwedge x \in x_i\ \beta(x, x_1, \ldots, x_n))),$$

que es lo mismo que $\bigwedge x_1 \cdots x_n \in M(\alpha^M \leftrightarrow \alpha)$.

Si γ es una fórmula Δ_1 en T, entonces existen fórmulas α y β de tipo Δ_0 tales que

$$\underset{T}{\vdash} \gamma \leftrightarrow \bigvee x\, \alpha, \qquad \underset{T}{\vdash} \gamma \leftrightarrow \bigwedge x\, \beta.$$

Como $M \vDash T$, esto implica que

$$\bigwedge x_1 \cdots x_n \in M(\gamma^M(x_1, \ldots, x_n) \leftrightarrow \bigvee x \in M \alpha^M(x, x_1, \ldots, x_n)),$$

$$\bigwedge x_1 \cdots x_n \in M(\gamma^M(x_1, \ldots, x_n) \leftrightarrow \bigwedge x \in M \beta^M(x, x_1, \ldots, x_n)).$$

Por lo tanto, fijados $x_1, \ldots, x_n \in M$, tenemos que

$$\gamma^M(x_1, \ldots, x_n) \to \bigvee x \in M \alpha^M(x, x_1, \ldots, x_n) \to \bigvee x \in M \alpha(x, x_1, \ldots, x_n) \to \bigvee x \alpha(x, x_1, \ldots, x_n) \to \gamma(x_1, \ldots, x_n),$$

donde hemos usado que α es absoluta para M. Usando las equivalencias con β obtenemos la implicación contraria. ∎

Observemos que si un término, como $x \cup y$, es absoluto para un modelo transitivo M, entonces, para $x, y \in M$ tenemos que $x \cup y = (x \cup y)^M \in M$, por el teorema 12.10, es decir, los modelos son cerrados para los términos absolutos, luego, en particular, si M es un modelo transitivo de ZF^*, tenemos que

$$\varnothing \in M, \quad x \cap y \in M, \quad x \cup y \in M, \quad \{x, y\} \in M, \quad (x, y) \in M, \quad x \times y \in M,$$

etc., pues todos son términos Δ_0.

Veamos ahora las condiciones que debe cumplir una clase transitiva M para satisfacer los axiomas de ZFC:

Extensionalidad La relativización del axioma de extensionalidad es

$$\bigwedge xy \in M(\bigwedge u \in M(u \in x \leftrightarrow u \in y) \to x = y).$$

Observemos que se cumple siempre que M es una clase transitiva. En efecto, esto equivale a

$$\bigwedge xy \in M(\bigwedge u(u \in x \leftrightarrow u \in y) \to x = y),$$

ya que si $u \in x \in M$, entonces $u \in M$ por transitividad, es decir, que si $x, y \in M$ tienen en común todos sus elementos que están en M, entonces tienen todos sus elementos en común (pues $x, y \subset M$), luego $x = y$.

Par La relativización del axioma del par es

$$\bigwedge xy \in M \bigvee z \in M \bigwedge u \in M(u \in z \leftrightarrow u = x \vee u = y).$$

Observemos que si M es cualquier clase transitiva esto equivale a

$$\bigwedge xy \in M \bigvee z \in M \bigwedge u(u \in z \leftrightarrow u = x \vee u = y),$$

pues tanto $u \in z$ como $u = x \vee u = y$, para $x, y, z \in M$, implican ya que $u \in M$, luego es redundante exigirlo. A su vez, esto equivale a

$$\bigwedge xy \in M\ \{x, y\} \in M.$$

Unión La relativización del axioma de la unión es

$$\bigwedge x \in M \bigvee y \in M \bigwedge u \in M(u \in y \leftrightarrow \bigvee v \in M(u \in v \wedge v \in x)).$$

Nuevamente, la mera transitividad de M permite eliminar dos acotaciones redundantes:

$$\bigwedge x \in M \bigvee y \in M \bigwedge u(u \in y \leftrightarrow \bigvee v(u \in v \wedge v \in x)),$$

y esto equivale a

$$\bigwedge x \in M\ \bigcup x \in M.$$

Infinitud Tras eliminar las acotaciones redundantes por transitividad, la relativización de AI se reduce a:

$$\bigvee x \in M(\bigvee u \in x \bigwedge v\ v \notin u \wedge \bigwedge u \in x \bigvee v \in x \bigwedge w(w \in v \leftrightarrow w \in v \vee w = v)).$$

Equivalentemente:

$$\bigvee x \in M(\varnothing \in x \wedge \bigwedge u \in x\ u' \in x).$$

Esto lo cumple cualquier clase que cumpla $\omega \in M$. La condición es también necesaria para modelos $M \vDash \mathrm{ZF}^* + \mathrm{AI} + \mathrm{V} = \mathrm{R}$, pues el término ω es Δ_0 en dicha teoría, luego es absoluto para M, luego $\omega = \omega^M \in M$.

Partes La relativización de AP es:

$$\bigwedge x \in M \bigvee y \in M \bigwedge u \in M(\bigwedge v \in M(v \in u \to v \in x) \to u \in y).$$

Al eliminar las acotaciones redundantes queda

$$\bigwedge x \in M \bigvee y \in M \bigwedge u \in M(\bigwedge v(v \in u \to v \in x) \to u \in y),$$

es decir,

$$\bigwedge x \in M \bigvee y \in M \bigwedge u \in M(u \subset x \to u \in y),$$

y esto lo cumple toda clase que cumpla

$$\bigwedge x \in M\ \mathcal{P}x \cap M \in M.$$

Nuevamente la condición es necesaria, ya que si $M \vDash \mathrm{ZF}^* + \mathrm{AP}$ hemos visto en el ejemplo previo a 12.15 que $\mathcal{P}x \cap M = \mathcal{P}^M x \in M$.

Regularidad La relativización del axioma de regularidad es

$$\bigwedge x \in M(\bigvee u \in M\ u \in x \to \bigvee u \in M(u \in x \wedge \neg \bigvee v \in M(v \in u \wedge v \in x))),$$

que se simplifica por transitividad a

$$\bigwedge x \in M(\bigvee u\ u \in x \to \bigvee u \in x \neg \bigvee v(v \in u \wedge v \in x)),$$

lo cual equivale a

$$\bigwedge x \in M(x \neq \varnothing \to \bigvee u \in x\ u \cap x = \varnothing),$$

y esto se cumple siempre que $M \subset R$.

Elección Observemos que la fórmula "f es una función de elección en x" es de tipo Δ_0, pues equivale a

$$f : x \longrightarrow V \wedge \bigwedge u \in x(u \neq \varnothing \to \bigvee v \in u\ (u, v) \in f).$$

Por lo tanto es absoluta para clases transitivas, y la relativización de AE se reduce a

$$\bigwedge x \in M \bigvee f \in M\ f \text{ es una función de elección en } x.$$

No hemos incluido el axioma de reemplazo porque la transitividad de M no da lugar a ninguna simplificación destacable. No obstante, incluimos su relativización en el resumen de la página siguiente. Ahora es fácil probar:

Teorema 12.17 (ZF*+AI) $V_\omega \vDash \ulcorner \mathrm{ZF}_{\mathrm{fin}} \urcorner$.

Relativizaciones de los axiomas de ZFC

Extensionalidad Se cumple en cualquier clase transitiva.

Par $\bigwedge xy \in M\ \{x,y\} \in M$.

Unión $\bigwedge x \in M\ \bigcup x \in M$.

Reemplazo Para toda fórmula $\phi(x, y, x_1, \ldots, x_n)$:

$$\bigwedge x_1 \cdots x_n \in M(\bigwedge xyz \in M(\phi^M(x,y) \wedge \phi^M(x,z) \to y = z)$$
$$\to \bigwedge a \in M \bigvee b \in M \bigwedge y(y \in b \leftrightarrow \bigvee x \in a\, \phi^M(x,y))).$$

Infinitud $\omega \in M$ (y entonces $\omega^M = \omega$).

Partes $\bigwedge x \in M\ \mathcal{P}x \cap M \in M$ (y entonces $\mathcal{P}^M x = M \cap \mathcal{P}x$).

Regularidad Se cumple siempre que $M \subset R$. En particular, si $V = R$ se cumple en toda clase transitiva.

Elección $\bigwedge x \in M \bigvee f \in M\ f$ es una función de elección en x.

Demostración: Es inmediato que V_ω cumple las condiciones anteriores para los axiomas de extensionalidad, par, unión y regularidad. Por el teorema 12.13, tenemos que lo mismo vale considerando estos axiomas como elementos de $\mathrm{Form}(\ulcorner\mathcal{L}_{\mathrm{tc}}\urcorner)$ en lugar de como fórmulas metamatemáticas. Si probamos que cumple también todos los casos del esquema de reemplazo tendremos demostrado que $V_\omega \vDash \ulcorner\mathrm{ZF}^*\urcorner$. Concretamente, tomamos $\phi \in \mathrm{Form}(\ulcorner\mathcal{L}_{\mathrm{tc}}\urcorner)$ y hemos de probar que

$$V_\omega \vDash (\bigwedge xyz(\phi(x,y) \wedge \phi(x,z) \to y = z)$$
$$\to \bigwedge a \bigvee b \bigwedge y(y \in b \leftrightarrow \bigvee x \in a\, \phi(x,y)))[v],$$

donde v es cualquier valoración definida sobre todas las variables que aparecen en la fórmula. Para ello suponemos

$$V_\omega \vDash \bigwedge xyz(\phi(x,y) \wedge \phi(x,z) \to y = z)[v],$$

es decir, que para todo $u, v, w \in V_\omega$, si $V_\omega \vDash \phi[v^{uv}_{xy}]$ y $V_\omega \vDash \phi[v^{uw}_{xy}]$, entonces $v = w$. Fijamos un conjunto $p \in V_\omega$ y hemos de encontrar otro $q \in V_\omega$ de modo que

$$V_\omega \vDash \bigwedge y(y \in b \leftrightarrow \bigvee x \in a\, \phi(x,y))[v^{pq}_{ab}],$$

lo cual, teniendo en cuenta la transitividad de V_ω, equivale a

$$\bigwedge v \in V_\omega(v \in q \leftrightarrow \bigvee u \in p\ V_\omega \vDash \phi[v^{uv}_{xy}]).$$

Y ahora basta aplicar el axioma de reemplazo a la fórmula $V_\omega \vDash \phi[v^{uv}_{xy}]$.

Falta probar que V_ω cumple CT y que todo conjunto es finito. De nuevo por el teorema 12.13 podemos considerar las fórmulas metamatemáticas correspondientes. Para CT hemos de probar que

$$\bigwedge x \in V_\omega \bigvee y \in V_\omega (x \subset y \wedge \bigcup y \subset y),$$

donde hemos usado que la unión y la inclusión son absolutas para modelos transitivos. Dado $x \in V_\omega$, existe un $n \in \omega$ tal que $x \in V_n$, y basta tomar $y = V_n \in V_{n+1} \subset V_\omega$.

Finalmente, si $x \in V_\omega$, entonces x es finito, luego existe $n \in \omega$ y $f : n \longrightarrow x$ biyectiva, y $n \times x \in V_\omega$, luego existe un $m \in \omega$ tal que $f \subset n \times x \in V_m$, luego $f \subset V_m$, luego $f \in V_{m+1} \subset V_\omega$, y entonces $(n \in \omega \wedge f : n \longrightarrow x \text{ biyectiva})^{V_\omega}$ porque la fórmula es Δ_0, luego es absoluta. Esto significa que $(x \text{ es finito})^{V_\omega}$, luego V_ω satisface la sentencia "todo conjunto es finito". ∎

Nota Es fácil comprobar que la prueba del teorema anterior sirve igualmente en KPI. Para aplicar el esquema de reemplazo 6.11 necesitamos cambiar ϕ por

$$\psi(x, y) \equiv \phi(x, y) \vee (\bigwedge z \neg\phi(x, z) \wedge y = w)$$

y tomar una valoración v tal que $V_\omega \vDash \phi[v_{xy}^{uv(w)}]$, para cierto $u \in p$. (Si esto es imposible sirve $q = \varnothing$.) ∎

Ejercicio: Probar en ZF que $V_{\omega+\omega} \vDash \mathrm{Z} + \bigwedge \alpha \in \Omega \bigvee n \in \omega(\alpha = n \vee \alpha = \omega + n)$. Por lo tanto, si ZF es consistente, en Z no puede probarse la existencia del ordinal $\omega + \omega$.

Veamos ahora que si ZFC−AP es consistente, también lo es $\mathrm{ZFC}_{\mathrm{num}}$, es decir, ZFC menos el axioma de partes más el axioma "todo conjunto es numerable" o, en otros términos, que sin el axioma de partes no es posible demostrar la existencia de conjuntos no numerables. Para ello, trabajamos en ZFC-AP:

Definición 12.18 Un conjunto x es *hereditariamente numerable* si $\mathrm{ct}\, x$ es numerable. Llamaremos HN a la clase de todos los conjuntos hereditariamente numerables.

Basta probar:

Teorema 12.19 (ZFC−AP) HN *es un modelo transitivo de* $\mathrm{ZFC}_{\mathrm{num}}$

Demostración: Observemos que la clase HN es transitiva, pues si se cumple $x \in y \in \mathrm{HN}$, entonces $x \in \mathrm{ct}\, y$, luego $x \subset \mathrm{ct}\, y$, luego $\mathrm{ct}\, x \subset \mathrm{ct}\, y$, luego la numerabilidad de $\mathrm{ct}\, y$ implica la de $\mathrm{ct}\, x$, luego $x \in \mathrm{HN}$.

Observemos que si $x \subset \mathrm{HN}$ es numerable, entonces $x \in \mathrm{HN}$, pues por el teorema 11.25 tenemos que

$$\mathrm{ct}\, x = x \cup \bigcup_{u \in x} \mathrm{ct}\, u,$$

y por el teorema 11.55 tenemos que $\mathrm{ct}\, x$ es numerable.

La transitividad de HN implica que cumple el axioma de extensionalidad y el de regularidad (pues suponemos $V = R$). Para probar el axioma del par tomamos $x, y \in \mathrm{HN}$ y observamos que, por 11.25,

$$\mathrm{ct}\,\{x, y\} = \{x, y\} \cup \mathrm{ct}\,x \cup \mathrm{ct}\,y,$$

luego es numerable y $\{x, y\} \in \mathrm{HN}$.

Para el axioma de la unión tomamos $x \in \mathrm{HN}$ y notamos que $\bigcup x \subset \mathrm{HN}$ (por transitividad) y es numerable (por ser unión numerable de conjuntos numerables), luego $\bigcup x \in \mathrm{HN}$.

