Capítulo 5

Specialized dictionaries and corpus linguistics in the translation of computer terminology

JORDI PIQUE-ANGORDANS
Universitat de València
SANTIAGO POSTEGUILLO
Universitat Jaume I de Castelló

1. INTRODUCTION

Those of us involved in the wide world of applied linguistics are convinced more than ever that and interdisciplinary approach must be and is intimately connected to our field of work. As Bygate and Kramsch (2000: 2) point out, "applied linguistics, in becoming more interdisciplinary, is better prepared for the principled handling of a range of real-world issues".

Our aim here is to write about the experience of "producing" a bilingual specialized dictionary, which also means to speak about translation. It is obvious that the production of a dictionary entails to analyze and understand, in the first place, all the elements that are involved in it. This is why Widdowson (2000: 4) says that before we can talk about applied linguistics, we must first of all know what linguistics is exactly before we can proceed to apply it. Similarly, to speak about dictionary production cannot be understood without speaking about translation since the former cannot be understood without the latter; and by extension, we cannot speak about building a specialized dictionary without understanding what a language for specific purposes (LSP) really is.

The term translation can be understood as the product –the text that has been translated– or the process –the act of producing the translation. This process between two written languages implies that the translator change an original written text (or source text [ST]) through the original verbal language (source language [SL]) into another written text (or target text [TT]) through a different verbal language (target language [TL]). This process corresponds to one of the three translation categories, as described

by Jakobson (1959/2000. 114) in his seminal paper "On linguistic aspect of translation", or "interlingual translation"; the other two being "intralingual translation", which occurs when we paraphrase a text in one and the same language, and "intersemiotic translation", or transmutation, in which verbal signals are interpreted through non-verbal signs. This interlingual translation is described, particularly in its semantic distinction between text, properly speaking, and language, in figure 1.

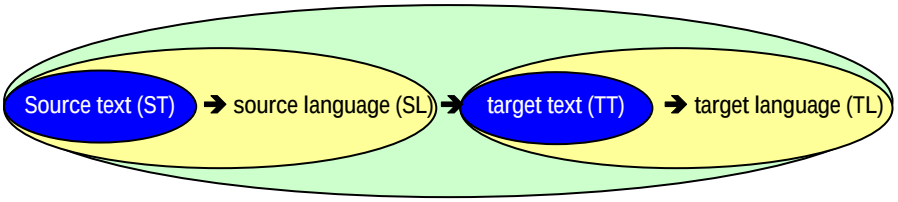

Figure 1. Interlingual translation: source language/source text vs. target language/text.

The source text (ST) is confronted with the target text (TT), each one within its own linguistic system, with its own rules and conventions and, in addition, their respective cultures. In other words, in the translation act we confront not only the source language with the target language, but also the source text with the target text.

Another element involved in translation is what Malinowski (1935) called "context of situation". This author, who has been quoted on many occasions, makes special reference to the concept of context in the exercise of translation. In the second volume of the *Coral Gardens and their Magic* (1935) he included an introduction where he analyzes speech acts in their relationship to the specific context in which language is used. He also points to the fact that language includes a magic element, which is encountered by the translator when translating literature, political oratory or publicity, and a pragmatic element of language when translating scientific and technical text. In other words, language must be fully embedded with context for a successful translation. As he says, "[t]he real linguistic fact is the full utterance within its context of situation" (Malinowski, 1935: 11).

In the same line, Darwish (1989) studies and understands translation as a double reflex mechanism. Translation should be understood as two sets of linguistic and cultural repertoires, each one with a sub-block of integrated components. Upon initiating the translation, he adds, the two parallel repertoires move to coincide and replace a series of situational, lexical, grammatical, stylistic, phonologic and, why not, cultural equivalents while, at the same time, these universal concepts are granted language

characteristics. According to Darwish, the text/discourse attributes travel from one repertoire to the other through what he calls the concept "lens", as illustrated in figure 2.

