

Selección de palabras de contenido referencial. Una propuesta metodológica

María Claudia Scurtu*, Vicente Manzano Arrondo
y Juan Francisco Rodríguez Testal

Universidad de Sevilla

En la investigación en Psicología, es habitual recurrir a listas de palabras que tienen contenido emocional específico. La literatura sobre este tema muestra disparidad de criterios y prácticas para construir y utilizar las listas. Este trabajo propone un procedimiento concreto orientado a facilitar la creación de conjuntos de palabras de contenido emocional. Con este objetivo, se realiza una revisión del estado de la cuestión y se articula un procedimiento en tres fases: (1) construcción inicial de una lista, que siga los criterios de especificidad, activación variable, consideración de constructos de control, diversidad formal y cantidad pertinente; (2) cribado experto, siguiendo las etapas de selección inicial de expertos, selección final, recogida de datos cuantitativos, y análisis mediante el índice I , y (3) selección final de elementos mediante muestreo de participantes. El procedimiento se ejemplifica sintéticamente y paso a paso, mediante un estudio empírico sobre pensamiento referencial.

La investigación experimental en Psicología requiere, en numerosas ocasiones, disponer de una lista de estímulos potencialmente generadores de respuestas emocionales. Entre la amplia variedad de materiales, es frecuente recurrir a conjuntos de palabras que varían, entre otras características, en valencia (emoción positiva, negativa o neutra) y la activación que provoca las mismas (Scott, O'Donnell, & Sereno, 2012).

Hasta la fecha, se ha realizado un esfuerzo notable para estudiar el contenido emocional de bases de palabras amplias, de tal forma que sirvan como banco de recursos para todo tipo de objetivos en la investigación acerca de las respuestas emocionales. Aunque hay bases más amplias,

* Correspondencia: Dr. María Claudia Scurtu. Departamento de Personalidad, Evaluación y Tratamiento Psicológicos. Facultad de Psicología. Universidad de Sevilla. c/ Camilo José Cela s/n. 41018, Sevilla, Spain; Telf: +34 697286043; E-mail: clauditza_psi@yahoo.com

- Warriner, Kuperman, & Brysbaert (2013), por ejemplo, propusieron una base con 13915 palabras - el más conocido y utilizado es el *Affective Norms for English Words* (ANEW), de Bradley, & Lang (1999a). En su trabajo, estos autores encadenaron varios estudios en los que estudiantes de Psicología evaluaban mediante la escala pictográfica SAM el contenido de listas de 100 a 150 palabras según agrado (figura sonriente/feliz hasta infeliz), activación (figura emocionada, con los ojos abiertos hasta relajada) y dominancia (figura amplia/control hasta figura pequeña/dominada), hasta completar un total de 1034 palabras. El ANEW se ha utilizado, en el recuerdo de la información emocional condicionada por el estado de ánimo (Hamman, & Mao, 2002; Lewis, Critchley, Smith, & Dolan, 2005), la actividad de los músculos faciales (Larsen, Norris, & Cacioppo, 2003), la memoria (Kensinger, Brierley, Medford, Growdon, & Corkin, 2002; Kensinger, & Corkin, 2003), o las cogniciones y actitudes políticas (Lodge, & Taber, 2005).

Redondo, Fraga, Comesaña & Perea (2007) tradujeron la base de palabras emocionales baremada del ANEW al español, añadiendo otros índices psicolingüísticos de los cuales cinco factores objetivos (número de letras, número de sílabas, clase gramatical, frecuencia de uso y vecindad ortográfica) y tres factores subjetivos (familiaridad, concreción y imaginabilidad). No obstante, diversos estudios muestran que: (1) el efecto emocional de las palabras en español o en inglés no son necesariamente coincidentes (Altarriba, & Bauer, 2004), y (2) una buena traducción de las palabras no garantiza la validez del resultado si éste no se somete de nuevo a estudios psicométricos específicos (Carretero-Dios, & Pérez, 2005).

En 2005, Redondo, Fraga, Comesaña, & Perea (2005) analizaron 478 sustantivos en castellano, con sus respectivos valores de valencia (desde lo agradable a lo desagradable), nivel o intensidad de activación (desde la calma hasta la excitación) y control o dominancia mediante, menos consistente y con un peso mucho menor (desde figura sobredimensionada hasta diminuta), mediante el Maniquí de Autoevaluación (Self-Assessment Manikin - SAM). Sin embargo consideramos que lo referencial se expresa y reconoce también a través de adjetivos y verbos que la lista baremada citada no los incluye.

A pesar de estas y otras iniciativas, y teniendo en cuenta la cercanía emocional entre las palabras de contenido depresivo, paranoide y referencial, las bases de palabras baremadas existentes nos resultaron generalistas, incompletas en cuanto a la categoría gramatical o poco discriminantes para investigar un constructo clínico concreto, como es el pensamiento referencial (PR). El concepto de PR abarca a diferentes tipos

de procesamientos autorreferentes (para una revisión: Lenzenweger, Bennett, & Lilenfeld, 1997, Senín-Calderón, Rodríguez-Testal, & Salvador-Perona, 2014). En particular interesa aquí aludir a las ideas de referencia o autorreferencias, como ejemplos de PR que se sitúan entre los síntomas básicos y prodrómicos de los trastornos psicóticos. Una autorreferencia es una atribución autodirigida sobre algo de lo que sucede en el entorno, por ejemplo, dos personas hablan: “Se están burlando de mí”. Hay listados de palabras de contenido depresivo y de tipo persecutorio (Jiménez, Vázquez, & Hernán, 1998; Nieto-Moreno, Hervás, & Vázquez, 2006), pero no específicamente acerca del PR, por lo que nuestra propuesta puede representar una aportación de interés clínico y de investigación. Así, en el estudio realizado por Nieto-Moreno, Hervás, & Vázquez (2006) en cuanto al contenido amenazante y depresivo de palabras en castellano, la palabra *vigilado* tiene un índice de especificidad amenazante muy alto (3,55), el mismo término encontrándose en la posición 15 de 41 (índice I_t de 0.71) en la lista de palabras de contenido referencial que proponemos (ver Anexo).