Para el axioma de reemplazo suponemos que

$$\bigwedge xyz \in \mathrm{HN}(\phi^{\mathrm{HN}}(x, y) \wedge \phi^{\mathrm{HN}}(x, z) \to y = z),$$

y tomamos $a \in \mathrm{HN}$. Por reemplazo existe $b = \{y \mid y \in \mathrm{HN} \wedge \bigvee x \in a\ \phi^{\mathrm{HN}}(x, y)\}$, de modo que $b \subset \mathrm{HN}$ y es numerable, pues $\{(x, y) \in a \times b \mid \phi^{\mathrm{HN}}(x, y)\}$ es una aplicación suprayectiva de un subconjunto de a en b, luego existe una aplicación inyectiva de b en un subconjunto de a. Por lo tanto $b \in \mathrm{HN}$ y es claro que cumple lo requerido por el axioma de reemplazo.

Claramente $\omega = \mathrm{ct}\,\omega \in \mathrm{HN}$, luego se cumple el axioma de infinitud. Si $x \in \mathrm{HN}$ y $f : x \longrightarrow \bigcup x$ es una función de elección, entonces $f \subset x \times \bigcup x \in \mathrm{HN}$, luego $f \subset \mathrm{HN}$ y es numerable, pues toda aplicación se puede biyectar con su dominio. Por lo tanto $f \in \mathrm{HN}$, y esto prueba el axioma de elección en HN.

Similarmente, si $x \in \mathrm{HN}$, entonces existe $f : x \longrightarrow \omega$ inyectiva, y $f \in \mathrm{HN}$ por el mismo razonamiento aplicado a la función de elección, y como ω es absoluto, tenemos que $(f : x \longrightarrow \omega \text{ inyectiva})^{\mathrm{HN}}$. Esto prueba que (todo conjunto es numerable)$^{\mathrm{HN}}$. ∎

Observaciones Notemos que el teorema anterior justifica realmente lo que hemos afirmado justo antes. En general, si en una teoría T (por ejemplo ZFC−AP) podemos demostrar que una clase (en este caso HN) es un modelo de otra teoría S (en este caso $\mathrm{ZFC}_{\mathrm{num}}$), tenemos una prueba constructiva de que si T es consistente también lo es S. En efecto, si tenemos una prueba de una contradicción en S, podemos prolongarla hasta una prueba de que $\bigvee x\ x \neq x$, por ejemplo, pero la prueba del teorema 12.12 nos dice explícitamente cómo construir una prueba en T de $\bigvee x \in M\ x \neq x$, con lo que nos encontramos con que T también es contradictoria.

En nuestro caso en concreto, si alguien demostrara en ZFC−AP que existe un conjunto no numerable, relativizando a HN la prueba punto por punto llegaríamos a una prueba en ZFC−AP de que (existe un conjunto no numerable)$^{\mathrm{HN}}$, mientras que hemos probado en ZFC−AP que (todo conjunto es numerable)$^{\mathrm{HN}}$, luego tenemos una contradicción en ZCF−AP.

Desde un punto de vista semántico, el teorema anterior prueba que si de un modelo de ZFC−AP eliminamos todos los conjuntos no hereditariamente numerables, lo que nos queda sigue siendo un modelo de ZFC−AP, donde además

no hay conjuntos no numerables. Por eso no puede probarse la existencia de conjuntos no numerables sin AP, porque, si pudiera probarse, tendría que haber conjuntos no numerables en todo modelo de ZFC−AP, y no es así, sino que, en caso de que los haya, se pueden eliminar. ∎

En ZFC tenemos que si $x \in \mathrm{HN}$, entonces $\operatorname{rang} x \in \mathrm{HN}$ (pues el rango es Δ_1, luego absoluto para modelos transitivos), luego $\operatorname{rang} x$ es un ordinal numerable. Ahora bien, en ZFC puede probarse la existencia de conjuntos no numerables, así como que todo conjunto puede biyectarse con un ordinal. Por lo tanto, podemos considerar el menor ordinal no numerable, que habitualmente se representa por ω_1, y acabamos de justificar que $\mathrm{HN} \subset V_{\omega_1}$, luego la clase HN es un conjunto, y el teorema 12.13 junto con una mínima variante de la prueba del teorema anterior en el caso del esquema de reemplazo (para probarlo para fórmulas $\phi \in \mathrm{Form}(\ulcorner \mathcal{L}_{\mathrm{tc}} \urcorner)$ en lugar de para fórmulas metamatemáticas) nos permite concluir que

$$\mathrm{HN} \vDash \ulcorner \mathrm{ZFC}_{\mathrm{num}} \urcorner,$$

luego concluimos que

$$\underset{\mathrm{ZFC}}{\vdash} \mathrm{Consis}\, \ulcorner \mathrm{ZFC}_{\mathrm{num}} \urcorner.$$

12.3 Consistencia e independencia del axioma de regularidad

Al introducir las teorías de conjuntos hemos separado de las teorías básicas ZF^* y NBG^*, que permiten demostrar las propiedades más elementales sobre clases y conjuntos, de los axiomas que aumentan significativamente la potencia de la teoría: el axioma de infinitud marca la diferencia entre la existencia o no de conjuntos infinitos (y también la diferencia entre las teorías cuya consistencia podemos demostrar fácilmente y las teorías de las que no conocemos una prueba de consistencia metamatemáticamente aceptable), el axioma de partes supone la introducción en la teoría de conjuntos no numerables, carentes de un contenido intuitivo preciso, y el axioma de elección introduce en la teoría conjuntos no definibles explícitamente.

También hemos incluido en este grupo al axioma de regularidad, pues implica propiedades estructurales muy fuertes sobre la clase de todos los conjuntos (ya que la estratifica en la jerarquía de conjuntos $\{V_\alpha\}_{\alpha\in\Omega}$, justifica la $\in$-inducción y la $\in$-recursión, etc.) Sin embargo, vamos a ver aquí que aceptar el axioma de regularidad dista mucho de ser un hecho relevante, en el sentido que lo es aceptar cualquiera de los otros tres axiomas, sino que se trata de un axioma trivial que podemos aceptar por conveniencia, aunque podríamos trabajar exactamente igual sin él. En realidad ya hemos discutido esta idea, pero ahora estamos en condiciones de precisarla en términos de modelos.

En esencia, vamos a demostrar la consistencia y la independencia del axioma de regularidad de los axiomas restantes de la teoría de conjuntos. La primera

equivale a que no es posible demostrar que existan conjuntos no regulares, mientras que la segunda equivale a que el axioma de regularidad no puede demostrarse a partir de los demás axiomas. La primera es una prueba simple, pero arquetípica de las pruebas de consistencia en teoría de conjuntos mediante modelos internos (transitivos), mientras que la segunda ilustra, también en un caso relativamente sencillo, el uso de modelos no naturales.

Teorema 12.20 *En* ZF^* (+ AI, + AP, + AE) *se demuestra que la clase* R *de los conjuntos regulares es un modelo transitivo de* $\mathrm{ZF}^* + V = R$ (+ AI, + AP, + AE).

Demostración: En la sección anterior hemos visto que el axioma de extensionalidad se cumple en cualquier clase, y el de regularidad se cumple porque $R \subset R$. El del par se cumple porque si $x, y \in R$, entonces $\{x, y\} \in \mathcal{P}R = R$, e igualmente sucede con el de la unión, pues si $x \in R$ entonces todo $u \in \bigcup x$ cumple $u \in v \in x \in R$, para cierto v, luego $u \in R$, luego $\bigcup x \in \mathcal{P}R = R$.

Para probar el esquema de reemplazo, fijados parámetros $x_1, \ldots, x_n \in R$, suponemos que $\bigwedge xyz \in R(\phi^R(x, y) \wedge \phi^R(x, z) \to y = z)$ y tomamos un $a \in R$. Aplicando el axioma de reemplazo a la fórmula

$$\psi(x, y) \equiv x \in R \wedge y \in R \wedge \phi^R(x, y)$$

concluimos que existe un b tal que

$$\bigwedge y(y \in b \leftrightarrow y \in R \wedge \bigvee x \in a\ \phi^R(x, y)).$$

En particular $b \in \mathcal{P}R = R$ y cumple lo requerido.

Si suponemos AI, entonces se cumple AI^R porque $\omega \in R$.

Si suponemos AP, entonces, dado $x \in R$, tenemos que $\mathcal{P}x \cap R \in \mathcal{P}R = R$, luego se cumple AP^R.

Finalmente suponemos AE y vamos a probar AE^R. Si $x \in R$ y f es una función de elección en x, podemos suponer que, en caso de que $\varnothing \in x$, se cumple $f(\varnothing) \in x$. Así, $f \subset x \times \bigcup x \subset R$, luego $f \in \mathcal{P}R = R$, y esto prueba AE^R. ∎

Como consecuencia inmediata:

Teorema 12.21 *Si la teoría* ZF^*+ AI (+ AP + AE) *es consistente, también lo es* $\mathrm{ZF}^* + V = R$ + AI (+ AP + AE)*, es decir, no puede demostrarse la existencia de conjuntos no regulares.*

Observaciones La situación es la misma que en la prueba de la consistencia de $\mathrm{ZFC}_{\mathrm{num}}$: Si alguien demostrara en $\mathrm{ZFC}{-}V = R$ que existe un conjunto no regular, relativizando a R la prueba punto por punto llegaríamos a una prueba en $\mathrm{ZFC}{-}V = R$ de que (existe un conjunto no regular)^R, mientras que hemos demostrado en $\mathrm{ZFC}{-}V = R$ que (todo conjunto es regular)^R, luego tenemos una contradicción en $\mathrm{ZCF}{-}V = R$.

Desde un punto de vista semántico, si en ZF pudiera probarse que existe un conjunto no regular, ello significaría que en todo modelo de ZF tendría que haber conjuntos no regulares, pero al demostrar que $R \vDash \mathrm{ZF}$, lo que hemos probado es que si un modelo tiene conjuntos no regulares y los eliminamos, es decir, pasamos a considerar únicamente sus conjuntos regulares, los conjuntos resultantes siguen siendo un modelo de ZF y además cumplen $V = R$, de modo que los posibles conjuntos no regulares son siempre prescindibles.

Dicho de otro modo: los conjuntos no regulares no son necesarios para que se cumplan los axiomas de ZF: si en un modelo hay conjuntos no regulares y hacemos que un matemático "vea" sólo los conjuntos regulares, no echará por ello nada en falta: si ve dos conjuntos, verá también su unión, y su intersección, y su conjunto de partes, etc., porque todas estas operaciones restringidas a conjuntos regulares dan conjuntos regulares, y, en general, el matemático "verá" todos los conjuntos que los axiomas de ZF le aseguran que tienen que estar ante su vista.

O también: aceptar el axioma de regularidad no supone exigir a los conjuntos propiedades más fuertes que si no lo aceptamos, sino simplemente descartar los posibles conjuntos no regulares de un modelo y llamar "conjuntos" a lo que, si no incluyéramos el axioma en la teoría, llamaríamos "conjuntos regulares". Es sólo una forma de decir "los conjuntos no regulares no nos interesan para nada, así que cuando hablemos de conjuntos, nos referimos siempre a los conjuntos regulares".

La prueba de la consistencia del axioma de regularidad (debida a von Neumann) sigue el mismo esquema que la prueba de Gödel de la consistencia del axioma de elección (que no veremos en este libro): Gödel definió la clase L de los *conjuntos constructibles,* que representa un papel análogo al de R en la prueba anterior: no puede probarse que $V = L$ (esto se conoce como *axioma de constructibilidad*, pero en ZF se prueba que L es un modelo de ZFC+$V = L$, de modo que si "eliminamos" los posibles conjuntos no constructibles, los que nos quedan siguen cumpliendo los axiomas de ZF y además cumplen el axioma de elección o, en términos sintácticos: si se pudiera probar una contradicción en ZFC, al relativizarla a L tendríamos la prueba de una contradicción en ZF. ∎

A continuación probaremos que sin el axioma de regularidad tampoco se puede demostrar que no existan conjuntos no regulares. Notemos que no podemos demostrar (por ejemplo en ZF$-V = R$) la existencia de un modelo transitivo M en el que se cumpla $V \neq R$, pues ello exigiría que $R \subsetneq M \subset V$, y con ello habríamos demostrado que existen conjuntos no regulares, cuando ya hemos probado que eso es imposible. Por consiguiente, el modelo que obtendremos no será transitivo y, de hecho, no será natural. En realidad vamos a mostrar una técnica general para construir modelos en los que falla el axioma de regularidad:

Teorema 12.22 (ZF*) *Sea $F : V \longrightarrow V$ biyectiva y $R = \{(x, y) \mid x \in F(y)\}$. Entonces $(V, R) \vDash \mathrm{ZF}^*$. Si suponemos* AI, AP *o* AE*, estos axiomas se cumplen también en (V, R).*

DEMOSTRACIÓN: Notemos que nada de lo que hemos visto para modelos transitivos nos ayuda ahora, de modo que hemos de calcular explícitamente las relativizaciones de los axiomas y demostrarlas. Convenimos en que las letras minúsculas representan conjuntos. Observemos que relativizar a (V, R) consiste simplemente en cambiar $\in$ por R.

EXTENSIONALIDAD: $\bigwedge xy(\bigwedge u(u\,R\,x \leftrightarrow u\,R\,y) \to x = y)$.

En efecto, la hipótesis es $\bigwedge u(u \in F(x) \leftrightarrow u \in F(y))$, lo cual implica claramente $F(x) = F(y)$, luego $x = y$.

PAR: $\bigwedge xy\bigvee z\bigwedge u(u\,R\,z \leftrightarrow u = x \vee u = y)$. Esto equivale a

$$\bigwedge xy\bigvee z\bigwedge u(u \in F(z) \leftrightarrow u \in \{x, y\}).$$

Basta tomar $z = F^{-1}(\{x, y\})$.

UNIÓN: $\bigwedge x\bigvee y\bigwedge u(u\,R\,y \leftrightarrow \bigvee v(u\,R\,v \wedge v\,R\,x))$. Esto equivale a

$$\bigwedge x\bigvee y\bigwedge u(u \in F(y) \leftrightarrow \bigvee v \in F(x)\ u \in F(v))$$

Basta tomar $y = F^{-1}\Big(\bigcup_{v\in F(x)} F(v)\Big)$.

VACÍO: $\bigvee x\bigwedge y\neg x\,R\,y$. Basta tomar $x = F^{-1}(\varnothing)$.