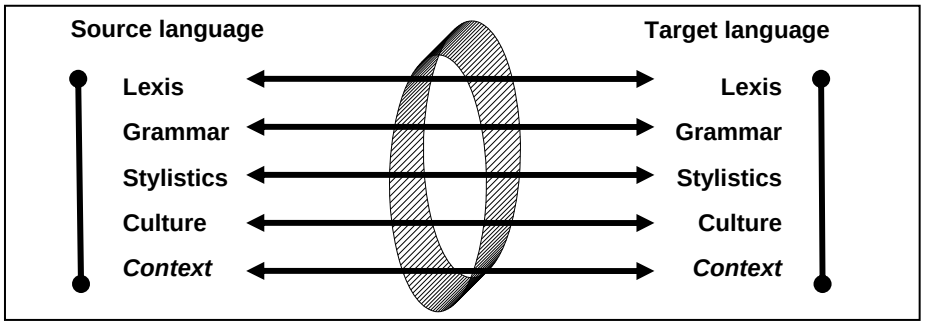

Figure 2. The concept "lens", adapted from Darwish (1989).

Through his diagram, Darwish tries to bring to the fore that translation is a two-way operation, even though translation may apparently occur in only one direction; it is also a part of a cumulative process in which a constant reflex movement is always present. However, to fully understand Darwish's diagram as a full explanation of the translation process the concept of context has been added. As pointed out by Posteguillo (2003: 95), Malinowski's words "seem to have been forgotten until pragmatists and specialised lexicographers brought them back to life in the last decades", particularly Alcaraz (1990, 1993).

2. SPECIALIZED DICTIONARIES

When we refer to a text, we can understand the adjective "specialized" from two points of view, as described by Cabré (1993: 135): (a) text as referring to "specialization through topic", or using Sager et al.'s (1980) label, as "subject specialized language"; and (b) text as referring to "specialization through «special» characteristics where the exchange of information takes place". Of the two approaches, we will focus on the first of the two, that is, texts thematically specialized.

2.1. SPECIALIZED LANGUAGE VS. GENERAL LANGUAGE

According to Cabré (1993: 127 ff.), language is made up of a set of subcodes which speakers use according to their dialectal modalities. General language, on the one hand, or "la langue tout entière", in Kocourek's (1982)

words, can be understood as a set of interrelated sets, their common link being the common language. Each one of these subsets can be one specialized language (Cabré, 1993: 129). Speakers then select one of these types according their needs of expression and the communicative context they are in. Also the language is conditioned by a set of morphologic, lexical, syntactic, phonologic, semantic and discursive rules that are common to all speakers of that language. The specialized language, on the other hand, is composed of a set of subcodes, partially coincidental with the common language, which is characterized by certain special peculiarities (Cabré, 1993: 128-129).

The amount of jargon which is used in the description of what a specialized language really is makes us think –as already many authors have already done so– that trying to define specialized languages is rather complicated, if not impossible, particularly in strict linguistic terms, precisely because what distinguishes it from general language are its pragmatic or extra-linguistic characteristics.

Authors do not seem to agree as to where to place one (general language) and the other (specialized language). For some, they are two separate entities; for others, one is inside the other; while others think that a sublanguage of speciality is half in and half out. See the three approaches as described and summarized by Ahmad et al. (1985) in figure 3.

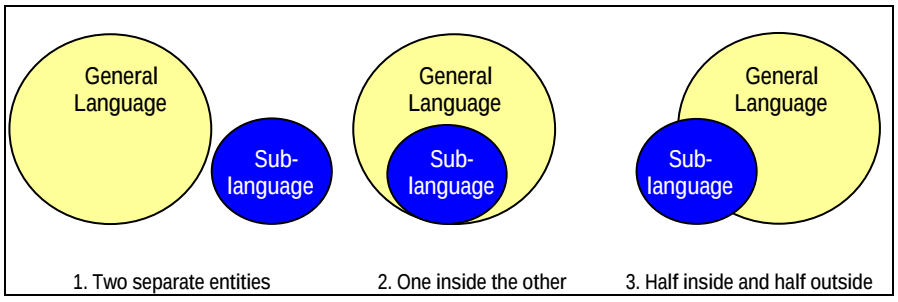

Figure 3. Models of relationship between general language and specialized languages.