Por lo que consideramos más útil elaborar conjuntos específicos según el objetivo que se pretenda en la investigación. Generalmente, los investigadores elaboran listas de palabras a partir de la identificación de palabras contenidas en cuestionarios, escalas o inventarios específicos - como ocurre en Foa, McNally, & Murdock (1989) - o recurriendo a los participantes en la investigación, bien sea en el transcurso de la situación experimental (Blanch, & Baños, 1996; Derry, & Kuiper, 1981; Ingram, Lumry, Cruel, & Sieber, 1987) o mediante una fase previa en la que los participantes seleccionan las palabras a partir de un conjunto inicial más extenso (McNally, Amir, Louro, Lukach, Riemann, & Calamari, 1994). No obstante, diversos estudios muestran que la exposición previa a las palabras genera efectos de activación emocional (priming) en los participantes en la situación experimental posterior (Brouillet, & Syssau, 2005; Carroll, & Young, 2005; Van Strien, & Morpurgo, 1992).

En general, las investigaciones que utilizan palabras como estímulos activadores de emociones se enfrentan a varios problemas metodológicos: la diversidad de circunstancias personales que matiza el contenido emocional de las palabras (Smith, & Ellsworth, 1985), la abundancia de estructuras de clasificación de emociones que dificulta comparar hallazgos (Yik, Russell, & Barrett, 1999), o la falta de acuerdo en el proceso de selección o formación del conjunto de palabras definitivo que será utilizado en la investigación. Por ello, y aunque cabe destacar una notable diversidad de situaciones de investigación, sería deseable contar con un procedimiento consensuado que sirva de guía general para la construcción de listas de palabras específicas. El objetivo de este trabajo es proponer un esquema de

actuación concreto que pueda cumplir esa función y que denominamos *procedimiento de selección de palabras* (PSP). La propuesta se articula a través de tres momentos secuenciales: construcción de una lista inicial, cribado de unidades a partir de juicios expertos, y selección final mediante muestreo de participantes equivalentes.

La exposición del PSP se acompaña de un ejemplo concreto: la construcción de una lista de palabras de contenido referencial.

Construcción de una lista de palabras inicial

Como se ha indicado, la construcción de un conjunto de palabras inicial puede seguir patrones diversos, bien sea de forma aislada o combinando fuentes:

- Equipo de investigación.
- Identificación de palabras a partir de cuestionarios específicos al uso.
- Selección de unidades desde un banco de palabras previo.
- Construcción del conjunto como resultado de una fase experimental, donde personas del mismo perfil suministran los elementos como resultado de una tarea específica.

Con independencia del procedimiento básico o combinado que se utilice para la construcción del conjunto inicial de unidades, es viable establecer una serie de criterios que condicionen el resultado, matizando las sugerencias contenidas en trabajos previos (Bradley, & Lang, 1999a; Greenberg, Vazquez, & Alloy, 1988; Kousta, Vinson, & Vigliocco, 2009; Mehrabian, & Russell, 1974; Mikels, Fredrickson, Larkin, Lindberg, Maglio, & Reuter-Lorenz, 2005; Smith, & Ellsworth, 1985; Verona, Sprague, & Sadeh, 2012; Yik, Russell, & Barrett, 1999). Estos criterios-guía son:

- *Especificidad*. Las palabras son seleccionadas atendiendo al constructo que se pretende estudiar.
- *Activación variable*. Las palabras han de estimular emociones con diferente intensidad. La versión mínima de este criterio establece el par *unidades activadoras* (suscitan emociones) – *unidades neutras* (no las suscitan).
- *Constructos de control*. Es difícil asumir la unidimensionalidad estricta de cada palabra seleccionada. Es muy posible que las unidades sean capaces de activar emociones diversas que interactúen entre sí. Asumir este fenómeno implica: (1) considerar al mismo tiempo varios constructos psicológicos que podrían explicar los efectos, y (2)

combinar subconjuntos de palabras específicas en la situación experimental.

- *Diversidad formal.* Con el objetivo de facilitar la generalización y el control, las unidades han de (1) ser diversas en características como longitud, familiaridad, categoría gramatical o dificultad de pronunciación, y (2) estar representadas adecuadamente en número, sin que covaríen estas características con las condiciones experimentales.
- *Cantidad pertinente.* Tanto la tarea a la que han de enfrentarse los participantes en el estudio, como la que han de llevar a cabo los expertos, aconsejan contar con el mínimo de palabras que cumplan con los criterios expuestos. Un exceso innecesario de unidades puede interferir en la ejecución de las tareas. No obstante, el mínimo ha de ser ampliado en la lista inicial, puesto que el proceso implicará presumiblemente pérdida de palabras. Como criterio general, se aconseja un 50% más de unidades iniciales de las que se considera apropiado para la situación final de investigación.