REEMPLAZO: Fijada una fórmula $\phi(x, y)$ quizá con otros parámetros, hemos de probar

$$\bigwedge x_1 \cdots x_n(\bigwedge xyz(\phi^{VR}(x, y) \wedge \phi^{VR}(x, z) \to y = z)$$
$$\to \bigwedge a\bigvee b\bigwedge y(y\,R\,b \leftrightarrow \bigvee x(x\,R\,a \wedge \phi^{VR}(x, y))).$$

Fijados $x_1, \ldots, x_n \in \mathrm{V}$, definimos la función $G : A \subset V \longrightarrow V$ mediante $G(x) = y \leftrightarrow \phi^{VR}(x, y)$. Hemos de probar que

$$\bigwedge a\bigvee b\bigwedge y(y \in F(b) \leftrightarrow \bigvee x \in F(a)\ y = G(x)).$$

Basta tomar $b = F^{-1}(G[F(a)])$.

PARTES: $\bigwedge x\bigvee y\bigwedge u(u\,R\,y \leftrightarrow \bigwedge v(v\,R\,u \to v\,R\,x))$. Esto equivale a

$$\bigwedge x\bigvee y\bigwedge u(u \in F(y) \leftrightarrow F(u) \subset F(x))$$

Basta tomar $y = F^{-1}(F^{-1}[\mathcal{P}F(x)])$.

INFINITUD: El axioma de infinitud es:

$$\bigvee x(\bigvee y(y \in x \wedge \bigwedge z\ z \notin y) \wedge \bigwedge y(y \in x \to$$
$$\bigvee z(z \in x \wedge \bigwedge u(u \in z \leftrightarrow u \in y \vee u = y)))).$$

La relativización es

$$\bigvee x(\bigvee y(y \in F(x) \wedge \bigwedge z\ z \notin F(y)) \wedge \bigwedge y(y \in F(x) \to$$
$$\bigvee z(z \in F(x) \wedge \bigwedge u(u \in F(z) \leftrightarrow u \in F(y) \vee u = y)))).$$

A su vez esto equivale a

$$\bigvee x(F^{-1}(\varnothing) \in F(x) \wedge \bigwedge y(y \in F(x) \to F^{-1}(F(y) \cup \{y\}) \in F(x))).$$

Como F es biyectiva, esto equivale a

$$\bigvee x(F^{-1}(\varnothing) \in x \wedge \bigwedge y(y \in x \to F^{-1}(F(y) \cup \{y\}) \in x)).$$

Definimos $y_0 = F^{-1}(\varnothing)$ y $\bigwedge n \in \omega\ y_{n+1} = F^{-1}(F(y_n) \cup \{y_n\})$. Es claro que $x = \{y_n \mid n \in \omega\}$ cumple lo pedido.

Elección: Con los axiomas ya probados el axioma de elección equivale a que para toda familia formada por conjuntos no vacíos disjuntos dos a dos existe un conjunto que contiene exactamente un elemento en común con cada elemento de la familia. Explícitamente:

$$\bigwedge x(\bigwedge u(u \in x \to \bigvee v\ v \in u) \wedge \bigwedge uv(u \in x \wedge v \in x \wedge u \neq v \to$$

$$\neg\bigvee z(z \in u \wedge z \in v)) \to \bigvee y\bigwedge u(u \in x \to \overset{1}{\bigvee} v(v \in y \wedge v \in u))).$$

La relativización es

$$\bigwedge x(\bigwedge u(u \in F(x) \to \bigvee v\ v \in F(u)) \wedge \bigwedge uv(u \in F(x) \wedge v \in F(x) \wedge u \neq v \to$$

$$\neg\bigvee z(z \in F(u) \wedge z \in F(v))) \to \bigvee y\bigwedge u(u \in F(x) \to \overset{1}{\bigvee} v(v \in F(y) \wedge v \in F(u)))).$$

Esto equivale a

$$\bigwedge x(\bigwedge u(u \in F(x) \to F(u) \neq \varnothing) \wedge \bigwedge uv \in F(x)(u \neq v \to F(u) \cap F(v) = \varnothing)$$

$$\to \bigvee y\bigwedge u \in F(x)\overset{1}{\bigvee} v \in F(y) \cap F(u)).$$

Usando que F es biyectiva esto equivale a

$$\bigwedge x(\bigwedge u(u \in x \to F(u) \neq \varnothing) \wedge \bigwedge uv \in x(u \neq v \to F(u) \cap F(v) = \varnothing)$$

$$\to \bigvee y\bigwedge u \in x\overset{1}{\bigvee} v \in y \cap F(u)),$$

lo cual se obtiene aplicando el axioma de elección (en la forma que estamos considerando) al conjunto $F[x]$. ∎

Considerando biyecciones adecuadas F podemos violar de mil maneras el axioma de regularidad. El ejemplo más simple es el siguiente:

Teorema 12.23 *Si* ZFC *es consistente también lo es* ZFC *menos el axioma de regularidad y más el axioma* $\bigvee a\ a = \{a\}$.

DEMOSTRACIÓN: Basta tomar la biyección $F : V \longrightarrow V$ definida mediante $F(0) = \{0\}$, $F(\{0\}) = 0$ y $F(x) = x$ en otro caso. Así, tomando $a = 0$, el modelo construido en el teorema anterior cumple

$$\bigwedge x(x\, R\, a \leftrightarrow x = a),$$

es decir, $(a = \{a\})^{\mathrm{V}R}$. ∎

Ejercicio: Demostrar la consistencia de que existan dos conjuntos x, y tales que $x = \{y\} \wedge y = \{x\}$.

Los conjuntos de la forma $a = \{a\}$ se llaman *átomos*. Vamos a probar que, de hecho, es consistente la existencia de infinitos átomos. La prueba nos obliga a tratar más a fondo con los modelos (V, R) que estamos considerando.

Teorema 12.24 *Si* ZFC *es consistente, también lo es* ZFC *sin el axioma de regularidad y más el axioma que afirma la existencia de un conjunto (infinito) numerable*[4] *de átomos.*

DEMOSTRACIÓN: Sea $F : V \longrightarrow V$ la aplicación dada por

$$F(x) = \begin{cases} \{\{\{n+1\}\}\} & \text{si } x = \{\{n+1\}\} \text{ con } n \in \omega, \\ \{\{n+1\}\} & \text{si } x = \{\{\{n+1\}\}\}, \text{ con } n \in \omega, \\ x & \text{en otro caso.} \end{cases}$$

Por abreviar representaremos $x_n \equiv \{\{n+1\}\}$, $y_n \equiv \{\{\{n+1\}\}\} = \{x_n\}$, para $n \in \omega$, de modo que lo que hace F es intercambiar cada x_n con y_n. La razón de tantas llaves es asegurar que $\bigwedge n \in \omega\ F(n) = n$, lo cual simplificará el argumento. Notemos que F fija a todos los conjuntos con más de un elemento.

Claramente, si $A = \{x_n \mid n \in \omega\}$, se cumple que $\bigwedge x \in A\ F(x) = \{x\}$, luego $\bigwedge x \in A \bigwedge u(u\, R\, x \leftrightarrow u = x)$, que, como $F(A) = A$, es lo mismo que

$$(\bigwedge x \in A(\bigwedge u(u \in x \leftrightarrow u = x)))^{VR}$$

o también $(\bigwedge x \in A(x = \{x\}))^{VR}$. Así pues, A es un conjunto de átomos en el modelo (V, R). Vamos a ver que es numerable^{VR}. (Obviamente A es numerable, pero no es eso lo que necesitamos, sino que

$$(\bigvee f\ f : \omega \longrightarrow A \text{ biyectiva})^{VR},$$

y tenemos que desarrollar esta fórmula para comprobar que realmente se cumple.)[5]

[4] No hemos desarrollado la teoría de cardinales cantoriana, pero el argumento se adapta fácilmente para probar la consistencia de que exista un conjunto de átomos de cualquier cardinal prefijado, $\aleph_1$, $\aleph_2$, etc.

[5] Informalmente, este proceso de relativizar consiste en "meterse en la piel" de alguien que "creyera" que la pertenencia entre conjuntos no es $\in$, sino R y estar al tanto de todas las "confusiones" (o no confusiones) a que esto le induciría. Por ejemplo, hemos de preguntarnos, ¿qué conjunto tomaría por el conjunto vacío alguien así? ¿y a qué conjunto llamaría ω? etc.

La forma de calcular la relativización de un concepto es relativizar un teorema que lo caracterice. Por ejemplo, como en ZFC se prueba que $\bigwedge x\ x \notin \varnothing$, también podemos probar su relativización, que es $\bigwedge x\ \neg x\, R\, \varnothing^{VR}$, que a su vez es lo mismo que $\bigwedge x\ x \notin F(\varnothing^{VR})$, luego $F(\varnothing^{VR}) = \varnothing = F(\varnothing)$, luego $\varnothing^{VR} = \varnothing$.

Veamos ahora la relativización de $x' \equiv x \cup \{x\}$. Como podemos probar que $\bigwedge xu(u \in x' \leftrightarrow u \in x \vee u = x)$, también tenemos la relativización de esta sentencia:

$$\bigwedge xu(u \in F((x')^{VR}) \leftrightarrow u \in F(x) \vee u = x),$$

luego $F((x')^{VR}) = F(x) \cup \{x\}$, luego $(x')^{VR} = F^{-1}(F(x) \cup \{x\})$. Teniendo en cuenta que F fija a los números naturales, es claro que $\bigwedge n \in \omega\ (n')^{VR} = n'$.

Ahora relativizamos $0 \in \omega \wedge \bigwedge n \in \omega\ n' \in \omega$, con lo que tenemos que

$$0 \in F(\omega^{VR}) \wedge \bigwedge n \in F(\omega^{VR})\ (n')^{VR} \in F(\omega^{VR}).$$

En particular, si un número natural $n \in \omega$ cumple $n \in F(\omega^{VR})$, tenemos que $n' = (n')^{VR} \in F(\omega^{VR})$. En definitiva, tenemos que

$$0 \in F(\omega^{VR}) \wedge \bigwedge n \in \omega(n \in F(\omega^{VR}) \to n' \in F(\omega^{VR})).$$

Por el principio de inducción (sin relativizar), concluimos que $\omega \subset F(\omega^{VR})$. Para probar la inclusión opuesta relativizamos el principio de inducción:

$$\bigwedge x(0 \in F(x) \wedge \bigwedge n\,(n \in F(\omega^{VR}) \wedge n \in F(x) \to (n')^{VR} \in F(x))$$

$$\to \bigwedge n(n \in F(\omega^{VR}) \to n \in F(x)).$$

Podemos aplicar esto a $x = \omega = F(\omega)$. Ciertamente sabemos que $0 \in \omega$ y, si $n \in F(\omega^{VR}) \wedge n \in \omega$, también $(n')^{VR} = n' \in \omega$, luego podemos concluir que $F(\omega^{VR}) \subset \omega$, y en total $F(\omega^{VR}) = \omega = F(\omega)$, luego $\omega^{VR} = \omega$.

Ahora consideramos $f = \{(n, x_n) \mid n \in \omega\}$, de modo que $f : \omega \longrightarrow A$ biyectiva, y vamos a probar que se cumple esto mismo relativizado a (V, R).

Es fácil ver que $\{n, x\}^{VR} = F^{-1}(\{n, x\})$ y, si $n \neq x$, como F fija todos los conjuntos con más de un elemento, $\{n, x\}^{VR} = \{n, x\}$. En particular esto vale si $n \in \omega$ y $x \in A$. Similarmente, $(n, x)^{VR} = F^{-1}(\{\{n\}, \{n, x\}\}) = (n, x)$.

Ahora es pura rutina comprobar que f cumple las relativizaciones de todas las condiciones de la definición de aplicación biyectiva. Por ejemplo, tenemos que

$$(\bigwedge z \in f \bigvee nx(n \in \omega \wedge x \in A \wedge z = (n, x)))^{VR},$$

pues, teniendo en cuenta que $F(f) = f$, $F(\omega) = \omega$ y $F(A) = A$, la sentencia es claramente absoluta. Dejamos al lector las comprobaciones restantes, que no ofrecen dificultad alguna.

Por lo tanto A es numerable^{VR} y, en total, hemos probado que

$$(V, R) \vDash \bigvee A\,(A \text{ numerable} \wedge \bigwedge x \in A\ x = \{x\}).$$

■

Veamos ahora que también es consistente la existencia de una sucesión inyectiva

$$\cdots \in x_4 \in x_3 \in x_2 \in x_1 \in x_0.$$

Más precisamente:

Teorema 12.25 *Si* ZFC *es consistente, también lo es* ZFC *menos el axioma de regularidad más la existencia de una sucesión inyectiva* $\{x_n\}_{n\in\omega}$ *tal que*

$$\bigwedge n \in \omega\ x_n = \{x_{n+1}\}.$$

Demostración: La prueba es una ligera variante de la demostración del teorema anterior. Ahora tomamos

$$F(x) = \begin{cases} \{\{\{n+1\}\}\} & \text{si } x = \{\{n\}\} \text{ con } n \in \omega, \\ \{\{n\}\} & \text{si } x = \{\{\{n+1\}\}\}, \text{ con } n \in \omega, \\ x & \text{en otro caso,} \end{cases}$$

consideramos $A = \{\{\{n\}\} \mid n \in \omega\}$ y la biyección $f = \{(n, \{\{n\}\}) \mid n \in \omega\}$. Exactamente igual que en el teorema anterior ve que $(f : \omega \longrightarrow A \text{ biyectiva})^{VR}$, pero ahora $(\bigwedge n \in \omega\ f(n) = \{f(n+1)\})^{VR}$, luego f es en (V, R) una sucesión con la propiedad requerida. ∎

El argumento del teorema anterior admite una variante de interés: si trabajamos en $\mathrm{ZF}_{\mathrm{fin}}$ (lo cual es legítimo, pues el teorema 12.22 se prueba en ZF^*) y consideramos el modelo (V, R) construido con la misma función F, ahora no podemos construir ni el conjunto A ni la función f, pues la clase ω no es un conjunto. No obstante, sigue siendo cierto que $0^{VR} = 0$ y que $\bigwedge n \in \omega\ (n')^{VR} = n$, de donde se sigue por inducción que $\bigwedge n \in \omega\ (n \in \omega)^{VR}$.

Más aún, los mismos argumentos empleados en la prueba del teorema 12.24 muestran que si x es un conjunto cualquiera y $f : n \longrightarrow x$ biyectiva, donde $n \in \omega$ (existe f porque estamos suponiendo que todo conjunto es finito), se prueba que $(\bigvee f n(n \in \omega \wedge f : n \longrightarrow a \text{ biyectiva}))^{VR}$, luego $(a \text{ es finito})^{VR}$. Así pues, en (V, R) también se cumple que todo conjunto es finito.