Sager et al. (1980) think that one of the fundamental characteristics of specialized languages is that they are used more self-consciously; it is in their level of use where we will find specific criteria of differences with the general language. Cabré (1993: 136), however, thinks that besides this level of use, the communicative context must also be taken into consideration, both from the speaker and the listener points of view. Specialized languages, therefore, have to be understood in terms of inclusion in reference to the

general language, and in terms of intersection, as far as the common language is concerned, with which both share characteristics and maintain a constant exchange of units and conventions (Cabré, 1993: 140).

Taking the third model of relationship (figure 3) as our point of departure, Cabré (1993: 140) develops this intersection relationship of the specialized sub-language in reference to the common language, which acts precisely as the common link in all sets, as well as an inclusion relationship with the general language since both share specific characteristics, units and conventions. Figure 4 illustrates Cabré's framework.

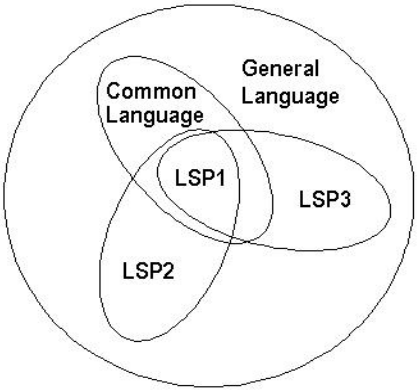

Figure 4. Interrelationship between global and general language, and specialized languages (Cabré 1993: 140).

This is where Cabré's (1993: 129) definition acquires full meaning when she talks of the general language as a "set of sets", that the common language is the "common link between all sets", and that each sub-set can become a "specialized language".

Taking this interrelationship between general language and specialized language, we can add a series of factors that Lehrberger (1986: 22) considers essential to define a specialized language, namely:

a) A specific topic
b) Lexical, semantic and syntactic restrictions
c) Abnormal grammatical rules
d) A higher frequency of specific constructions
e) Its own specific structure
f) A specific use of symbols

2.2. CULTURE AND SPECIALIZED DICTIONARIES

According to Bergenholtz and Tarp (1995: 60), we can only say that a specialized dictionary is culture bound when "historically and culturally the dictionary has developed its own specific features within delimited geographical areas" and, as an example, they mention the Spanish-Danish Law Dictionary, since the Spanish law code is based on the Roman law, while the Danish law code is based on the Germanic law. Nevertheless, they add, bilingual dictionaries of science and technology, particularly computer dictionaries, offer problems which are completely different. According to Bergenholtz and Tarp (1995: 70), these dictionaries "comprise an extremely broad spectrum of subject fields and topics with widely differing user requirements for linguistic and encyclopedic information. The only common denominator in this host of possibilities is that the subject fields involved are invariably independent of culture".

Although we basically agree with this opinion, it is also true that some cultural notes will invariably appear in a dictionary. A practical example – although not everybody would agree with their inclusion– is Collins et al.'s (2004) *Spanish Computing Dictionary* in which we have included two quotes which we understand to be totally culture bound, both taken from the Spanish newspaper supplement *Ciberp@ís*, the first one in reference to the dictionary entry "teledemocracia":

> Callús (Barcelona), Carreño (Asturias) y Jun (Granada) son los tres vértices del triángulo tecnológico. Jun (gobernado por el PSOE) fue el primer municipio de la Unión Europea en declarar el acceso a Internet como un derecho de sus ciudadanos. También apostaron por la **teledemocracia**: sus vecinos no sólo votan en los plenos municipales por la red sino que son los que más han rellenado el censo en Internet.

Under "nanociencia", the following quotation was also included (where CSIC stands for Consejo Superior de Investigaciones Científicas, a Spanish institute of scientific research):

> El CSIC trabaja tanto en **nanociencias** (conocimiento de las propiedades de los materiales a escala nanométrica) como en nanotecnología (la aplicación práctica).