Aplicación

La utilización de PSP para elaborar un listado de palabras de contenido referencial comenzó con una colección inicial de unidades desde varias fuentes: palabras referenciales escogidas de la escala SPQ (*Scale for the Assessment of Schizotypal Personality*, Raine, 1991), del Cuestionario REF (*The Referential Thinking Scale*, Lenzenweger, et al., 1997), del SSQ (*The Schizotypic Syndrome Questionnaire*, van Kampen, 2006), del diccionario de la lengua española (vigésima edición, Real Academia Española, 2001), del diccionario de sinónimos y antónimos (Wordreference, 2005) y del Manual LEXESP (Sebastián, Martí, Carreiras, & Cuetos, 2000). Además de palabras de contenido referencial, la lista contemplaba también términos de contenido depresivo, señalados en la literatura específica por su relación con el PR (Wing, Cooper, & Sartorius, 1974). La lista inicial fue inspeccionada por el equipo de investigación y reducida a 164 unidades. Se consideró contar con unas 40 palabras para la prueba definitiva, y una pérdida aceptable del 50% para cada fase (de la lista inicial al cribado de palabras por juicio experto; y de éste a la muestra equivalente), lo que implicaba una lista de partida con alrededor de 160 unidades.

Juicios expertos

Existe un amplio consenso a la hora de considerar el juicio de personas expertas como un buen recurso para fundamentar los elementos

que formarán parte de un instrumento de medida para variables psicológicas (van der Salm, van Knippenberg, & Daamen, 1997). PSP parte de este recurso, basándose en los consejos de Escobar-Pérez, & Cuervo-Martínez (2008), Landeta (2001), Neijens, Ridder, & Saris (1992), Skjong, & Wentworth (2001) y Weinstein (1993) para la etapa 1 del trabajo de expertos, que va desde el diseño de la consulta hasta la recogida de datos. En la etapa 2, de cribado de expertos, PSP propone la selección de personas cuyo comportamiento evaluador vaya en la línea del conjunto, utilizando para ello el Alfa de Cronbach. En la etapa 3, de cribado de palabras, PSP propone la utilización del índice I . Para una visión más amplia sobre posibilidades de análisis en el manejo de información proveniente de expertos o jueces, puede consultarse el texto de Ato, Benavente, & López (2006).

Etapas 1. Selección de expertos y recogida de datos

El trabajo con personal experto implica varios cometidos: selección, instrucción y recogida de datos. La selección ha de estar guiada por tres categorías de criterios: (1) experticia, es decir, dominio del asunto, tanto respecto al conocimiento suficiente de los constructos estudiados como de la experiencia práctica en su manejo; (2) viabilidad metodológica, lo que a su vez implica disponibilidad para aceptar y llevar a cabo la tarea en el tiempo requerido, y motivación para realizarla correctamente; y (3) diversidad, es decir, consideración de distintos perfiles, de tal forma que se garantice un mínimo de variabilidad de perspectivas o experiencias individuales.

La instrucción se concreta en un texto breve y claro donde se especifica: (1) el objetivo de la investigación, (2) la fase en la que interviene el personal experto, (3) la tarea específica que han de realizar, y (4) el modo de relación con el equipo del estudio. La instrucción no es solo un vehículo para posibilitar la tarea, es también una información precisa que permite al personal experto valorar si cuenta con la experticia y la viabilidad metodológica que se requiere.

La recogida de datos se lleva a cabo mediante un instrumento diseñado habitualmente para ese fin. En el caso de construcción de una lista de palabras, el instrumento consta básicamente de una matriz donde las filas están ocupadas por las palabras propuestas para la valoración del constructo específico investigado, mientras que las columnas contienen la escala de valoración. Cada experto valora en qué medida esa unidad suscita esa emoción o hace referencia a ese constructo concreto, utilizando para ello una misma escala, que PSP aconseja sea cuantitativa.

Aplicación. En la elaboración del listado de palabras de contenido referencial, se recurrió a profesionales de centros de salud y a la red de contactos personales de los miembros del equipo de investigación, hasta completar un conjunto de 20 personas con experticia (con una media de 28 años de experiencia profesional) y viabilidad metodológica. Siguiendo el criterio de diversidad, se procuró contar con expertos de las dos disciplinas con mayor competencia en este campo (14 de Psicología clínica y 6 de Psiquiatría), así como mujeres (11) y hombres (9), y edades comprendidas entre 26 y 63 años. En el proceso de captación, los expertos fueron informados sobre la naturaleza del estudio y el cometido específico que se les pedía. Esta información se incluyó de nuevo, de forma sintética, en el documento de trabajo, que contenía: (1) los objetivos de la investigación, (2) una descripción del PR, aunque todos los expertos contaban con experiencia en psicosis y estados mentales de alto riesgo, (3) instrucciones para cumplimentar el cuestionario, y (4) el cuestionario matriz con las 164 palabras (de contenido depresivo y referencial) en filas y en columnas una escala de 1 a 10, en cuyas casillas debían indicar el contenido referencial de las palabras. Los 20 expertos respondieron a la tarea en los plazos establecidos.

Etapa 2. Cribado de expertos

Es importante analizar las respuestas del grupo de expertos para identificar en qué medida existe acuerdo o son identificables varias perspectivas. Este segundo caso es más propio de entornos de predicción e incertidumbre, abordados en la técnica Delphi (Landeta, 2002). En la tarea de identificar el contenido emocional de una lista de palabras, es más difícil esperar que existan diferentes perspectivas, que afecten al conjunto del listado. Por este motivo, es recomendable realizar algún tipo adecuado de análisis de datos que permita identificar excepciones individuales a la norma grupal. Si fuera el caso, habría que prescindir de las respuestas de los expertos señalados por el análisis, para continuar el proceso con suficientes garantías.