Veamos ahora que, excepcionalmente, (V, R) cumple el axioma de regularidad en este caso. Tomamos un conjunto $(x \neq \varnothing)^{VR}$, que es lo mismo que $x \neq \varnothing$, con lo que $F(x) \neq \varnothing$. Queremos probar que x tiene $\in$-minimal^{VR}, es decir, que existe un u tal que $u\,R\,x \wedge \neg\bigvee v(v\,R\,u \wedge v\,R\,x)$ o, equivalentemente,

$$u \in F(x) \wedge \neg\bigvee v(v \in F(u) \wedge v \in F(x)).$$

Distinguimos dos casos:

1) Existe un $n \in \omega$ tal que $\{\{n\}\} \in F(x)$. Por reemplazo podemos formar el conjunto $\{n \in \omega \mid \{\{n\}\} \in F(x)\}$, y como todo conjunto es finito, tiene que tener un máximo elemento n. Llamamos $u = \{\{n\}\}$. Entonces $u \in F(x)$, pero si $v \in F(u)$, entonces $v = \{\{n+1\}\} \notin F(x)$, luego u es un $\in$-minimal de x en (V, R).

2) No existe un $n \in \omega$ tal que $\{\{n\}\} \in F(x)$. Sea u un $\in$-minimal de $F(x)$. Esto significa que

$$u \in F(x) \wedge \neg\bigvee v(v \in u \wedge v \in F(x)).$$

Si $F(u) = u$ ya tenemos lo requerido. En caso contrario, $u = \{\{\{n+1\}\}\}$ para cierto $n \in \omega$, y $F(u) = \{\{n\}\}$. Si u no cumple lo requerido es porque $\{n\} \in F(x)$. A su vez, si $\{n\}$ no cumple lo requerido es porque $n \in F(x)$, en cuyo caso podemos tomar el mínimo número natural que cumple $n \in F(x)$, y claramente cumple lo requerido.

Así pues, (V, R) es un modelo de ZFC menos el axioma de infinitud más todo conjunto es finito. Sin embargo, el axioma CT es falso en (V, R), pues $x = \{\{\{0\}\}\}$ no tiene clausura transitiva en (M, R).

En efecto, supongamos que $(y = \operatorname{ct} x)^{VR}$. Como $\{\{0\}\}\, R\, x$, se cumple $\{\{0\}\}\, R\, y$, es decir, $\{\{0\}\} \in F(y)$. Si se cumple $\{\{n\}\} \in F(y)$, puesto que $\{\{n+1\}\}\, R\, \{\{n\}\}\, R\, y$, también $\{\{n+1\}\}\, R\, y$, es decir, $\{\{n+1\}\} \in F(y)$, luego por inducción $\bigwedge n \in \omega\ \{\{n\}\} \in F(y)$, pero el axioma de reemplazo nos da que $\{n \in \omega \mid \{\{n\}\} \in F(y)\}$ es un conjunto, es decir, que ω es un conjunto, mientras que estamos suponiendo que todo conjunto es finito.

En conclusión hemos demostrado lo siguiente:

Teorema 12.26 $\mathrm{ZF}_{\mathrm{fin}}-\mathrm{CT}$ *es consistente, es decir, sin el axioma de infinitud, en* ZFC *no puede demostrarse la existencia de clausura transitiva.*

Más precisamente, hemos construido un modelo de $\mathrm{ZF}_{\mathrm{fin}}$ en el que todo conjunto no vacío tiene $\in$-minimal, pero en el que hay clases no vacías que no tienen $\in$-minimal. Es la existencia de clausura transitiva la que permite probar que las clases no vacías tienen $\in$-minimal partiendo de que lo cumplen los conjuntos.

12.4 Teorías de conjuntos con átomos

Todas las teorías de conjuntos que hemos considerado hasta ahora tienen una propiedad en común cuya "naturalidad" sería cuestionable: sus elementos son conjuntos y nada más que conjuntos (o clases, que también son colecciones de conjuntos), pero lo único que puede ser un elemento de una clase o conjunto es otro conjunto. Si aceptamos el axioma de regularidad, la jerarquía $\{V_\alpha\}_{\alpha\in\Omega}$ pone en evidencia que todos los conjuntos "están hechos" en última instancia del conjunto vacío. Todos son de la forma:

$$\varnothing, \quad \{\varnothing\}, \quad \{\{\varnothing\}\}, \quad \{\varnothing, \{\{\varnothing\}\}\}, \quad \{\varnothing, \{\varnothing\}, \{\{\{\varnothing\}\}\}\}, \quad \ldots$$

aunque $\varnothing$ pueda aparecer infinitas veces, pero no hay nada más que "vacíos y llaves". Sería natural considerar una teoría que hablara de conjuntos y "cosas" que no son conjuntos, de manera que los conjuntos fueran conjuntos de cosas, sin perjuicio de que unos conjuntos pudieran ser también elementos de otros.

La única razón por la que esta posibilidad no es popular entre los matemáticos es que no se gana nada con ella, pero nada impide construir teorías de conjuntos que admitan la existencia de objetos que no son conjuntos. Una forma de hacerlo sin romper con los esquemas que hemos trazado hasta ahora es convenir en que un átomo, es decir, un conjunto de la forma $x = \{x\}$ es una forma razonable de representar un objeto que en realidad no es un conjunto. Así, aunque desde cierto punto de vista un átomo puede considerarse como una aberración, también podemos verlo como un simple (y elegante) convenio, no muy diferente del convenio por el que el número 3 (que en principio no tiene nada de conjuntista) se identifica con el conjunto $3 = \{0, 1, 2\}$. Dado que en las teorías de conjuntos que manejamos todo tiene que ser un conjunto, pues ello está implícito en su mismo planteamiento, podemos convenir que los objetos que no son conjuntos se representen en la teoría como átomos, es decir, como conjuntos en los que la pertenencia no dice nada realmente, sino que se reduce a un mero formulismo.

Ya hemos visto que es consistente con ZFC$-V = R$ la existencia de átomos, y ahora vamos a ver que ésta es compatible con una versión ligeramente debilitada del axioma de regularidad. Trabajamos en ZF$-V = R$:

Definición 12.27 Definimos la clase $R(A)$ de los *conjuntos regulares respecto a un conjunto* A como la dada por

$$R_0(A) = \operatorname{ct} A \quad \wedge \quad \bigwedge\alpha\ R_{\alpha+1}(A) = \mathcal{P}R_\alpha(A) \quad \wedge \quad \bigwedge\lambda\ R_\lambda(A) = \bigcup_{\delta<\lambda} R_\delta(A),$$

$$R(A) = \bigcup_{\alpha\in\Omega} R_\alpha(A).$$

Observemos que si $A = \varnothing$ entonces $R(A)$ se reduce a la clase R de los conjuntos regulares (alternativamente, podemos definir $R = R(\varnothing)$ o, directamente, definir R mediante la definición anterior omitiendo A). Una simple inducción prueba que

$$\bigwedge AB(A \subset B \to \bigwedge\alpha\ R_\alpha(A) \subset R_\alpha(B)),$$

luego

$$\bigwedge AB(A \subset B \to R(A) \subset R(B)).$$

En particular $R \subset R(A)$.

También se comprueba inmediatamente por inducción que cada $R_\alpha(A)$ es un conjunto transitivo. En efecto, es claro para $\alpha = 0$ y para ordinales límite, mientras que si $R_\alpha(A)$ es transitivo y $u \in v \in R_{\alpha+1}(A)$, entonces tenemos que $u \in v \subset R_\alpha(A)$, luego $u \in R_\alpha(A)$, luego $u \subset R_\alpha(A)$, luego $u \in R_{\alpha+1}(A)$.

Consecuentemente $R(A)$ es una clase transitiva.

Para cada $u \in R(A)$ existe un ordinal α tal que $u \in R_\alpha(A)$, luego $u \subset R_\alpha(A)$, luego podemos definir el *rango* de u como el menor ordinal α tal que $u \subset R_\alpha(A)$. Tenemos así una aplicación $\operatorname{rang}_A : R(A) \longrightarrow \Omega$.

Observemos que, para $\alpha > 0$, se cumple[6]

$$R_\alpha(A) = \{u \in R(A) \mid \mathrm{rang}_A(u) < \alpha\}.$$

En efecto, si $u \in R_\alpha(A)$ y $\alpha = \delta + 1$, entonces $u \subset R_\delta(A)$, mientras que si α es un ordinal límite, existe $\delta < \alpha$ tal que $u \in R_\delta(A)$, luego $u \subset R_\delta(A)$. En ambos casos $\mathrm{rang}_A(u) \leq \delta < \alpha$. Recíprocamente, si $\mathrm{rang}_A(u) = \delta < \alpha$, entonces $u \subset R_\delta(A)$, luego $u \in R_{\delta+1}(A) \subset R_\alpha(A)$.

El rango satisface la siguiente relación recurrente: si $\mathrm{rang}_A(u) > 0$, entonces

$$\mathrm{rang}_A(u) = \bigcup_{v \in u} (\mathrm{rang}_A(v) + 1).$$

Es decir, el rango de un conjunto (si es no nulo) es el menor ordinal estrictamente mayor que los rangos de todos sus elementos.[7]

En efecto, si $\mathrm{rang}_A(u) = \alpha > 0$ y $v \in u \subset R_\alpha(A)$, entonces $v \in R_\alpha(A)$, luego $\mathrm{rang}_A(v) < \alpha$, luego $\mathrm{rang}_A(v) + 1 \leq \alpha$, y tenemos la desigualdad

$$\beta = \bigcup_{v \in u} (\mathrm{rang}_A(v) + 1) \leq \alpha.$$

Por otra parte, cada $v \in u$ cumple $\mathrm{rang}_A(v) < \beta$, luego $v \in R_\beta(A)$, luego $u \subset R_\beta(A)$, luego $\alpha = \mathrm{rang}_A(u) \leq \beta$.

Observemos ahora que $\mathcal{P}R(A) = R(A)$. La inclusión $R(A) \subset \mathcal{P}R(A)$ se debe simplemente a la transitividad de $R(A)$. Para probar la contraria observamos que si $u \in \mathcal{P}R(A)$, el axioma de reemplazo nos da que la clase

$$\{\mathrm{rang}_A(v) \mid v \in u\}$$

es un conjunto de ordinales, luego tiene supremo α, luego $u \subset R_\alpha(A)$, luego $u \in R_{\alpha+1}(A) \subset R(A)$.

De aquí obtenemos un principio de $\in$-inducción:

$$\bigwedge X(\mathrm{ct}(A) \subset X \wedge \mathcal{P}X \subset X \to R(A) \subset X).$$

Más detalladamente: si queremos probar que todo elemento de $R(A)$ está en una clase X (es decir, tiene la propiedad que la define), basta probar que todos los elementos de A están en X y que, bajo la hipótesis de que los elementos de un conjunto arbitrario $u \in R(A)$ están en X, probar que $u \in X$.

Esto es un caso particular del teorema general de inducción transfinita, aplicado a la relación

$$E_A = \{(x, y) \in R(A) \times R(A) \mid x \in y \wedge y \notin \mathrm{ct}(A)\}.$$

[6] Obviamente, la igualdad vale para $\alpha = 0$ si y sólo si $A = \varnothing$.

[7] Si $A = \varnothing$ el único conjunto de rango 0 es $\varnothing$ y la fórmula de los rangos se cumple trivialmente.

Observemos que esta relación está bien fundada en $R(A)$ pues, dado cualquier conjunto $x \subset R(A)$ no vacío, cualquier $u \in x$ de rango mínimo es un E_A-minimal de x, ya que $v\, E_A\, u \to \mathrm{rang}_A(v) < \mathrm{rang}_A(u)$.

Por lo tanto, el teorema 11.38 nos da un teorema de recursión para $R(A)$: para definir una función $F : R(A) \longrightarrow V$ basta definirla sobre los elementos de $R_0(A)$ y, para cualquier otro conjunto x, definir $F(x)$ en función de x y $F|_x$.

En general, la aplicación rang_A puede ser más o menos "caótica", porque si A (o su clausura transitiva) contiene conjuntos regulares, éstos aparecen artificialmente "antes de tiempo" en la sucesión $\{R_\alpha(A)\}_{\alpha\in\Omega}$, y ello distorsiona la jerarquía. Esto no sucede si A contiene sólo conjuntos que no aparecerían de no haber sido incluidos en la base de la jerarquía, por ejemplo átomos.

A partir de aquí suponemos que A es un conjunto de átomos. Entonces A es transitivo, luego $R_0(A) = A$ y los únicos átomos de $R(A)$ son los de A. Más aún, si $u \in R(A)$ cumple $u \in u$, es porque $u = \{u\} \in A$. En efecto, tiene que ser $\mathrm{rang}_A(u) = 0$, pues en caso contrario $u \in u$ implicaría $\mathrm{rang}_A(u) < \mathrm{rang}_A(u)$.

Observemos que $R_\alpha(A) \cap R = R_\alpha$. En efecto, para $\alpha = 0$ es $A \cap R = \varnothing$, lo cual se cumple, por ejemplo, porque acabamos de probar que $R(\varnothing)$ no tiene átomos. Si vale para α y $u \in R_{\alpha+1}(A) \cap R$, entonces $u \subset R_\alpha(A) \cap R = R_\alpha$, luego $u \in R_{\alpha+1}$. La inclusión opuesta vale en cualquier caso, y el caso límite es inmediato.

Similarmente se prueba que $R_\alpha(A) \cap \Omega = \alpha$, luego $\Omega \subset R(A)$. En particular concluimos que, para $u \in R$, se cumple $\mathrm{rang}_A(u) = \mathrm{rang}(u)$.

Definición 12.28 Llamaremos ZFCA (resp. NBGA) a la teoría de conjuntos ZFC (resp. NBG) sin el axioma de regularidad más el axioma[8]

$$\mathrm{cto\,A} \wedge \mathrm{V} = \mathrm{R(A)},$$

donde $A \equiv \{x \mid x = \{x\}\}$.

Notemos que este axioma no contradice a V = R, sino que equivale a éste si añadimos que A = ∅. Por otra parte, la teoría de conjuntos con átomos puede completarse con varios axiomas, como por ejemplo $\bigvee f\ f : \omega \longrightarrow \mathrm{A}$ biyectiva, que nos da un conjunto numerable de átomos.

Según la costumbre, al trabajar con este axioma de regularidad relajado, escribiremos $V_\alpha(A)$ en lugar de $R_\alpha(A)$.

En 12.24 hemos probado que si ZFC es consistente, también lo es la teoría ZFC$-V = R$ más la existencia de un conjunto numerable de átomos. Ahora podemos refinar este resultado:

[8] Para que esta sentencia tenga sentido como axioma de ZFCA hemos de observar que puede reformularse omitiendo toda referencia a clases propias. Concretamente:

$$\bigvee x(\bigwedge y(y \in x \leftrightarrow y = \{y\}) \wedge \bigwedge u \bigvee \alpha\ u \in \mathrm{R}_\alpha(x)).$$

Teorema 12.29 *Si* ZFC *es consistente también lo es* ZFCA+A *es (infinito) numerable.*

DEMOSTRACIÓN: Basta probar en ZFC+ "existe un conjunto numerable de átomos" que si A_0 es dicho conjunto, la clase $R(A_0)$ es un modelo de ZFCA+A es numerable.