2.3. PASSIVE VS. ACTIVE DICTIONARIES

Although Bergenholtz and Tarp (1995) do no speak directly of the concept of specialized bilingual active and passive dictionaries, they indeed

do indirectly in their chapter on linguistic information that should be included in this type of dictionaries. These authors fully develop the typical contents that a good specialized bilingual active dictionary should contain. Nevertheless, we should further delimit this concept following Salerno's (1999) footsteps. According to her, a good active dictionary must fulfill a two-fold service: it must help both to code and decode a text, and it is precisely this coding function that makes a specialized bilingual dictionary an indispensable tool for the translator, as figure 5 shows.

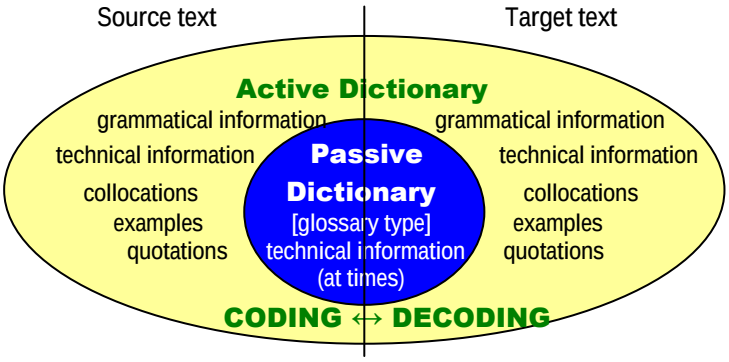

Figure 5. Active bilingual dictionary versus passive in a specialized context.

A more practical attitude is taken by Gak (1992: 335) when he says that an active dictionary should be more than just a simple dictionary: "il doit être une sorte d'encyclopédie de la lange dans laquelle on traduit". Fuertes and Velasco (2001: 37), in their evaluation of bilingual dictionaries of economics, add the following to their concept of an active dictionary:

> The construction of an active (Spanish-English) dictionary has to be based on the 'translation principle'. This means primarily two things: giving a real translation and taking into consideration phraseology and meaning in context.

If we may say so, in our dictionary we have tried to do precisely this, and as Fuertes (2005: 154) comments in his review article of our dictionary, it is adequate for computer students, translators and professionals as well. In addition, he underscores its usefulness for the translation of user's manuals, so common nowadays.

2.4. ARE SPECIALIZED DICTIONARIES NECESSARY?

Technological advances, along with the speed with which these are produced particularly in the last thirty years, have brought about a black hole as to what refers to available and good bilingual computer dictionaries. New concepts and terms daily invade us and dictionary updating, or the appearance of new ones, is absolutely necessary, or it is rather slow when this occurs. Without them, the translation products of our professionals are often based on improvisation and something rather crafty. What is even worse, is the harmful effects on different languages in the world, since most advances are derived from an English-speaking environment. Hens and Vella (1992: 362), in the 8th AESLA Conference, already pointed to the undesirable effects for our language which is gradually suffering from absolute pure import of terms and structures (calques, neologisms, *extranjerismos*, etc.) which in the end becomes a serious communication obstacle. Elsewhere we made a comparison among some of the main bilingual computer dictionaries in the Spanish market which, although incomplete, gives an idea, both in quantity of entries as well as in the information provided (Piqué-Angordans et al., 2006) and reproduced in table 1.

	Online glossaries	Printed glossaries	Oxford, 1983	McGraw-Hill, 1993	Paraninfo, 1995	Prentice-Hall, 1999	Ariel, 2001	Collin et al., 2004
Approximate no. of entries	<1,000	<1,000	±4,000	±4,000	±2,500	±4,000	±6,200	±35,000
Technical information	√ [*]	√ [*]	√	√	[*]	√	√	√
Gramatical information					√			√
Collocations					√		√	√
Examples					√			√
Internet terminology	√ [*]	√ [*]				√	√	√
Quotations					√			√
Complete Spanish Section					√			√

[*] Information not always included

Table 1. Spanish Computer Dictionaries available: Comparison (Piqué-Angordans et al., 2006).