Al recurrir, como se ha indicado más arriba a una escala cuantitativa para expresar el grado en que una palabra activa una emoción específica o hace referencia a un constructo en concreto, entonces cada una de las k palabras, respecto a cada emoción considerada, cuenta con un conjunto de n puntuaciones, cada una de ellas proveniente de cada uno de los n expertos. Ello implica la utilización de una matriz de datos con n filas y k columnas. Con esta información de partida, la propuesta PSP es:

1. Transponer la matriz, es decir, generar un resultado de k filas y n columnas, donde la columna i contiene los k juicios del experto i para cada una de las k palabras, y la fila j muestra los n juicios emitidos sobre la palabra j por parte de cada uno de los n expertos.
2. Aplicar el Alfa de Cronbach (1951) de consistencia interna, utilizando cada experto como si fuera un ítem de un cuestionario, de tal modo que Alfa expresa el grado en que el conjunto de los ítems-expertos comparten un juicio normativo en el conjunto de palabras.
3. Poner en marcha el proceso iterativo al uso (Ledesma, Molina, & Valero, 2002), en el que se eliminan expertos hasta que el valor de Alfa es suficientemente elevado o no puede ser mayor.

Como refieren, entre otros, Carretero-Dios & Pérez (2005), existe un hábito generalizado contraproducente respecto al uso del Alfa de Cronbach. Por lo general, se van eliminando ítems de un cuestionario hasta que progresivamente el valor de Alfa vaya tocando techo. No obstante, este hábito genera en no pocas ocasiones que los ítems finales constituyan una mera variación lingüística de una misma aseveración, de tal forma que un valor superior a 0,9 comienza a ser *sospechoso* más que *recomendable*. No obstante, en la aplicación a los juicios de experto, este peligro no existe. No tiene sentido concebir a cada experto como una variación de una forma única, pues cada uno representa la aplicación de su experticia a la valoración del contenido emocional de un conjunto de palabras. En esta situación, cualquier variación al alza de la consistencia interna es recomendable. Si ocurriera que se observan variaciones muy evidentes o sensibles entre expertos, este resultado indicaría otro tipo de estructura interna en los juicios y aconsejaría la aplicación de otras estrategias, como podría ser un análisis factorial exploratorio.

Este proceso cuantitativo no elimina la posibilidad de un acercamiento más cualitativo. La observación de que algún experto lleva a cabo valoraciones distintas a la norma puede ser un indicador, no ya de problemas asociados a la preparación de este profesional para realizar la tarea, sino de algún aspecto que se le escapa al equipo de investigación y requiere un estudio más profundo asociado a quien se aleja de lo esperado.

Aplicación. Respeto al estudio sobre palabras de contenido referencial, el Alfa de Cronbach correspondiente a los 20 expertos, dispuestos como ítems de un cuestionario en la matriz traspuesta de juicios, suministró un valor de .867, una cuantía satisfactoria. Las correlaciones corregidas ítem-total (experto-total) se dispusieron en el recorrido .262 a

.667, sin que prescindir de algún experto implicara un aumento en el valor del estadístico. Este resultado refuerza la conclusión de que la selección de expertos fue acertada y que existe cierto comportamiento normativo en los juicios.

Etapa 3. Cribado de palabras

Con la matriz original de *expertos x palabras*, en la que se han eliminado los juicios de los expertos iniciales cuyas valoraciones se alejan de la norma, procedemos a poner en marcha un cribado de palabras. El cribado consiste en seleccionar las palabras que cumplan en mejor medida dos criterios: (1) mayor acuerdo entre expertos y (2) mayor capacidad de activación del constructo investigado. Cuando la respuesta de los expertos se adapta a una escala cuantitativa, los criterios de mayor acuerdo y de mayor activación se corresponden respectivamente con una desviación tipo (S) mínima y una media aritmética (d) máxima. Para combinar ambos criterios en un mismo recurso, se agradece que sean expresados en la misma escala. La propuesta PSP es utilizar sendos índices acotados en el intervalo abierto (0,1) y operar con ellos mediante una ponderación simple:

$$I_t = I_d w_d + I_s w_s \quad (1)$$

donde, I_t es el índice que expresa la combinación de los criterios de máxima activación y máximo acuerdo; I_d expresa la intensidad de activación (media acotada); I_s indica el grado de acuerdo (a partir de la desviación tipo acotada); y w_d y w_s son los pesos que permiten ponderar respectivamente la importancia de la intensidad y del acuerdo en el criterio conjunto I_t . En el desarrollo de estos elementos, ha de cumplirse que:

$$0 \leq I_t, I_d, I_s, w_d, w_s \leq 1 \quad \text{y} \quad w_d + w_s = 1$$

Para la acotación del promedio, donde el recorrido de la variable (los valores posibles de respuesta de los expertos) es (m, M), se procede de este modo:

$$m \leq d \leq M \Rightarrow 0 \leq d - m \leq M - m \Rightarrow 0 \leq \frac{d - m}{M - m} \leq 1 \Rightarrow I_d = \frac{d - m}{M - m} \quad (2)$$

En el recorrido (m, M), la máxima varianza se obtiene cuando los n datos se encuentran repartidos en los extremos, de tal forma que c datos coinciden con el valor M y, por tanto, $n - c$ con m (Manzano-Arrondo, 2002). Para acotar S en estas condiciones, se requiere expresar c y $n - c$ en función de la media, d :

$$d = \frac{cM + (n - c)m}{n} \Rightarrow c = n \frac{d - m}{M - m} \quad (3)$$

$$y \quad n - c = n - n \frac{d - m}{M - m} = n \left(\frac{M - d}{M - m} \right) \quad (4)$$