Trabajamos, pues, en ZFC más la existencia de un conjunto numerable A_0 formado por átomos. No afirmamos que A_0 coincida con la clase A de todos los átomos, pero sabemos que los únicos átomos en $R(A_0)$ son los de A_0.

La prueba del teorema 12.20 se adapta trivialmente para probar que $R(A_0)$ es un modelo transitivo de ZFC$-V = R$. En efecto, la prueba se basa esencialmente en que $\mathcal{P}R = R$, y ahora tenemos igualmente que $\mathcal{P}R(A_0) = R(A_0)$. La prueba de que R cumple $V = R$ no se generaliza porque ahora no podemos afirmar que $R(A_0) \subset R$.

Sólo nos falta probar que $R(A_0)$ es un modelo de la teoría del enunciado. Concretamente, debemos probar que

$$(\bigvee A(\bigwedge x(x \in A \leftrightarrow x = \{x\}) \wedge A \text{ es numerable}$$

$$\wedge \bigwedge x \bigvee \alpha(\alpha \text{ es un ordinal} \wedge x \in R_\alpha(A)))^{R(A_0)}.$$

Como $A_0 \in R(A_0)$, basta probar

$$(\bigwedge x(x \in A_0 \leftrightarrow x = \{x\}) \wedge A_0 \text{ es numerable}$$

$$\wedge \bigwedge x \bigvee \alpha(\alpha \text{ es un ordinal} \wedge x \in R_\alpha(A_0)))^{R(A_0)}.$$

La relativización de la primera parte es $\bigwedge x \in R(A_0)(x \in A_0 \leftrightarrow x = \{x\})$, lo cual ya hemos observado que es cierto.

La fórmula "α es un ordinal" es absoluta para modelos transitivos si suponemos $V = R$ porque entonces se reduce a "ser transitivo y $\in$-conexo", que son propiedades Δ_0, pero en nuestro contexto falta probar que "ser un conjunto bien fundado" también es absoluto. Concretamente,

$$\bigwedge x(x \text{ está bien fundado} \leftrightarrow \bigwedge y(y \subset x \wedge y \neq \varnothing \to \bigvee u \in y(u \cap y = \varnothing))).$$

Al relativizar a $R(A_0)$ queda

$$\bigwedge x \in R(A_0)(x \text{ está bien fundado}^{R(A_0)} \leftrightarrow$$

$$\bigwedge y(y \subset x \wedge y \neq \varnothing \to \bigvee u \in y(u \cap y = \varnothing))).$$

Notemos que no hace falta poner $y \in R(A_0)$ porque si $y \subset x$, entonces $y \in \mathcal{P}R(A_0) = R(A_0)$. En definitiva tenemos que

$$\bigwedge x \in R(A_0)(x \text{ está bien fundado}^{R(A_0)} \leftrightarrow x \text{ está bien fundado}).$$

Así pues, ser un ordinal es absoluto para $R(A_0)$, y claramente entonces también es absoluto ser un ordinal sucesor, un ordinal límite, un ordinal límite, y también ω.

Con esto podemos probar la segunda parte: tomamos $f : \omega \longrightarrow A_0$ biyectiva. Como $\omega, A_0 \in R(A_0)$, también $\omega \times A_0 \in R(A_0)$, luego $f \subset R(A_0)$ y por lo tanto $f \in \mathcal{P}R(A_0) = R(A_0)$. Como la fórmula "$f : x \longrightarrow y$ biyectiva" es absoluta (porque es Δ_0) y ω también lo es,

$$(f : \omega \longrightarrow A_0 \text{ biyectiva})^{R(A_0)},$$

pero esto es lo mismo que $(A_0 \text{ es numerable})^{R(A_0)}$.

Por otra parte, si $x \in R(A_0)$, entonces $\mathcal{P}x \subset \mathcal{P}R(A_0) = R(A_0)$, luego

$$(\mathcal{P}x)^{R(A_0)} = \mathcal{P}x \cap R(A_0) = \mathcal{P}x,$$

es decir, $\mathcal{P}x$ también es absoluto para $R(A_0)$. Ahora demostramos que

$$\bigwedge \alpha \in \Omega \; R_\alpha(A)^{R(A_0)} = R_\alpha(A_0).$$

En efecto, para $\alpha = 0$, relativizando $\bigwedge x(x \in R_0(A) \leftrightarrow x = \{x\})$ obtenemos que $\bigwedge x \in R(A_0)(x \in R_0(A)^{R(A_0)} \leftrightarrow x = \{x\})$, y ya hemos visto que esto equivale a $\bigwedge x(x \in R_0(A)^{R(A_0)} \leftrightarrow x \in A_0)$ (donde hemos usado la transitividad de $R(A_0)$ para eliminar la cota del generalizador), y por consiguiente llegamos a que $R_0(A)^{R(A_0)} = R_0(A_0)$.

Si se cumple para α, relativizando $\bigwedge \alpha \; R_{\alpha+1}(A) = \mathcal{P}R_\alpha(A)$, teniendo en cuenta que $\mathcal{P}x$ es absoluto, concluimos que

$$R_{\alpha+1}(A)^{R(A_0)} = \mathcal{P}R_\alpha(A)^{R(A_0)} = \mathcal{P}R_\alpha(A_0) = R_{\alpha+1}(A_0).$$

Finalmente, si se cumple para todo $\delta < \lambda$, relativizando que

$$\bigwedge \lambda \bigwedge x(x \in R_\lambda(A) \leftrightarrow \bigvee \delta < \lambda \; x \in R_\delta(A))$$

concluimos que

$$\bigwedge x(x \in R_\lambda(A)^{R(A_0)} \leftrightarrow \bigvee \delta < \lambda \; x \in R_\delta(A)^{R(A_0)}).$$

Aplicando la hipótesis de inducción llegamos a que

$$R_\lambda(A)^{R(A_0)} = \bigcup_{\delta<\lambda} R_\delta(A_0) = R_\lambda(A_0).$$

De este modo, como trivialmente $\bigwedge x \in R(A_0) \bigvee \alpha \; x \in R_\alpha(A_0)$, concluimos que $(\bigwedge x \bigvee \alpha \; x \in R_\alpha(A))^{R(A_0)}$, pero esto es $(\mathrm{V} = R(A))^{R(A_0)}$. Así pues, $R(A_0)$ es un modelo de ZFCA más "A es numerable". ∎

12.5 El teorema de reflexión

El teorema que presentamos en esta sección implica, entre otras cosas, que ZF (supuesto que sea consistente) no es finitamente axiomatizable. En su forma más general es demostrable en ZF*+AI.

Diremos que una sucesión $\theta_1, \ldots, \theta_k$ de expresiones de $\mathcal{L}_{\mathrm{tc}}$ es *adecuada* si toda subexpresión de cada θ_i es de la forma θ_j, con $j < i$ y si $\theta_i \equiv x|\alpha$, entonces existe un $j < i$ tal que $\theta_j \equiv \bigvee\!\!\!\!\!^{1} x\alpha$.

Es claro que toda sucesión finita de fórmulas de $\mathcal{L}_{\mathrm{tc}}$ se puede extender hasta una sucesión adecuada de expresiones de $\mathcal{L}_{\mathrm{tc}}$.

Teorema 12.30 *Sea $\theta_1, \ldots, \theta_k$ una sucesión adecuada de expresiones de $\mathcal{L}_{\mathrm{tc}}$. Si $M \subset N$ son dos clases cualesquiera tales que $\varnothing \in M$ y para toda fórmula θ_i que sea de la forma $\bigwedge x\alpha(x, x_1, \ldots, x_n)$ se cumple*

$$\bigwedge x_1 \cdots x_n \in M(\bigvee x \in N \neg\alpha^N(x, x_1, \ldots, x_n) \to \bigvee x \in M \neg\alpha^N(x, x_1, \ldots, x_n)),$$

entonces todas las expresiones de la sucesión son absolutas para $M - N$.

Demostración: Vamos a demostrar por inducción sobre i que cada θ_i es absoluta para $M - N$. De hecho, todas las posibilidades para θ_i se tratan de forma evidente salvo los casos $\theta_i \equiv \bigwedge x\alpha$ y $\theta_i \equiv x|\alpha$. Veamos, pues, estos dos.

Si $\theta_i \equiv \bigwedge x\alpha$, por hipótesis de inducción tenemos que

$$\bigwedge x x_1 \cdots x_n \in M(\alpha^M \leftrightarrow \alpha^N).$$

Obviamente entonces, $\bigwedge x_1 \cdots x_n \in M((\bigwedge x\alpha)^N \to (\bigwedge x\alpha)^M)$. Por otra parte, al combinar la hipótesis del teorema con la hipótesis de inducción tenemos la implicación contraria.

Supongamos ahora que $\theta_i \equiv x|\alpha$. El hecho de que la sucesión dada sea adecuada nos da entonces la hipótesis de inducción para $\bigvee\!\!\!\!\!^{1} x\alpha$, es decir,

$$\bigwedge x_1 \cdots x_n \in M(\bigvee\!\!\!\!\!^{1} x \in M\alpha^M \leftrightarrow \bigvee\!\!\!\!\!^{1} x \in N\alpha^N).$$

Fijados $x_1, \ldots, x_n \in M$, o bien no se da ninguna de las dos equivalencias, en cuyo caso $\theta_i^M = \varnothing = \theta_i^N$, o bien se dan ambas. Sea $x \in M$ el único que cumple $\alpha^M(x, x_1, \ldots, x_n)$. Por la hipótesis de inducción sobre α también $\alpha^N(x, x_1, \ldots, x_n)$, es decir, el único elemento de M que cumple α^M es el único elemento de N que cumple α^N, luego $(x|\alpha)^M = (x|\alpha)^N$. ∎

El teorema de reflexión es un caso particular del teorema siguiente:

Teorema 12.31 *Sean $\phi_1, \ldots, \phi_r$ fórmulas de $\mathcal{L}_{\mathrm{tc}}$. Sea $\{Z_\alpha\}_{\alpha\in\Omega}$ una sucesión de conjuntos tal que $\bigwedge\alpha\beta(\alpha \leq \beta \to Z_\alpha \subset Z_\beta)$ y $\bigwedge\lambda\ Z_\lambda = \bigcup_{\delta<\lambda} Z_\delta$. Sea $Z = \bigcup_{\alpha\in\Omega} Z_\alpha$. Entonces*[9] *para cada ordinal α existe un ordinal límite $\lambda > \alpha$ tal que $\phi_1, \ldots, \phi_r$ son absolutas para $Z_\lambda - Z$.*

[9] Con el convenio usual de que la descripción impropia es $\varnothing$ tenemos que suponer además que $\varnothing \in Z$.

DEMOSTRACIÓN: La idea de la prueba es la misma que la del teorema de Löwenheim-Skolem, sólo que no necesitamos el axioma de elección porque vamos a elegir ordinales. Extendemos la sucesión de fórmulas dada a una sucesión adecuada de expresiones $\theta_1, \ldots, \theta_k$.

Para cada índice i tal que $\theta_i \equiv \bigwedge x\psi(x, x_1, \ldots, x_n)$, con $n \geq 1$, definimos la función $G_i : Z^n \longrightarrow \Omega$ tal que $G_i(x_1, \ldots, x_n)$ es el mínimo ordinal η tal que $\bigvee x \in Z_\eta \neg\psi^Z(x, x_1, \ldots, x_n)$ si existe tal η y $G_i(x_1, \ldots, x_n) = 0$ en caso contrario.

Definimos $F_i : \Omega \longrightarrow \Omega$ mediante

$$F_i(\xi) = \bigcup_{y \in Z_\xi^n} G_i(y).$$

Si $\theta_i \equiv \bigwedge x\psi(x)$, sin variables libres, definimos $F_i : \Omega \longrightarrow \Omega$ de modo que $F_i(\xi)$ es el mínimo ordinal η tal que $\bigvee x \in Z_\eta \neg\psi^Z(x)$ o bien $F_i(\xi) = 0$ si no existe tal η.

Para los índices i tales que θ_i no es una generalización definimos F_i como la función nula. Dado $\alpha \in \Omega$, llamamos α_0 al menor ordinal $> \alpha$ tal que $\varnothing \in Z_{\alpha_0}$. Definimos una sucesión $\{\beta_p\}_{p\in\omega}$ mediante

$$\beta_0 = \alpha_0, \qquad \beta_{p+1} = (\beta_p + 1) \cup F_1(\beta_p) \cup \cdots \cup F_k(\beta_p).$$

Como la sucesión es estrictamente creciente, $\lambda = \bigcup_{p\in\omega} \beta_p$ es un ordinal límite, obviamente $\lambda > \alpha$. Vamos a probar que cumple lo pedido. De la construcción se sigue que si $\delta \leq \epsilon$ entonces $F_i(\delta) \leq F_i(\epsilon)$. A su vez esto implica que si $\delta < \lambda$ entonces $F_i(\delta) < \lambda$. En efecto, existe un $p \in \omega$ tal que $\delta < \beta_p$, con lo que $F_i(\delta) \leq F_i(\beta_p) \leq \beta_{p+1} < \lambda$.

Para probar que las expresiones θ_i son absolutas para $Z_\lambda - Z$ aplicamos el teorema anterior. Suponemos que $\theta_i \equiv \bigwedge x\psi(x, x_1, \ldots, x_n)$ (tal vez $n = 0$) y hemos de probar que

$$\bigwedge x_1 \cdots x_n \in Z_\lambda(\bigvee x \in Z \neg\psi^Z(x, x_1, \ldots, x_n) \to \bigvee x \in Z_\lambda \neg\psi^Z(x, x_1, \ldots, x_n)).$$

Fijamos $x_1, \ldots, x_n \in Z_\lambda$ (en el caso en que $n \neq 0$). Entonces existe un ordinal $\delta < \lambda$ tal que $x_1, \ldots, x_n \in Z_\delta$. Sea $\eta < \lambda$ el mínimo ordinal tal que $\bigvee x \in Z_\eta \neg\psi^Z(x, x_1, \ldots, x_n)$.