Pérez Hernández (2002), in her discussion on the relationship between lexicography and terminology, and specialized lexicography as a bridge

discipline between the two, poses the following situation when faced with the hypothetical need of asking oneself as to what dictionary should a user consult in case information on the word "leukaemia" may be warranted. We would propose a similar situation with an apparently simpler word, such as "bomb", but in a computer text. What we are really implying here is the dichotomy between technical words, which are monosemic in meaning, and semi-technical words, that is, polysemic in nature, which entail a more complicated area to both students and professional translators. The translation is simple, "bomba", but the translation is not sure of its real meaning in the computer world and there must be some extra meaning attached to it than simply a destruction tool used by terrorists. The 22nd edition of the *DRAE* (2001), like most general monolingual dictionaries, is of no help to the translator. It presents over 30 different meanings and collocations, but none of them refers to computer science. An encyclopedic dictionary, Larouse type, is of no help either. One must resort to a specialized dictionary to obtain the correct information, either monolingual in the field of informatics, or a truly active specialized bilingual dictionary to obtain adequate information that will satisfy the need demanded by the translation. Bilingual online glossaries simply translate the word and seldom provide additional information for users.

Users need additional information to be satisfied. This is what was meant when we were talking about a black hole in dictionary help for information technology. Online monolingual glossaries will indeed give you sufficient information and that alone will sufficient for the translator's understanding of what "bomb" in a specialized computer context stands for, but that is not necessarily the case with online bilingual glossaries. A fully developed specialized bilingual dictionary should quickly solve this information gap.

More difficulty would constitute a word that has undergone changes over time. Such is the case of the term "hacker". The numerous translations offered for this term range from "computer enthusiast" (*entusiasta de los ordenadores*) and "computer expert" (*experto en informática*) to "informatics pirate" (*pirata informático*). Internet glossaries tend to give this term a more friendly definition, as is the case of the Internet glossary "Glossarist" (http://www.glossarist.com/glossaries/science/) we extracted the following definition of *hacker*: "Slang term for a technically sophisticated computer user who enjoys exploring computer systems and programs, sometimes to the point of obsession". The *Diccionario Oxford de Informática* (1983: 213) defines it as "retocador excesivo del programa" and "pulidor excesivo del programa", which is a rather farfetched and unreal definition. It is interesting to learn, however, that hackers themselves prefer

to be called "crackers". Nevertheless, the term has recently acquired –one could say that almost by popular demand– a pejorative meaning and most bilingual dictionaries will prefer "pirata o intruso informático".

Two very common anglicisms, which have already been accepted by the *DRAE* (2001), are the English terms *software* and *hardware*. Attempts have been made by such bilingual dictionaries as Collins and Larouse at introducing expressions such as "quincallería" for *hardware*, but were not very successful. Also by different style books by major periodicals and news agencies (e.g. *El País*, Agencia EFE) also tried to introduce the expressions "soporte físico" and "soporte lógico" for hardware and software, respectively. Although still present in some current bilingual dictionaries (e.g., KLETT by Cambridge University Press), writers and particularly informatics practitioners, however, do not believe such a proposal to be a practical solution and these attempts seem to have failed.

In the recently published *Diccionario Panhispánico de Dudas* (2005), other less fortunate proposals have been introduced in which the only consideration has been towards the pronunciation of the term in question and they have been transcribed simply by how they sound; see, for example, terms like "deuvedé" for *DVD*, "zum" for *zoom*, "baipás" for *bypass*, or "pirsin" for *piercing*. As Emilio Lorenzo said, in his prologue to Aguado's (Paraninfo, 1996: ix) commented dictionary, there are still many computer science terms that are still waiting to be granted "lexical entity".

Another area we were concerned about was to provide our dictionary was to increase collocational information, one of the main reflections of language in use, particularly in specialized contexts. As Firth observed, "[y]ou shall know a word by the company it keeps" (Palmer, 1968: 179). These two or more words within a short space of each other in a text, as Sinclair and Carter (1991) would define collocations, seldom occur in general language; in specialized languages, however, their frequency increases considerably, as has been shown in several studies (Frantzi, 2003: 219-220).