Aplicando ambos resultados en el cálculo de la varianza:

$$\begin{aligned} S^2 &= \frac{\sum (x_i - \bar{X})^2}{n} = \frac{c(M - d)^2 + (n - c)(d - m)^2}{n} = \frac{\frac{n(d - m)(M - d)^2}{M - m} + \frac{n(M - d)(d - m)^2}{M - m}}{n} = \\ &= \frac{(d - m)(M - d)^2}{M - m} + \frac{(M - d)(d - m)^2}{M - m} = (d - m)(M - d) \quad (5) \end{aligned}$$

La expresión (5) se refiere a la máxima varianza, mientras que nos preocupa la mínima desviación tipo, lo que implica:

$$I_s = 1 - \frac{S}{\sqrt{(d - m)(M - d)}} \quad (6)$$

La expresión (6) no tiene solución cuando el valor d coincide con cualquiera de los extremos, situación en la que no solo se anula el denominador sino también el numerador, puesto que $S = 0$. Las expresiones (2) y (6) se combinan en (1) para obtener el índice acotado, I_t . No obstante, para facilitar su exposición, en lugar de los valores mínimo (m), máximo (M) y promedio (d), la expresión final puede simplificarse utilizando $a = d - m$ y $b = M - d$, es decir, respectivamente las distancias de la media a los extremos mínimo y máximo. En tal caso:

$$I_t = \frac{aw_d}{a + b} - \frac{Sw_s}{\sqrt{ab}} + w_s$$

Para la decisión $w_s = w_d = .5$, se tiene que:

$$I_t = \frac{1}{2} \left[\frac{a}{a + b} - \frac{S}{\sqrt{ab}} + 1 \right]$$

Aplicación. Para la elaboración del listado de contenido referencial, se decidió ponderar con el mismo peso los criterios de acuerdo inter-jueces y de valoración sobre la activación de este constructo emocional ($w_s = w_d = .5$). Con las puntuaciones de los 20 jueces y siguiendo el criterio ya especificado de seleccionar el 50% de las mejores unidades en cada fase, se recurrió a la mediana (palabra en la posición 83 del listado inicial de 164 palabras) con el valor $I_t = .64$.

Selección mediante muestra equivalente

Las dos fases anteriores (construcción de la lista inicial y cribado de expertos), permiten contar con una lista de palabras, basada en la experticia del equipo y de los jueces expertos. Este resultado puede ser suficiente como propuesta final: una lista de estímulos específicos de una emoción, que se propone para un amplio espectro de objetivos prácticos. No obstante, la vocación de PSP es responder a situaciones menos generalistas, caracterizadas por: (1) contar con unas condiciones de investigación o aplicación concretas, y (2) orientarse a un perfil determinado de participantes. Pongamos por caso que se requiere una colección de palabras que activen emocionalmente ante la existencia de claustrofobia, con el objetivo de servir como base para una prueba diagnóstica. La prueba ha de discriminar entre quienes padecen claustrofobia y quienes no, y ha de conseguirlo mediante un procedimiento concreto de medida.

Con esta pretensión específica, PSP establece una tercera fase o estudio con muestra equivalente, que se define por:

1. Selección de participantes, con presencia/ausencia de la característica central (claustrofobia, PR...) que justifica la respuesta emocional ante las palabras-estímulo. La identificación de los participantes como representantes de cada perfil, surge de criterios clínicos o bien de la aplicación de alguna prueba específica, como ocurre con un test sobre la característica central.
2. Aplicación de la prueba, es decir, presentación de la lista de palabras seleccionada en la fase anterior, junto con otras, supuestamente neutras. Las palabras de contenido neutro (no activadoras de emociones) cumplen dos funciones. Por un lado, se comportan como control, es decir, como grupo de datos de comparación o contraste respecto a los datos que provienen de las palabras activadoras. Por otro lado, ejercen de colchón o amortiguación, rebajando el nivel de estimulación que supondría enfrentarse únicamente a estímulos que suscitan emociones concretas. La disposición interna de ambos tipos de palabras ha de ser aleatoria en esencia o apariencia, es decir, ha de evitarse cualquier tendencia a *prever* la llegada de un estímulo activador.
3. Selección final de las palabras que formarán parte de la lista propuesta específica. Para ello, se aplican las pruebas estadísticas que resulten pertinentes en cada caso, seleccionando: (1) las palabras que realmente se comportan como neutras, es decir, cuya exposición mantiene la respuesta emocional en un nivel basal, y (2) las palabras de contenido emocional que permiten discriminar entre los perfiles de participantes,

es decir, cuya presentación genera sensiblemente una respuesta de emoción más intensa en el perfil que presenta la característica central.

Al finalizar esta tercera fase, se ha conseguido una lista de palabras que responde a: (1) la experticia del equipo de investigación, (2) el juicio de un grupo de expertos con una opinión normativa, y (3) los resultados de una muestra equivalente que garantiza la capacidad discriminante de la lista.

Aplicación. Acerca de la elaboración del listado de contenido referencial en concreto, las 83 palabras valorados por los expertos y ordenadas alfabéticamente fueron presentadas a un número de 38 estudiantes de los cuales 28 mujeres (73.68%) y 10 hombres (26.32%), con edad entre 20 y 48 años ($M=22.50$ y $DT=.871$, de entre ellos 28 tenían una puntuación relativamente alta en pensamiento referencial (cuartil superior = suma igual o superior a 6, al que nombramos grupo referencial) según la Escala REF de Lenzenweger, et al., (1997) y un grupo compuesto por 10 alumnos con las puntuaciones relativamente más bajas en REF (cuartil inferior = suma igual o inferior a 1, nombrado grupo control). Se aplicó el mismo recurso del Alfa de Cronbach correspondiente a los 38 estudiantes y se obtuvo un valor de .861.