Si $n > 0$ entonces $\eta = G_i(x_1, \ldots, x_n) \leq F_i(\delta) < \lambda$ y si $n = 0$ entonces $\eta = F_i(0) < \lambda$. En cualquier caso tenemos que $Z_\eta \subset Z_\lambda$, luego concluimos que $\bigvee x \in Z_\lambda \neg\psi^Z(x, x_1, \ldots, x_n)$, como teníamos que probar. ∎

Si suponemos el axioma de regularidad y el axioma de partes, es decir, en ZF, el teorema anterior se aplica a la sucesión

$$\mathrm{V} = \bigcup_{\alpha\in\Omega} \mathrm{V}_\alpha.$$

Así tenemos el teorema de reflexión propiamente dicho:

Teorema 12.32 (Teorema de reflexión) *Si $\phi_1, \ldots, \phi_r$ son fórmulas de $\mathcal{L}_{\mathrm{tc}}$, entonces para todo ordinal α existe un ordinal límite $\lambda > \alpha$ tal que las fórmulas dadas son absolutas para V_λ.*

En particular, puesto que V es un modelo de ZF, si Γ es un conjunto finito de axiomas de ZF, en ZF se demuestra que $V \vDash \Gamma$, luego por el teorema anterior existe un conjunto $M = V_\lambda$ tal que $M \vDash \Gamma$, y por el teorema 12.13 esto equivale a $M \vDash \ulcorner\Gamma\urcorner$. Así pues:

Teorema 12.33 *Si Γ es un conjunto finito de axiomas de* ZF*, en* ZF *se demuestra que existe un conjunto M que es un modelo transitivo de $\ulcorner\Gamma\urcorner$. En particular $\vdash_{\mathrm{ZF}} \mathrm{Consis}\,\ulcorner\Gamma\urcorner$.*

Como consecuencia inmediata:

Teorema 12.34 *Ninguna extensión consistente de* ZF *es finitamente axiomatizable.*

Demostración: Supongamos que Γ es un conjunto finito de sentencias de $\mathcal{L}_{\mathrm{tc}}$ entre cuyas consecuencias estén todos los axiomas (y, por consiguiente, todos los teoremas) de ZF. Vamos a probar que Γ es contradictorio. Por el teorema de reflexión (que es demostrable a partir de Γ) existe un modelo $M = V_\lambda$ tal que las sentencias de Γ son absolutas para $M - V$, luego $M \vDash \Gamma$, luego también $M \vDash \ulcorner\Gamma\urcorner$. Así pues, $\Gamma \vdash \bigvee M\ M \vDash \ulcorner\Gamma\urcorner$, luego $\Gamma \vdash \mathrm{Consis}\,\ulcorner\Gamma\urcorner$, y el segundo teorema de incompletitud de Gödel implica entonces que Γ es contradictorio. ∎

12.6 Consideraciones finales

Tal vez la principal conclusión que el lector debería extraer de este capítulo es que diseñar una teoría de conjuntos es precisamente eso, una cuestión de diseño, y no una cuestión de "capturar" en algún sentido, la "esencia" de lo que son los conjuntos. Podemos diseñar teorías de conjuntos con o sin conjuntos infinitos, con y sin conjuntos no numerables, con y sin átomos, con y sin funciones de elección, etc. Entre otras variantes que no hemos discutido aquí podemos destacar las teorías de conjuntos no estándar, que incorporan números naturales infinitamente grandes y números reales infinitamente pequeños, es decir, conceptos que formalizan el concepto de "magnitud infinitesimal" que los matemáticos y los físicos han manejado informalmente durante varios siglos.

De entre toda una amplia gama de posibilidades ZFC y NBG tienen el nivel de complejidad adecuado para que se consideren hoy en día las teorías de conjuntos "estándar", en el sentido de que una afirmación matemática se considera "demostrada" si puede probarse en cualquiera de estas dos teorías equivalentes. En esta elección no sólo se tiene en cuenta la capacidad expresiva de las teorías (ciertamente, en ZFC puede formalizarse cualquier argumento que cualquier matemático juzgue "convincente") sino también la simplicidad a la hora de trabajar en ellas. Por ejemplo, en KPI+AP+AE puede formalizarse la mayor parte

de la matemática actual, y es una teoría bastante más débil que ZFC, pero al matemático no familiarizado con la lógica le resulta artificial tener que pararse a analizar la complejidad sintáctica de sus afirmaciones (es decir, situarlas en la jerarquía de Lévy) y, desde luego, le resulta totalmente injustificable que se le prohíba afirmar la existencia de un conjunto sólo porque la fórmula con que pretende definirlo es demasiado compleja. Lo mismo sucede con las teorías de conjuntos no estándar, en las que no caer en contradicciones requiere un control continuo sobre la estructura sintáctica de las afirmaciones que se manejan, algo ajeno a la práctica matemática habitual y que, por consiguiente, reduce drásticamente el nivel de popularidad que pueden alcanzar estas teorías alternativas. Por el contrario, cualquier argumento matemático usual se formaliza en ZFC o NBG "de forma natural", en el sentido de que una gran parte de los matemáticos profesionales escriben teoremas de ZFC sin conocer siquiera los detalles técnicos de ZFC, y dicho desconocimiento no provoca nunca errores análogos a los que podría cometer alguien que creyera haber demostrado algo en KPI y su prueba fuera errónea por haber utilizado, por ejemplo, especificación para una fórmula Σ_1 no Δ_1. Obviamente, un matemático que trabaje en ZFC también puede cometer errores, pero, salvo que el asunto sea una cuestión muy íntimamente relacionada con la teoría de conjuntos, dichos errores los puede detectar cualquier matemático competente aunque tampoco conozca los detalles técnicos de ZFC.

El lector no debería quedarse con la idea de que ZFC o NBG son las teorías "más completas" de entre todas las posibles. Son las más completas de entre las que considera "indiscutibles" la mayor parte de la comunidad matemática, pero ZFC puede extenderse con una gran variedad de axiomas adicionales, que a su vez tienen consecuencias no triviales (y a menudo mutuamente contradictorias) en varias ramas de la matemática, no sólo la teoría de conjuntos propiamente dicha, sino también la topología, el análisis o el álgebra.

Aunque puede especularse sobre qué teorías de conjuntos describen más fielmente nuestra noción intuitiva de conjunto, en realidad esto no es especialmente importante, como puede entenderse pensando en la geometría: no cabe duda de que la única geometría que se ajusta exactamente a nuestros conceptos intuitivos de punto, recta y plano (y todos los relacionados, como perpendicularidad, paralelismo, etc.) es la geometría tridimensional euclídea (y la bidimensional, si nos restringimos a la geometría plana), pero esto no es razón para considerar que las geometrías no euclídeas (o las geometrías euclídeas de dimensiones superiores) tienen menos valor matemático o son, en cualquier sentido, "geometrías de segunda clase". Del mismo modo, aunque puede cuestionarse, por ejemplo, que el concepto de "conjunto no numerable" tenga algún contenido intuitivo, eso no es razón para desconfiar del axioma de partes de ZFC. Al fin y al cabo, un conjunto no numerable no es nada más extraño que un plano en el que no hay rectas paralelas.

De los cuatro axiomas que hemos discutido separadamente en las teorías ZFC y NBG: los axiomas de infinitud, partes, regularidad y elección, los dos primeros son de naturaleza muy distinta a los dos últimos. Los dos primeros "añaden conjuntos", mientras que los dos últimos "eliminan conjuntos". Con esto

queremos decir que, por ejemplo, si partimos de un modelo M de ZFC menos $V = R$, podemos obtener un modelo de ZFC sin más que tomar M_0 como el conjunto de los objetos de M que satisfacen la fórmula "x es regular", con la misma relación de pertenencia. Este argumento puede formalizarse en ZF*+AI de forma totalmente natural, con lo que

$$\underset{\mathrm{ZF}^*+\mathrm{AI}}{\vdash} \mathrm{Consis}\,\ulcorner\mathrm{ZFC} - \mathrm{V} = \mathrm{R}\urcorner \leftrightarrow \mathrm{Consis}\,\ulcorner\mathrm{ZFC}\urcorner.$$

Lo mismo sucede con el axioma de elección: Puede probarse que a partir de cualquier modelo M de ZF puede obtenerse un modelo M_0 de ZFC sin más que quedarse con los objetos de M que cumplen una propiedad "x es constructible" definida adecuadamente, y la prueba se formaliza sin dificultad en ZF*+AI, por lo que

$$\underset{\mathrm{ZF}^*+\mathrm{AI}}{\vdash} \mathrm{Consis}\,\ulcorner\mathrm{ZF}\urcorner \leftrightarrow \mathrm{Consis}\,\ulcorner\mathrm{ZFC}\urcorner.$$

En cambio, si partimos de un modelo M de ZFC$-$AP, en principio puede suceder que en él no haya conjuntos no numerables, luego para construir un modelo de ZFC a partir de él habría que añadir de un modo y otros "nuevos conjuntos" no numerables (sin perjuicio de que sólo fueran no numerables internamente, pero el modelo completo fuera numerable, tal y como explicamos al tratar la paradoja de Skolem). Y no está claro cómo construir un modelo de ZFC a partir de un modelo de ZFC$-$AP. Más aún, no es que "no esté claro", sino que podemos probar que no es posible construir un modelo de ZFC a partir de un modelo de ZFC$-$AP por un procedimiento formalizable en ZFC, pues en tal caso tendríamos que

$$\underset{\mathrm{ZFC}}{\vdash} \mathrm{Consis}\,\ulcorner\mathrm{ZFC} - \mathrm{AP}\urcorner \leftrightarrow \mathrm{Consis}\,\ulcorner\mathrm{ZFC}\urcorner,$$

pero hemos visto que $\underset{\mathrm{ZFC}}{\vdash} \mathrm{Consis}\,\ulcorner\mathrm{ZFC} - \mathrm{AP}\urcorner$, por lo que llegaríamos a que

$$\underset{\mathrm{ZFC}}{\vdash} \mathrm{Consis}\,\ulcorner\mathrm{ZFC}\urcorner,$$

y el segundo teorema de incompletitud de Gödel nos daría que ZFC es contradictorio (y también ZFC$-$AP). Así pues, dado que cualquier argumento conjuntista "convincente" es formalizable en ZFC, podemos concluir que, si ZFC$-$AP es consistente, es imposible dar una prueba metamatemática "convincente" de que la consistencia de ZFC$-$AP implica la consistencia de ZFC, justo al revés de lo que sucede con $V = R$ o AE.

Lo mismo vale para el axioma de infinitud. Se trata de un axioma que "añade conjuntos" porque en un modelo de ZFC$-$AI no tiene por qué haber conjuntos infinitos, y en uno de ZFC tiene que haberlos. Y el hecho de que $\underset{\mathrm{ZFC}}{\vdash} \mathrm{Consis}\,\ulcorner\mathrm{ZFC} - \mathrm{AI}\urcorner$ hace que si ZFC es consistente, es imposible probar que lo es aun suponiendo la consistencia de ZFC$-$AI, por el mismo argumento empleado con AP.

Así pues, AI y AP son los dos axiomas "fuertes" de ZFC (en realidad el axioma del reemplazo pertenece a esta misma categoría con respecto a la teoría

de Zermelo), mientras que los axiomas de regularidad y elección son axiomas "débiles" en el sentido de que añadirlos aumenta el conjunto de teoremas, pero no el conjunto de teoremas aritméticos (y en particular no puede dar lugar a una demostración de la sentencia $0 \neq 0$ salvo que ya pudiera ser demostrada sin el nuevo axioma).

Apéndice A

Conceptos elementales de la teoría de conjuntos

En este apéndice recogemos por completitud los conceptos que conforman el vocabulario básico en torno a los conjuntos.

A.1 Definiciones básicas

En esta sección representaremos por letras minúsculas *objetos* arbitrarios $x, y, z \ldots$ y por letras mayúsculas *colecciones* de objetos $A, B, C, \ldots$, y escribiremos $x \in A$ cuando el objeto x sea uno de los integrantes de la colección A. En caso contrario escribiremos $x \notin A$. Usaremos la notación $\{x \mid \phi(x)\}$ para referirnos a la colección de todos los objetos x que cumplen la propiedad $\phi(x)$. La notación $\{x_1, \ldots, x_n\}$ representará a la colección formada exactamente por los objetos $x_1, \ldots, x_n$.

Esto (al igual que todo cuanto vamos a exponer en esta sección) es tan general que admite interpretaciones muy diversas. Podemos entender que los objetos a los que nos referimos son objetos metamatemáticos bien definidos informalmente, como los números naturales, los signos de un lenguaje formal, las sucesiones finitas de signos, etc., y que las colecciones de objetos son criterios bien definidos informalmente que especifican algunos de estos objetos (como la colección de los números pares, o la de las fórmulas de un lenguaje formal, etc.) y que las propiedades $\phi(x)$ son propiedades bien definidas informalmente, (como "ser un número par" o "ser una sucesión finita de números naturales", etc.), pero también podemos entender todo cuanto vamos a decir como definiciones y teoremas de diversas teorías formales. Puesto que sólo vamos a dar definiciones elementales y presentar consecuencias inmediatas, resulta evidente que todos los resultados que presentamos aquí son formalizables en cualquier teoría axiomática que reúna los requisitos mínimos para ello, que iremos explicitando según los vayamos necesitando.

De momento sólo necesitamos que la teoría en cuestión proporcione un sentido preciso a las expresiones $x \in A$ y $\{x \mid \phi(x)\}$ (y no a todas las expresiones de este tipo, sino sólo para los casos concretos de propiedades $\phi(x)$ que vamos a considerar) y que nos permita afirmar que si dos colecciones de objetos tienen los mismos elementos, entonces son la misma colección de objetos. Formalmente:

$$\bigwedge XY(\bigwedge u(u \in X \leftrightarrow u \in Y) \to X = Y),$$

pero también podemos interpretar informalmente expresiones como ésta, entendiendo que significan lo que obviamente pretenden significar (en este caso, que dos colecciones con los mismos elementos son iguales).[1] Aunque ningún nombre se adecúa a todas las interpretaciones posibles, emplearemos la palabra "clase" para referirnos a las colecciones de objetos, si bien, según el contexto, "clase" deberá entenderse como "colección metamatemática", "conjunto de ZF*", "clase de ZF*", "conjunto de AP_2", etc., e igualmente "propiedad" deberá entenderse como "propiedad metamatemática bien definida" o como "fórmula de un determinado lenguaje formal".