3. CORPUS LINGUITICS AND THE *SPANISH DICTIONARY OF COMPUTING*: A PRACTICAL CASE

We are not being original if we say that the expression corpus linguistics has arrived through Anglo Saxon linguistics; however, the first steps taken in this area were taken by an Italian researcher, Father Roberto

Busa of the Università La Sapienza, with a number of Latin texts from the works of Thomas Aquinas (Marello, 2004: 351); using the old punched cards, in 1951 he elaborated and published his *Index Thomisticus* in an IBM machine. More recently, in 1972 the Croatian Louis Milić, did something similar with half million words extracted from St. Augustine's prose. Milić, however, does not appear to be very happy with his findings since in 1991 he said that he was rather disappointed, particularly with what had been achieved with informatics in the area of the humanities (Aarseth, 1997).

When we speak of corpus linguistics in modern linguistics we tend to imply different things, such as text machine readable, sampling, representativity, open or closed corpus, etc. We tend to see it rather as a series of methods, procedures and resources that deal with empirical data in linguistics. Indeed, the corpus in the elaboration of our dictionary has simply been just another tool, although necessary, in our work.

3.1. CRITERIA FOR THE COMPILATION OF A CORPUS

As suggested by Pearson (1998: 43), the fact that so many different definitions exist for the definition of corpus is the reason why the linguistic community has not decided for one or the other; however, there is consensus in saying that it is an artifact, elected or assembled in electronic form, according to some explicit criteria. In addition, it is a series of language pieces that have not been edited and that occur naturally. The corpus, therefore, can be used because it is a "language sample", as Sinclair (1991) would say, or because it is "representative of a given language", as pointed out by Francis (1993).

Obviously, most authors coincide in saying that the more generous a corpus is, more believable it will be. Maximum representativity is what we are looking for, although this is not always possible due to several reasons. Nevertheless, we must undertake the construction of a corpus from the point of view of the aim for which we are constructing it: it can be a corpus with a purely lexical aim, or a corpus for a genre study, or a corpus for a study on a purely grammatical aspect.

According to Sinclair (1991: 17), a corpus "is a collection of material which is broadly homogeneous, but which is gathered from a variety of sources so that the individuality of a source is obscured unless the researcher isolates a particular text". However, given the complexity of our corpus, perhaps we should call it a "special purpose corpus" (Pearson, 1998: 48),

given the fact that it is composed of diverse sub-corpora, as we will see below, although much smaller than the one Pearson (1998: 56) believes a "special purpose corpus" should be.

The size of a corpus depends very much on the type of research being carried out. Biber (1993: 256) says that its size perhaps is not the main consideration for a corpus to be representative; its adequacy depends rather on the applicability for which it was elaborated. Pearson (1998: 51) confirms this idea in the following manner:

> if one wishes to carry out linguistic studies on a subset of the language, size may be less important but it will still be important for the corpus to be representative of the subset in question and, consequently, the larger it is, the more representative it is likely to be.

In the case of specialized corpora, as is our case, the volume must be relatively limited in order to be operative, and this was done with a monolingual specialized dictionary as our starting point, which became our base vocabulary, to which numerous new terms were then added.

In regard to written texts, the recommendation is that it should not be centered on only one author; it is recommended a diversity of authors and published texts. From a technical viewpoint, texts must be as uniform as possible in regard to a specific discipline, and this is exactly what we did with computer science in general. This uniformity should also be maintained as far as text type. Texts can be informative, didactic or normative, and to mix these three types would simply enhance the lack of uniformity in a corpus; however, our aim was not uniformity of texts from this point of view, but rather to obtain an ample variety of commonly used terms in one discipline, computer science, although this discipline appear in different media, for instance, written texts and hypertexts.

3.2. OUR CORPUS: DATA COLLECTION INSTRUMENTS

(a) The translation team

Our team was made up of four people, two in Valencia, Spain –who carried out the preparation of the corpus, search of new terms, and the translation properly speaking, aided by a computer professional who helped us solve difficult areas– and two in London –a professor from the University

of Surrey-Roehampton and a representative of the firm Peter Collin Publishing. They carried out the revision of the translation and the edition.