La tabla muestra los estudiantes que se comportan de modo menos acorde con el conjunto, de tal modo que al excluirlas del análisis, mejoraría el valor del Alfa de Cronbach (finalmente el valor para los 32 elementos se quedó en .898).

Al eliminar	r*8	c*30	r11	r6	c31	r24
Alfa es	.874	.884	.890	.893	.894	.898

Nota: r* - sujetos perteneciendo al grupo referencial

c* - sujetos perteneciendo al grupo de control

Con las puntuaciones de los 32 estudiantes restantes se llevó a cabo el cálculo del índice I_r , distinguiendo el grupo referencial del control. Se seleccionaron el conjunto de palabras cuya diferencia entre los grupos referencial y control no fuera inferior a -0.1 (una décima del valor de I_r a favor del grupo control). Como resultado, se eliminaron 22 palabras.

A continuación, se fusionaron los archivos de los estudiantes y expertos, contando como listado el total de las 61 palabras que quedaban desde los estudiantes.

Estadísticos descriptivos

	N	Mínimo	Máximo	Media	Desv. tip.
<i>I_r</i> Estudiantes	61	.46	.76	.6442	.07322
<i>I_r</i> Expertos	61	.65	.77	.7029	.03307

Se observa que ocurre lo esperado: como las palabras que se han seleccionado para presentar a los estudiantes son las de mayor puntuación por parte de los expertos, estos puntúan por término medio más que los estudiantes (.70 frente a .64), aunque la dispersión sea menor (siendo ambas muy tolerables).

Con ello se calcula la diferencia absoluta y la media entre los índices I_r de ambas procedencias. El criterio de selección consiste en aplicar el mismo criterio de la décima aplicado en el caso de la comparación entre grupo control y referencial. Como resultado, se eliminan 20 palabras y las 41 restantes palabras de contenido referencial vienen recogidas en el Anexo.

CONCLUSIONES Y DISCUSIÓN

La investigación en Psicología requiere en numerosas ocasiones contar con listas de palabras potencialmente activadoras de emociones específicas y otros procesos a estudiar. La revisión de la literatura sobre el tema muestra que no existe un procedimiento consensuado para generar estas listas. Si bien cada situación de estudio puede aconsejar técnicas concretas, utilizar un procedimiento general suficientemente compartido añadiría más garantías metodológicas a estas investigaciones. Este documento presenta una propuesta procedimental concreta, con el acrónimo PSP. Esquemáticamente, el procedimiento implica seguir varias etapas de manera secuencial:

1. Construcción de un listado de palabras inicial, a partir de la literatura sobre el tema, de cuestionarios específicos, o del conocimiento del equipo de investigación.
2. Cribado de expertos, que permita un primer análisis acerca de las palabras seleccionadas.
3. Cribado de palabras. A partir del cálculo del índice I_r para cada palabra y selección de las unidades a partir de los valores I_r más elevados.
4. Muestra equivalente. Obtención de dos muestras de participantes, según su valor en la variable de contenido emocional sobre la que versa la activación de las palabras.

Si bien el trabajo se ha realizado en todo momento utilizando palabras, PSP es susceptible de ser aplicado a otras unidades de activación emocional, como pueden ser sonidos (véase, por ejemplo, Bradley, & Lang, 1999b) o imágenes (véase, por ejemplo, Lang, 1980). En ningún momento del proceso es indispensable que los elementos de la lista se encuentren impresos, sino dispuestos de algún modo que pueda ser revisado por expertos y presentado finalmente a una muestra de participantes.

PSP no se propone a la comunidad científica y profesional con el objetivo de estimular una estandarización forzada. Cada situación concreta exigirá a su vez respuestas específicas. La motivación de PSP es servir de guía metodológica, de tal forma que su aplicación genere confianza respecto a los resultados obtenidos. Por esto mismo, los autores han procurado no establecer una lista excesivamente larga. Puede considerarse que una exigencia metodológica que complique el proceso, reduce drásticamente las posibilidades de aplicación real. Así, por ejemplo, los programas de análisis de datos al uso en las ciencias del comportamiento contemplan el cálculo del Alfa de Cronbach de una manera sencilla. Del mismo modo, el índice I_t tiene una implementación fácil en una hoja de cálculo e incluso en el software estadístico programable. De este modo, PSP puede ser asumido como un procedimiento que provee confianza en los resultados sin que ello implique un nivel poco realista en las operaciones metodológicas.

A la vez el trabajo recalca en proporcionar una lista de palabras de contenido referencial útil en el estudio del pensamiento referencial. Del mismo modo, puede complementar a las existentes sobre palabras depresivas y de tipo persecutorio, lo que permitiría en las futuras investigaciones discriminar los procesos que tienen que ver con los delirios persecutorios versus referenciales.

ABSTRACT

Selection procedure for referential content words. A methodological proposal. In Psychology research, it is common to use lists of words that have specific emotional content. The literature on this subject shows disparity of views and practices to make lists and use them. This work proposes a specific procedure designed to facilitate the creation of sets of emotional words. To this end we have made a review of the status of the matter and we designed a selection process divided in three phases: (1) drawing up an initial list which follows the criteria of specificity, variable activation level, control concepts, formal diversity and relevant number; (2) experts ratings of initial list, final selection, quantitative data collection and analysis using I_t index; and (3) final selection of items using the assessment

made by a sample of participants. The procedure is illustrated synthetically and step by step, through an empirical study on referential thinking.