El álgebra de clases Dadas dos clases A y B, se define su *unión* y su *intersección*, su *complemento* y su *diferencia* respectivamente como

$$A \cup B \equiv \{x \mid x \in A \vee x \in B\}, \qquad A \cap B \equiv \{x \mid x \in A \wedge x \in B\},$$

$$\bar{A} = \{x \mid x \notin A\}, \qquad A \setminus B = \{x \mid x \in A \wedge x \notin B\}.$$

La *clase universal* y la *clase vacía* se definen como:

$$V = \{x \mid x = x\}, \qquad \varnothing = \{x \mid x \neq x\}.$$

Se dice que una clase A *es una subclase de* B o que *está incluida en* una clase B si cumple

$$A \subset B \equiv \bigwedge x(x \in A \to x \in B).$$

Así, toda clase A cumple que $\varnothing \subset A \subset V$. Es inmediato comprobar las propiedades siguientes:

$$A \cup (B \cup C) = (A \cup B) \cup C, \quad A \cup B = B \cup A, \quad A \subset A \cup B,$$

$$A \cap (B \cap C) = (A \cap B) \cap C, \quad A \cap B = B \cap A, \quad A \cap B \subset A,$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C), \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

[1] A lo largo de este libro hemos expuesto numerosas teorías axiomáticas muy distintas entre sí que permiten formalizar (siempre trivialmente) los conceptos que vamos a presentar aquí, desde teorías de conjuntos, como la teoría básica de conjuntos B, o la teoría ZF*, en la que podemos interpretar las colecciones de objetos como conjuntos o también como clases de conjuntos definidas por fórmulas, o teorías como NBG* en la que las clases están también formalizadas, o teorías como AP_2 en la que los objetos son números naturales y las colecciones de objetos son los conjuntos de la teoría, etc.

Aplicaciones A partir de aquí necesitamos contar con que, para cada par de objetos x, y, existe otro objeto, que llamaremos *par ordenado* (x, y), de forma que se cumpla la relación fundamental:

$$\bigwedge xyzw((x, y) = (z, w) \leftrightarrow x = z \wedge y = w).$$

La definición concreta de los pares ordenados es irrelevante para todo lo que vamos a definir a continuación.[2]

Se define el *producto cartesiano* de dos clases A y B como la clase

$$A \times B \equiv \{x \mid \bigvee ab(a \in A \wedge b \in B \wedge x = (a, b))\}.$$

En general, usaremos la notación

$$\{(a, b) \mid \phi(a, b)\} \equiv \{x \mid \bigvee ab(x = (a, b) \wedge \phi(a, b))\}.$$

Por ejemplo, en estos términos podemos escribir, más brevemente:

$$A \times B \equiv \{(a, b) \mid a \in A \wedge b \in B\}.$$

Una clase F es una *función* si sus elementos son todos pares ordenados y un mismo conjunto x no aparece como primera componente de dos pares distintos en F, es decir,

$$F \text{ es una función} \equiv \bigwedge x \in F \bigvee uv\, x = (u, v)$$

$$\wedge \bigwedge uvw((u, v) \in F \wedge (u, w) \in F \to v = w).$$

Se define el *dominio* (rango) de una clase A como la clase de las primeras (segundas) componentes de los pares ordenados que pertenezcan a A, es decir,

$$\mathcal{D}A \equiv \{x \mid \bigvee y\,(x, y) \in A\}, \qquad \mathcal{R}A \equiv \{x \mid \bigvee y\,(y, x) \in A\}.$$

De este modo, si F es una función y $x \in \mathcal{D}F$, existe un único conjunto y tal que $(x, y) \in F$. Lo representaremos por $y = F(x)$ y lo llamaremos *imagen* de x por F. Formalmente, definimos

$$F(x) \equiv y \mid (x, y) \in F,$$

teniendo en cuenta que $F(x)$ puede ser una descripción impropia, pero sabemos que es propia siempre que F es una función y $x \in \mathcal{D}F$.

[2]La definición usual en las teorías de conjuntos es $(x, y) = \{\{x\}, \{x, y\}\}$, pero en otras teorías podemos tener definiciones alternativas. Por ejemplo, en la aritmética de Peano podemos definir pares ordenados aritméticos de números naturales(definición 5.14) que sirven como interpretación posible de los pares ordenados que vamos a considerar aquí, pues cumplen igualmente la relación fundamental. Si consideramos objetos definidos informalmente, no necesitamos definir los pares ordenados de ninguna forma particular: simplemente, siempre que tenemos dos objetos bien definidos x, y, podemos considerar que el par (x, y) es un nuevo objeto bien definido.

Una clase F es una *aplicación* de una clase A en una clase B si cumple[3]

$$F : A \longrightarrow B \equiv F \text{ es una función} \wedge \mathcal{D}F = A \wedge \mathcal{R}F \subset B.$$

De este modo, una aplicación $F : A \longrightarrow B$ asigna a cada $x \in A$ una única imagen $F(x) \in B$.

Una aplicación $F : A \longrightarrow B$ es *inyectiva* si $\bigwedge xy \in A(F(x) = F(y) \to x = y)$, es decir, si elementos distintos en A tienen imágenes distintas en B.

Cuando F es una función y $F(x) = y$, se dice también que x es una *antiimagen* de y por F.

Si $F : A \longrightarrow B$, cada elemento de b puede tener varias antiimágenes en A o no tener ninguna. Se dice que F es *suprayectiva* si $\mathcal{R}F = B$, es decir, si cada elemento de B tiene al menos una antiimagen en A.

Una aplicación $F : A \longrightarrow B$ es *biyectiva* si es a la vez inyectiva y suprayectiva, es decir, si cada elemento de A se corresponde con un único elemento de B y viceversa.

Notas A veces se define la *gráfica* de una aplicación $F : A \longrightarrow B$ como la clase $\{(x, F(x)) \mid x \in A\}$, pero conviene tener presente que, de acuerdo con las definiciones que hemos dado, la gráfica de F coincide con F.

También conviene observar que si $F : A \longrightarrow B$ y $B \subset C$, entonces también $F : A \longrightarrow C$, por lo que la noción de suprayectividad no depende únicamente de F, sino de F y de B. ∎

Si $F : A \longrightarrow B$ y $C \subset A$, se define $F[C] = \{F(x) \mid x \in C\} \subset B$. Formalmente,

$$F[C] \equiv \{b \in B \mid \bigvee c \in C\, b = F(c)\}.$$

Similarmente, si $D \subset B$ se define $F^{-1}[D] = \{x \in A \mid F(x) \in D\}$.

Es fácil probar que

$$F^{-1}[D_1 \cup D_2] = F^{-1}[D_1] \cup F^{-1}[D_2], \qquad F^{-1}[D_1 \cap D_2] = F^{-1}[D_1] \cap F^{-1}[D_2],$$

Así mismo $F[C_1 \cup C_2] = F[C_1] \cup F[C_2]$, pero $F[C_1 \cap C_2] \subset F[C_1] \cap F[C_2]$ y en general no se da la igualdad.

En general, para una clase arbitraria A definimos $A^{-1} \equiv \{(x, y) \mid (y, x) \in A\}$. De este modo, si $F : A \longrightarrow B$ biyectiva, se cumple que $F^{-1} : B \longrightarrow A$ biyectiva, y se dice que F^{-1} es la *aplicación inversa* de F.

Notemos que en este contexto tenemos dos definiciones distintas de $F^{-1}[D]$, pero ambas son equivalentes.

[3]Considerar a las aplicaciones como clases de pares ordenados es un convenio conjuntista necesario para formalizar el concepto en determinadas teorías de conjuntos, pero que no es imprescindible para que tenga sentido lo que sigue. En realidad basta con que para cada objeto $x \in A$ esté bien definido un único objeto $F(x) \in B$.

Dada una clase A, la aplicación $I_A : A \longrightarrow A$ dada por $\bigwedge x \in A\, I_A(x) = x$, se llama *identidad* en A. Si $A \subset B$, la identidad en A considerada como aplicación $A \longrightarrow B$ recibe el nombre de *inclusión* de A en B.

Si $F : A \longrightarrow B$ y $C \subset A$, se define la *restricción* de F a B como la clase $F|_C = F \cap (C \times B)$, de modo que $F|_C : C \longrightarrow B$ y $\bigwedge c \in C\, F|_C(c) = F(c)$.

En general, dadas dos clases A y B, definimos su *composición* como la clase

$$A \circ B \equiv \{(x, y) \mid \bigvee z((x, z) \in A \wedge (z, y) \in B)\}.$$

Es fácil ver que $(A \circ B) \circ C = A \circ (B \circ C)$. Si $F : A \longrightarrow B$ y $G : B \longrightarrow C$, entonces $F \circ G : A \longrightarrow C$ y se cumple[4] que $\bigwedge x \in A\, (F \circ G)(x) = G(F(x))$.

Relaciones Una clase R es una *relación* en una clase A si $R \subset A \times A$. En tal caso, en lugar de escribir $(x, y) \in R$, se escribe $x\, R\, y$ y se lee "x está relacionado con y". En estas condiciones:[5]

1. R es *reflexiva* si $\bigwedge x \in A\, x\, R\, x$.
2. R es *irreflexiva* si $\bigwedge x \in A\, \neg x\, R\, x$.
3. R es *simétrica* si $\bigwedge xy \in A(x\, R\, y \to y\, R\, x)$.
4. R es *antisimétrica* si $\bigwedge xy \in A(x\, R\, y \wedge y\, R\, x \to x = y)$.
5. R es *asimétrica* si $\bigwedge xy \in A(x\, R\, y \to \neg y\, R\, x)$.
6. R es *transitiva* si $\bigwedge xyz \in A(x\, R\, y \wedge y\, R\, z \to x\, R\, z)$.
7. R es *conexa* si $\bigwedge xy \in A(x\, R\, y \vee y\, R\, x)$.

Relaciones de equivalencia Una *relación de equivalencia* en una clase A es una relación reflexiva, simétrica y transitiva en A.

Si R es una relación de equivalencia en A y $a \in A$, se define la *clase de equivalencia* de a respecto de R como la clase $[a]_R \equiv \{x \in A \mid x\, R\, a\}$. Si no hay confusión suprimiremos el subíndice R.

De la reflexividad se sigue que $a \in [a]$, por lo que $[a] \neq \varnothing$. Así mismo es fácil probar que

$$\bigwedge xy \in A(x\, R\, y \leftrightarrow [x] = [y]),$$

$$\bigwedge xy \in A(\neg x\, R\, y \leftrightarrow [x] \cap [y] = \varnothing).$$

[4]Es frecuente definir $F \circ G$ de modo que $(F \circ G)(x) = F(G(x))$, pero, cuando se trabaja con muchas aplicaciones, es mucho más fácil invertir el orden cuando se deshace una composición que cuando se ha de plasmar por escrito una composición cuyo esquema está claro mentalmente.

[5]Como en el caso de las aplicaciones, considerar que una relación es una clase de pares ordenados no es imprescindible en todos los contextos. Para que lo que sigue tenga sentido basta con que, para cada par de objetos $a, b \in A$, esté bien definida la afirmación $a\, R\, b$.

Relaciones de orden Una *relación de orden* en una clase A es una relación reflexiva, antisimétrica y transitiva en A. A veces se dice también que R es una *relación de orden parcial,* mientras que una *relación de orden total* es una relación de orden conexa.

Es habitual representar las relaciones de orden mediante el signo $\leq$, entendiendo que éste hace referencia a relaciones distintas según el contexto.

Una *relación de orden estricto* en una clase A es una relación irreflexiva, asimétrica y transitiva. Es claro que si $\leq$ es una relación de orden no estricto en una clase A, entonces la relación dada por $x < y \equiv (x \leq y \wedge x \neq y)$ es una relación de orden estricto en A y, recíprocamente, si $<$ es una relación de orden estricto en A, entonces la relación dada por $x \leq y \equiv (x < y \vee x = y)$ es una relación de orden no estricto en A, por lo que ambos conceptos son equivalentes.

Sea A una clase ordenada por la relación $\leq$ y sea $B \subset A$. Entonces:

1. $M \in A$ es una *cota superior* de B si $\bigwedge x \in B\, x \leq M$,
2. $m \in A$ es una *cota inferior* de B si $\bigwedge x \in B\, m \leq x$,
3. $M \in A$ es un *maximal* de B si $M \in B$ y $\bigwedge x \in B(M \leq x \to M = x)$.
4. $m \in A$ es un *minimal* de B si $m \in B$ y $\bigwedge x \in B(x \leq m \to x = m)$.
5. $M \in A$ es el *supremo* de B si M es una cota superior de B y $\bigwedge x \in A(x$ es una cota superior de $B \to M \leq x)$.
6. $m \in A$ es el *ínfimo* de B si m es una cota inferior de B y $\bigwedge x \in A(x$ es una cota inferior de $B \to x \leq m)$.
7. $M \in A$ es el *máximo* de B si $M \in B$ y M es una cota superior de B.
8. $m \in A$ es el *mínimo* de B si $m \in B$ y m es una cota inferior de B.

Es fácil ver que en un conjunto totalmente ordenado todo maximal es máximo y todo minimal es mínimo. Si un conjunto tiene máximo o mínimo, supremo o ínfimo, entonces éstos son únicos. El supremo (ínfimo) de una clase es máximo (mínimo) si y sólo si pertenece a la clase.

Cuando tenemos una clase A ordenada por una relación $\leq$ y una subclase $B \subset A$, consideramos, aunque no se indique explícitamente, que B está ordenada por la restricción de $\leq$ a B, es decir, con la intersección de $\leq$ con $B \times B$, de modo que si x, $y \in B$, se cumple $x \leq y$ como elementos de B si y sólo si se cumple como elementos de A.

Es inmediato comprobar que esta restricción es un orden en B. Más aún, B está totalmente ordenada si A lo está.

Una aplicación $F : A \longrightarrow B$ entre dos clases ordenadas por respectivas relaciones de orden $\leq_1$ y $\leq_2$ es *monótona creciente* o, simplemente, *creciente* (respecto a dichas relaciones), si

$$\bigwedge xy \in A(x \leq_1 y \to F(x) \leq_2 F(y)).$$

Se dice que F es *monótona decreciente* o *decreciente* si cumple

$$\bigwedge xy \in A(x \leq_1 y \to F(y) \leq_2 F(x)).$$

Se dice que F es *estrictamente* monótona creciente o decreciente si se cumple esto mismo cambiando las desigualdades no estrictas $\leq$ por desigualdades estrictas $<$.

A.2 Otros conceptos conjuntistas

Mientras los resultados de la sección anterior son formalizables en cualquiera de las teorías que hemos considerado en este libro dotadas de una relación de pertenencia, recogemos aquí unos pocos resultados adicionales cuya formalización puede requerir algunos supuestos adicionales.

Clases cociente Si R es una relación de equivalencia en una clase A, definimos la *clase cociente* de A respecto a R como la clase A/R de todas las clases de equivalencia de R, es decir,

$$A/R \equiv \{[x]_R \mid x \in A\}.$$

La dificultad que presenta la formalización de este concepto es que requiere que unas clases (en este caso las clases de equivalencia de R) puedan ser elementos de otras clases (en este caso de la clase cociente A/R) y no todas las teorías permiten esto.