(b) Term selection criteria used

Lexicographers tend to be very precise when establishing selection criteria in regard the preliminary steps towards the elaboration of a specialized bilingual dictionary. This is the case with Bergenholtz and Tarp (1995: 103) who would advocate the following criteria:

(a) LSP terms only,
(b) both LSP terms and other non-common language expressions,
(c) all expressions which must be assumed to occur regularly,
(d) expressions having a certain minimum frequency in a given LSP corpus.

Alcaraz and Hugues, in the prologue of their *Diccionario de términos económicos, financieros y comerciales* (4ª rev. ed., 2004: XVII-XVIII), describe these criteria in three main terms: pertinence, clarity and economy. They explain them as follows:

(i) **Criterion of pertinence or relevance** – specialized terms of real use in the areas of computer science and the Internet; this criterion is based on three types of words: firstly, words which are exclusive or almost exclusive of the discipline, such as *buffer*, *CPU*, etc.; secondly, words that have a number of meanings of the general language and others from specialized language; and thirdly, words from the ordinary lexis which have a higher rate of presence in the special text, such as *download*, *upload*, etc.

(ii) **Criterion of clarity** – according to Alcaraz and Hughes, this criterion is fundamental; it helps both the translator and the student with numerous examples, collocations, phrases, etc.

(iii) **Criterion of economy** – not all words from different disciplines can be included in a dictionary, unless they are widely used in that discipline. Among these terms we should include the adj+noun compounds, or nouns with adjectival or attributive function. It would be endless if we were to include all of them in a dictionary since it has obvious limits, besides those imposed by editorials.

(c) Our corpus

For our dictionary, we were provided with a vocabulary base which derived from the 3rd edition of the *Dictionary of Computing* (Collin, 1998), published by Peter Collin Publishing, a firm dedicated almost exclusively to the edition of dictionary, particularly specialized dictionaries, and which at first was to also publish ours. This fact alone was sufficient guarantee that we were going to work with people with experience in that endeavor. At the same time, this made our task easier in terms of having an already published set of terms as base corpus to which we would incorporate new terms from different sources. This made our final list of terms a sufficiently diversified corpus, so much so that we can talk about a four-fold corpus which included the following:

Sub-corpus 1: Vocabulary base (Collin, 1998), complemented with additional terms selected from other dictionaries (Collin, 1996, 1997, 2000).

Sub-corpus 2: Texts extracted from the Internet.

Sub-corpus 3: Reference bilingual texts from which examples and collocations were also extracted for the Spanish-English section of the dictionary from two well-known computer science textbook (Hamacher, 1996; Tanenbaum, 1997) which were first scanner and then placed in parallel in a Word document; also text about informatics technology from the European Union, downloaded from the Internet (http://europa.eu.int/index_es.htm), also placed in a parallel Word document.

Sub-corpus 4: Reference texts for quotations from other publications (Cebrián, 2000; Castells, 2001; Millán, 2001; Yus, 2001, among others), including texts from popular publications in Spanish, namely the weekly supplement *Ciberp@ís* of *El País*, and Spanish editions of popular journals, such as *Computer Hoy*, *PC Plus*, and the like, published during 2001 and 2002.

Not much can be added to the composition of Sub-corpus 1, since it was a decision previously taken when the project was initiated; however, it was also agreed to complement this initial list with additional terms, given the constant production of new terms in computer science. Sub-corpus 2, however, implied a production of randomly selected chunks of text dealing with computer science and downloaded from the Internet. This search produced a whole plethora of new terms that were then compared to sub-corpus 1 and a decision was taken as to which ones were to be added. It is indeed a fact that the Internet, as pointed out by Wooldridge (2005), has a

great access speed and thus many linguistic phenomena can be tracked down. Its main advantage is precisely this dynamic nature of constant renewal of the language: "Tout comme la langue ell-même, il se renouvelle sans cesse et offre ainsi des instantanés de l'état actual de la langue, avec sa dose d'usages nouveaux, établis ou vieillissants" (Wooldridge, 2005: 208). However, there are other aspects related to the Internet, particularly because the terms we were deciding to add to the main list had to be tested against currently available specialized glossaries and dictionaries, as well as against other texts extracted from professional journals. This operation was a necessary step we had to take, particularly because of its lack of system the way in which terms are introduced in the Internet; in addition, not all texts can be taken as reliable enough, as pointed out by Barbadilla (2002); in addition, as Posteguillo (2003: 83) has rightly pointed out, in the Internet reigns a sort of a "terminological cyber-anarchy".