REFERENCIAS

- Altarriba, J., & Bauer, L. M. (2004). The distinctiveness of emotion: A comparison between emotion, abstract, and concrete words. *American Journal of Psychology*, *117*, 389-410.
- Ato, M., Benavente, A., & López, J. J. (2006). Análisis comparativo de tres enfoques para evaluar el acuerdo entre observadores. *Psicothema*, *18*(3), 638-645.
- Blanch, M. T., & Baños, R. M., (1996). Estímulos verbales y trastornos emocionales: Un estudio sobre palabras con contenido emocional. *Revista de Psicopatología y Psicología Clínica*, *1*, 137-157.
- Bradley, M. M., & Lang, P. J. (1999a). *Affective norms for English words (ANEW): Instruction manual and affective ratings*. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.
- Bradley, M. M., & Lang, P. J. (1999b). *International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings*. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.
- Brouillet, T., & Syssau, A. (2005). Connection between the evaluation of positive or negative valence and verbal responses to a lexical decision making task. *Canadian Journal of Experimental Psychology*, *59*, 255-261. doi: 10.1037/h008805.
- Carroll, N. C., & Young, A. W. (2005). Priming of emotion recognition. *Quarterly Journal of Experimental Psychology*, *58A*, 1173-1197. doi: 10.1080/02724980443000539
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.
- Derry, P. A., & Kuiper, N. A. (1981). Schematic processing and self-reference in clinical depression. *Journal of Abnormal Psychology*, *90*, 286-297. doi: 10.1037/0021-843X.90.4.286
- Escobar-Pérez, J., & Cuervo-Martínez, Á. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medicina*, *6*, 27-36.
- Foa, E. B., McNally, R. J., & Murdock, T. B. (1989). Anxious mood and memory. *Behaviour Research and Therapy*, *27*, 141-147.
- Greenberg, M. S., Vazquez, C. V., & Alloy, L. B. (1988). Depression versus anxiety: Differences in self and other schemata. In L. B. Alloy (Ed.), *Cognitive processes in depression*, 109-142. New York: Guilford Press.
- Hamann, S., & Mao, H. (2002). Positive and negative emotional verbal stimuli elicit activity in the left amygdala. *NeuroReport*, *13*, 15-19.
- Ingram, R. E., Lumry, A, Cruel, D., & Sieber, W. (1987). Attentional processes in depression disorders. *Cognitive Therapy and Research*, *11*, 351-360.
- Jiménez, F., Vázquez, C., & Hernan Gómez, L. (1998). Adjetivos en castellano de contenido depresivo autorreferente y de contenido neutral: Normas de emocionalidad y frecuencia subjetiva de uso. *Revista de Psicopatología y Psicología Clínica*, *3*, 199-215.
- Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory and Cognition*, *31*(8), 1169-1180.

- Kousta, S. T., Vinson, D., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, *112*, 473-481. doi: 10.1016/j.cognition.2009.06.007.
- Landeta, J. (2002). *El método Delphi. Una técnica de previsión del futuro*. Barcelona: Ariel.
- Lang, P. J. (1980). Behavioural treatment and bio-behavioural assessment: Computer applications. In J.B. Sidowski, J.H. Johnson y T.A. Williams (Eds.), *Technology in mental health and delivery systems*, 119-137, Norwood, New Jersey: Ablex.
- Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, *40*, 776-785.
- Ledesma, R., Molina, G., & Valero, P. (2002). Análisis de consistencia interna mediante alfa de Cronbach: un programa basado en gráficos dinámicos. *Psico-USF*, *7*(2), 143-152. doi:10.1590/S1413-82712002000200003.
- Lenzenweger, M. F., Bennett, M. E., & Lilienfeld, L. R. (1997). The referential thinking scale as a measure of schizotypy: Scale development and initial construct validation. *Psychological Assessment*, *9*(4), 452-463. doi: 10.1037/1040-3590.9.4.452.
- Lewis, P. A., Critchley, H. D., Smith, H. D., & Dolan, R. J. (2005). Brain mechanisms for mood congruent memory facilitation. *NeuroImage*, *25*, 1214-1223. doi: 10.1016/j.neuroimage.2004.11.053.
- Lodge, M., & Taber, C. S. (2005). The automaticity of affect for political leaders, groups, and issues: An experimental test of the hot cognition hypothesis. *Political Psychology*, *26*, 455. doi: 10.1111/j.1467-9221.2005.00426.x.
- Manzano-Arrondo, V. (2002). Variación y posición de variables: estrategias para su estudio exploratorio. *Metodología de Encuestas*, *4*(1), 45-61.
- McNally, R. J., Amir, N., Louro, C. E., Lukach, B. M., Riemann, B. C., & Calamari, J. E. (1994). Cognitive processing of idiographic information in panic disorder. *Behaviour Research and Therapy*, *32*, 119-122.
- Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. Cambridge MA, US, The MIT Press.
- Meyer E. C., & Lenzenweger, M. F. (2009). The specificity of referential thinking: A comparison of schizotypy and social anxiety. *Psychiatry Research*, *165*, 78-87. doi: 10.1016/j.psychres.2007.10.015.
- Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., & Reuter-Lorenz, P. A. (2005). Emotional category data on images from the International Affective Picture System. *Behavior Research Methods*, *37*, 626-630.
- Neijens, P., Ridder, J. A., & Saris, W. E. (1992). An instrument for collecting informed opinions. *Quality & Quantity*, *26*, 245-258. doi: 10.1007/BF00172428.
- Nieto-Moreno, M., Hervás, G., & Vázquez, C. (2006). Palabras de contenido paranoide y depresivo en castellano: Especificidad, frecuencia de uso y grado de emocionalidad. *Revista de Psicopatología y Psicología Clínica*, *11*, 165-178.
- Raine, A. (1991). The SPQ: A Scale for the Assessment of Schizotypal Personality Based on DSM-III-R Criteria. *Schizophrenia Bulletin*, *17*(4), 555-564. doi: 10.1093/schbul/17.4.555.
- Real Academia Española (2001). *Diccionario de la lengua española (DRAE)*. 22ª Edición.
- Redondo, J., Fraga, I., Comesaña, M., & Perea, M. (2007). The Spanish adaptation of ANEW (Affective Norms for English Words). *Behavior Research Methods*, *39*(3), 600-605.