En general, una clase C (cuyos elementos sean clases) es una *partición* de una clase A si cumple

1. $\bigwedge x \in C(x \subset A \wedge x \neq \varnothing)$,
2. $\bigwedge a \in A \bigvee x \in C\, a \in x$,
3. $\bigwedge xy \in C(x = y \vee x \cap y = \varnothing)$.

Tenemos, pues, que si R es una relación de equivalencia en una clase A, entonces la clase cociente A/R es una partición de A.

Ejercicio: Probar que si C es una partición de una clase A, entonces existe una relación de equivalencia R en A tal que $C = A/R$.

Buenos órdenes Un *buen orden* en una clase A es una relación de orden $\leq$ en A respecto a la cual toda subclase de A no vacía tiene un mínimo elemento. (También se dice que A está *bien ordenada* por $\leq$.) Todo buen orden es un orden total, pues si x, $y \in A$, tendremos $x \leq y$ o $y \leq x$ según quién sea el mínimo de $\{x, y\}$.

La dificultad de este concepto es que no todas las teorías axiomáticas permiten formalizar el concepto de "toda subclase". Por ejemplo, cuando en la

teoría B demostramos que todo ordinal x está bien ordenado por la inclusión (teorema 3.14), hay que entender que todo subconjunto de x no vacío tiene un mínimo elemento, y cuando en 3.18 demostramos que la clase Ω de todos los ordinales es un ordinal, el resultado previo debe entenderse en principio como que todo subconjunto de Ω (no toda subclase) no vacío tiene mínimo elemento, pues sólo esto puede expresarse mediante una fórmula de $\mathcal{L}_{\mathrm{tc}}$ y demostrarse en la teoría básica B. Ahora bien, luego resulta que, para cada subclase $A \subset \Omega$ no vacía (definida mediante una fórmula $\phi(x)$) puede probarse que tiene mínimo elemento, pues si $\alpha \in A$, o bien α es el mínimo de A, o bien $A \cap \alpha$ es un subconjunto de Ω no vacío, que tendrá mínimo elemento, y dicho mínimo será también claramente el mínimo de A.

A.3 La jerarquía de Lévy

En esta sección recopilamos los resultados sobre la complejidad de los conceptos básicos de la teoría de conjuntos respecto de la jerarquía de Lévy definida en 6.1.

Conceptos $\boldsymbol{\Delta_0}$ en la teoría básica B (luego también en KP y en ZF^*):

1. $z = x \cup y \leftrightarrow \bigwedge u \in z\,(u \in x \vee u \in y) \wedge \bigwedge u \in x\ u \in z \wedge \bigwedge u \in y\, u \in z$,
2. $z = x \cap y \leftrightarrow \bigwedge u \in z\,(u \in x \wedge u \in y) \wedge \bigwedge u \in x(u \in y \to u \in z)$,
3. $z = x \setminus y \leftrightarrow \bigwedge u \in z\,(u \in x \wedge u \notin y) \wedge \bigwedge u \in x(u \notin y \to u \in z)$,
4. $z = \varnothing \leftrightarrow \bigwedge u \in z\ u \neq u$,
5. $z = \bigcup x \leftrightarrow \bigwedge u \in z \bigvee v \in x\ u \in v \wedge \bigwedge v \in x \bigwedge u \in v\ u \in z$,
6. $x \subset y \leftrightarrow \bigwedge u \in x\ u \in y$,
7. $w = \{u, v\} \leftrightarrow v \in w \wedge v \in w \wedge \bigwedge x \in w\,(x = u \vee x = v)$,
8. $w = (u, v) \leftrightarrow \bigvee rs \in w(r = \{u\} \wedge s = \{u, v\})$
   $\wedge \bigwedge x \in w\ (x = \{u\} \vee x = \{u, v\})$,
9. $y = x' \leftrightarrow \bigwedge u \in y\ (u \in x \vee u = x) \wedge \bigwedge u \in x\ u \in y \wedge x \in y$,
10. x es transitivo $\leftrightarrow \bigwedge u \in x\ u \subset x$,
11. x es $\in$-conexo $\leftrightarrow \bigwedge uv \in x(u \in v \vee v \in u \vee u = v)$,
12. r es una relación $\leftrightarrow \bigwedge z \in r \bigvee w \in z \bigvee uv \in w\ z = (u, v)$,
13. r es una relación en $a \leftrightarrow \bigwedge z \in r \bigvee uv \in a\ z = (u, v)$,
14. f es una función $\leftrightarrow f$ es una relación $\wedge \bigwedge xy \in f \bigwedge r \in x \bigwedge s \in y$
    $\bigwedge uv \in r \bigwedge w \in s(x = (u, v) \wedge y = (u, w) \to v = w)$,
15. $f : x \longrightarrow y \leftrightarrow f$ es una función $\wedge \bigwedge z \in f \bigvee u \in x \bigvee v \in y\ z = (u, v)$
    $\wedge \bigwedge u \in x \bigvee v \in y \bigvee z \in f\ z = (u, v)$.

Dejamos como ejercicio probar el carácter Δ_0 de las fórmulas "$f : x \longrightarrow y$ es una aplicación inyectiva, suprayectiva, biyectiva", así como "r es una relación en a reflexiva, irreflexiva, simétrica, antisimétrica, transitiva, conexa, de equivalencia, de orden, de orden total". Ninguna presenta dificultad teniendo en cuenta los resultados precedentes.

Ordinales La fórmula "x está bien fundado" es Π_1 en la teoría B, pues equivale a

$$\bigwedge u(u \subset x \wedge u \neq \varnothing \to \bigvee v \in u \bigwedge w \in v \; w \notin u).$$

y lo mismo vale para $x \in \Omega$. Sin embargo, en cualquier teoría en la que se demuestre que todo conjunto está bien fundado (por ejemplo, en KP o en ZF^* más el axioma de regularidad) tenemos que $x \in \Omega$ es Δ_0, pues

$$x \in \Omega \leftrightarrow x \text{ transitivo y } \in\text{-conexo}.$$

A su vez, en cualquier teoría en la que $x \in \Omega$ sea Δ_0, también son Δ_0 las fórmulas siguientes:

1. x es un ordinal sucesor $\leftrightarrow x \in \Omega \wedge \bigvee y \in x \; x = y'$,
2. x es un ordinal límite $\leftrightarrow x \in \Omega \wedge x \neq \varnothing \wedge x$ no es un ordinal sucesor,
3. $x \in \omega \leftrightarrow x \in \Omega \wedge \bigwedge u \in x(u = 0 \vee u \text{ sucesor}) \wedge (x = 0 \vee x \text{ sucesor})$.

Observemos que la relación de orden (estricto) en los ordinales es Δ_0 porque no es más que la inclusión (resp. la pertenencia).

Si suponemos el axioma de infinitud (de modo que ω es un conjunto) tenemos además que ω es Δ_0, pues

$$y = \omega \leftrightarrow y \text{ es un ordinal límite} \wedge \bigwedge u \in y \; (u = 0 \vee u \text{ es sucesor}).$$

Conceptos Δ_0 en cualquier extensión de B que pruebe su existencia (por ejemplo en KP o en Z^*):

1. $z = x \times y \leftrightarrow \bigwedge u \in x \bigwedge v \in y \bigvee w \in z \; w = (u, v)$
   $\wedge \bigwedge w \in z \bigvee u \in x \bigvee v \in y \; w = (u, v)$,
2. $z = \mathcal{D}x \leftrightarrow \bigwedge w \in z \bigvee u \in x \bigvee r \in w \bigvee v \in r \; w = (u, v)$
   $\wedge \bigwedge u \in x \bigvee w \in z \bigvee r \in w \bigvee v \in r \; w = (u, v)$,
3. $z = \mathcal{R}x \leftrightarrow \bigwedge w \in z \bigvee v \in x \bigvee r \in w \bigvee u \in r \; w = (u, v)$
   $\wedge \bigwedge v \in x \bigvee w \in z \bigvee r \in w \bigvee u \in r \; w = (u, v)$,
4. $y = f[x] \leftrightarrow \bigwedge v \in y \bigvee u \in x \bigvee z \in f \; z = (u, v)$
   $\wedge \bigwedge z \in f \bigwedge u \in x \bigwedge w \in z \bigwedge v \in w(z = (u, v) \to v \in y)$,

5. $x = f^{-1}[y] \leftrightarrow \bigwedge u \in x \bigvee v \in y \bigvee z \in f\ z = (u,v))$
   $\wedge \bigwedge v \in y \bigwedge z \in f \bigwedge w \in z \bigwedge u \in w(z = (u,v) \to u \in x)$,
6. $y = x^{-1} \leftrightarrow \bigwedge z \in y \bigvee w \in z \bigvee uv \in w \bigvee z' \in x(z = (u,v) \wedge z' = (v,u))$
   $\wedge \bigwedge z \in x \bigwedge w \in z \bigwedge uv \in w(z = (u,v) \to \bigvee z' \in y\ z' = (v,u))$,
7. $g = f|_x \leftrightarrow g \subset f \wedge \bigwedge z \in f \bigwedge w \in z \bigwedge uv \in w(z = (u,v) \wedge u \in x \to z \in g)$
   $\wedge \bigwedge z \in g \bigvee w \in z \bigvee u \in x \bigvee v \in w\ z = (u,v)$,
8. $h = f \circ g \leftrightarrow \bigwedge z \in h \bigvee z' \in f \bigvee z'' \in g \bigvee w' \in z' \bigvee w'' \in z'' \bigvee uu' \in w'$
   $\bigvee u'' \in w''(z = (u,u'') \wedge z' = (u,u') \wedge z'' = (u',u''))$
   $\wedge \bigwedge z' \in f \bigwedge z'' \in g \bigvee z \in h \bigvee w' \in z' \bigvee w'' \in z'' \bigvee uu' \in w' \bigvee u'' \in w''$
   $(z = (u,u'') \wedge z' = (u,u') \wedge z'' = (u',u''))$.

Conjuntos finitos La definición de conjunto finito es Σ_1:

$$x \text{ es finito} \leftrightarrow \bigvee f n(n \in \omega \wedge f : n \longrightarrow x \text{ biyectiva}),$$

mientras que la definición de conjunto D-finito (11.6) es Π_1:

$$x \text{ es } D\text{-finito} \leftrightarrow \neg \bigvee f(f : x \longrightarrow x \text{ inyectiva no suprayectiva}).$$

Bajo el axioma de elección ambas definiciones son equivalentes, por lo que la finitud se vuelve Δ_1.

No obstante, incluso sin AE, los términos $x^{<\omega}$ y $\mathcal{P}^f x$, que son definibles tanto en ZF* + AI como en KPI, son Δ_1, pues

$$y = x^{<\omega} \leftrightarrow \bigvee f w(w = \omega \wedge f : w \longrightarrow V \wedge y = \bigcup \mathcal{R} f \wedge f(0) = \{\varnothing\} \wedge$$
$$\bigwedge n \in w \bigwedge s \in f(n+1) \bigvee t \in f(n) \bigvee a \in x\ s = t \cup \{(n,a)\} \wedge$$
$$\bigwedge n \in w \bigwedge x \in f(n) \bigwedge a \in x \bigvee t \in f(n+1)\, t = s \cup \{(n,a)\})$$

y

$$y = \mathcal{P}^f x \leftrightarrow \bigvee z(z = x^{<\omega} \wedge \bigwedge u \in y \bigvee v \in z\ u = \mathcal{R} v \wedge \bigwedge v \in z \bigvee u \in y\ u = \mathcal{R} v).$$

(Esto prueba que ambos términos son Σ_1, luego Δ_1).

Σ_1-recursión En la prueba del teorema 12.4 se ve que la fórmula $y = \mathrm{ct}\, x$ es Δ_1 en KP. El teorema 12.7 afirma que las funciones definidas en KP por recurrencia a partir de funciones Σ_1 son Σ_1 y (si su dominio es Σ_1, en particular si es un conjunto) son Δ_1. De aquí se sigue que la suma y el producto de números naturales, así como la suma de ordinales, son Δ_1 en KP.

Bibliografía

[1] Baker, A. *Breve Introducción a la Teoría de Números,* Alianza Universidad, Madrid, 1986.

[2] Barwise, J. *Admissible sets and structures,* Springer 1975.

[3] Barwise, J. (editor), *Handbook of Mathematical Logic,* North Holland, Amsterdam, 1977.

[4] Beckmann, A. y Buss, S.R. *Corrected upper bounds for free-cut elimination* Theor. Comput. Sci.(2011) 5433–5445

[5] Bernays, P. y Fraenkel, A. *Axiomatic Set Theory,* North Holland, Amsterdam, 1958.

[6] Buss, S.R. (editor) *Handbook of Proof Theory,* Elsevier, Amsterdam, 1998.

[7] Cohen, P. *Set Theory and the Continuum Hypothesis,* W.A.Benjamin inc. reading, New York, 1966.

[8] Davis, M. *Hilbert's Tenth Problem is Unsolvable,* Am. Math. Monthly **80** (1973) pp. 233–269.

[9] Devlin, K. J. *Fundamentals of Contemporary Set Theory.* Springer, New York.

[10] Enderton, H. B. *Elements of Recursion Theory,* (en Barwise).

[11] Gödel, K. *Obras completas,* Alianza Universidad, Madrid, 1981.

[12] — *Sobre Sentencias Formalmente Indecidibles de Principia Mathematica y Sistemas Afines,* (1931).

[13] — *Sobre Sentencias Indecidibles de Sistemas Formales Matemáticos,* (1934).

[14] — *La Consistencia del Axioma de Elección y de la Hipótesis Generalizada del Continuo con los Axiomas de la Teoría de Conjuntos,* (1940).

[15] Hájek, P. y Pudlák, P. *Metamathematics of First-Order Arithmetic,* Springer, Berlín, 1998.

[16] Hamilton, A. G. *Lógica para Matemáticos,* Paraninfo, Madrid, 1981.

[17] Jech, T.J. *The Axiom of Choice,* North Holand, Amsterdam, 1973.

[18] — *Set Theory.* Academic Press, New York, 1978.

[19] Kaye, R. *Models of Peano Arithmetic,* Clarendon Press, Oxford, 1991.

[20] Kleene, S. C. *Introducción a la Metamatemática,* Tecnos, Madrid, 1974.

[21] Kunen, K. *Set Theory. An Introduction to Independence Proofs,* North Holland, Amsterdam, 1985.

[22] Moschovakis, Y.N. *Descriptive Set Theory,* (segunda edición) AMS 2009.

[23] Mosterín, J. *Lógica de Primer Orden,* Ariel, Barcelona, 1970.

[24] — *Teoría Axiomática de Conjuntos,* Ariel, Barcelona, 1971.

[25] Smorynski, C. *The Incompleteness Theorems,* (en Barwise).

Índice de Materias