- Redondo, J., Fraga, I., Comesaña, M., & Perea, M. (2005). Estudio normativo del valor afectivo de 478 palabras españolas. *Psicológica*, 26, 317-326. doi: 10.3758/s13428-015-0572-5 · 2.93.
- Rodríguez-Testal, J. F., Valdés-Díaz, M., Benítez-Hernández, M. M., Fuentes-Márquez, S., Fernández-Jiménez, E., & Senín-Calderón, M. C. (2009). Stability and reliability of the assessment of referential thinking by the REF scale. *World Psychiatry*, 8(1), 297-297.
- Scott, G. G., O'Donnell, P. J., & Sereno, S. C. (2012). Emotion words affect eye fixations during reading. *Journal of Experimental Psychology, Learning Memory and Cognition*, 38(3), 783-792. doi: 10.1037/a0027209.
- Sebastián, N., Martí, M. A., Carreiras, M., & Cuetos, F. (2000). *LEXESP. Léxico informatizado del español*. Barcelona: Edicions Universitat de Barcelona.
- Senín-Calderón, M. C., Rodríguez-Testal, J. F., Fernández-Jiménez, E., Benítez-Hernández, M. M., Fuentes-Márquez, S., & Valdés-Díaz, M. (2010). Usefulness of likert format in assessment of referential thinking. *European Psychiatry*, 25(1), 1551. doi: 10.1016/S0924-9338(10)71534-0.
- Senín-Calderón, M. C., Rodríguez-Testal, J. F., & Perona-Garcelán, S. (2014). *El pensamiento referencial: aspectos psicopatológicos y del desarrollo*. CreateSpace Independent Publishing Platform, 156. ISBN 978-1500487942.
- Skjong, R., & Wentworth, B. H. (2001). Expert Judgement and Risk Perception. In R. Skjong, & B.H. Wentworth (Eds.), *Offshore and Polar Engineering Conference* (pp. 537-544). Stavanger: ISOPE.
- Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48, 813-838.
- Van der Salm, C. A., Van Knippenberg, D., & Daamen, D. D. (1997). A critical test of the choice questionnaire for collecting informed public opinion. *Quality & Quantity*, 31, 193-917.
- Van Kampen, D. (2006). The Schizotypy Syndrome Questionnaire (SSQ): Psychometrics, validation and norms. *Schizophrenia Research* 84, 305-322. doi: 10.1016/j.schres.2006.02.011.
- Van Strien, J. W., & Morpurgo, M. (1992). Opposite hemispheric activations as a result of emotionally threatening and non-threatening words. *Neuropsychologia*, 30, 845-848.
- Verona, E., Sprague, J., & Sadeh, N. (2012). Inhibitory control and negative emotional processing in psychopathy and antisocial personality disorder. *Journal of abnormal psychology* 121 (2), 498-510. doi: 10.1037/a0025308.
- Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behaviour Research Methods*, 45, 1191-1207. doi: 10.3758/s13428-012-0314-x.
- Weinstein, B. D. (1993). What is an expert? *Theoretical Medicine*, 14(1), 57-73.
- Wing J. K., Cooper J. E., & Sartorius N. (1974). *The description classification of psychiatric symptoms. An instruction manual for the PSE and CATEGO system*. London: Cambridge University Press.
- Wordreference (2005). *Diccionario de sinónimos y antónimos*. Espasa Calpe.
- Yik, M. S. M., Russell, J. A., & Barrett, L. F. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology*, 77, 600-619.

ANEXO

Lista de palabras de contenido referencial

Nº	Palabra	Índice It	Nº	Palabra	Índice It
1	murmurar	0,75	22	burlando	0,70
2	aludido	0,74	23	alusión	0,69
3	risita	0,74	24	sonrisilla	0,69
4	cuchicheo	0,74	25	hipócrita	0,68
5	criticando	0,73	26	observar	0,68
6	rumor	0,73	27	comentarios	0,68
7	burlas	0,73	28	vergüenza	0,68
8	chismorreos	0,73	29	calumniar	0,67
9	miradita	0,72	30	cotilla	0,67
10	malpensado	0,72	31	sarcástico	0,67
11	insinuar	0,71	32	irónico	0,66
12	ridículo	0,71	33	liante	0,66
13	insinuación	0,71	34	disimular	0,66
14	secretos	0,71	35	mirón	0,66
15	vigilado	0,71	36	entrometido	0,65
16	crítica	0,70	37	mentiroso	0,65
17	crítico	0,70	38	metomentodo	0,65
18	susurrar	0,70	39	hiriente	0,63
19	chismoso	0,70	40	burlón	0,62
20	sonrisita	0,70	41	fingido	0,60
21	habladurías	0,70			

(Manuscrito recibido: 1 Septiembre 2015; aceptado: 27 Enero 2016)