Sub-corpus 3 entailed the scanning of the two texts mentioned above, in their two versions English original and Spanish translation, and were placed in parallel format in a Word document. Its first function was to be used as reference corpus, through which we could search for different words, contextualize them, or ascertain meaning is use and double-check uncertain uses. A second operation with this sub-corpus was the selection of specific words to which we would select collocates and examples with the idea of supplying them in the Spanish-English section of the dictionary. This was done with the help of the *FileMaker Pro* program in which we elaborated a database for the bilingual information being drawn from the two texts (see Figures 7 and 8 below for two layout formats of the same record). Sub-corpus 4 was actually a complement to sub-corpus 3 because it served a similar purpose, also for the Spanish-English section of the dictionary, although the source material was all in Spanish to which we provided our own translation when necessary. Figure 6 illustrates a bilingual record.

This model record, which in addition had space available for quotations (CITAS) and commentaries (COMENTARIOS) for the Spanish-English side of the dictionary, was elaborated for the purpose of providing additional entry material in Spanish, particularly new examples in order to avoid repeating the same examples used in the first part (English-Spanish), and also quotations from Spanish texts listed in the corpus description. This provided the material necessary to complete the second part of the dictionary, once the doubling of the first part had been finished.

This record layout, with its different field possibilities, allowed us to complement the second part (Spanish-English) of the dictionary, once the screen dump of the first part (English-Spanish) was done.

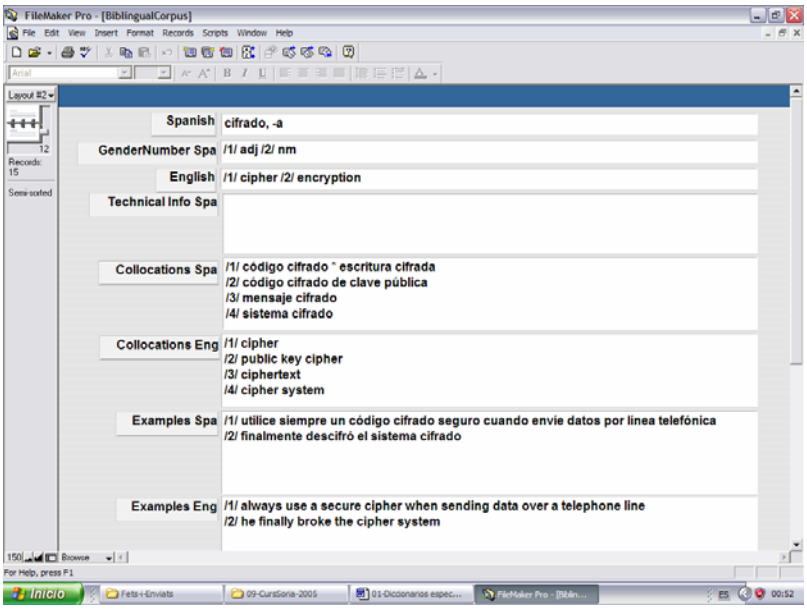

Figure 6. Layout of a bilingual record for the Spanish-English version.

4. ELABORATION PROCESS OF THE DICTIONARY

In the initial proposition to work on this dictionary we had very much in time available material in bookstores, while at the same time we made sure we had sufficient and necessary freedom to provide the dictionary with as much information as possible, although having in mind the obvious editorial needs and demands.

4.1. CONTENTS OF THE DICTIONARY ENTRIES

Our project was to provide, whenever necessary, each article of the dictionary with the information that a specialized bilingual dictionary should provide, as outlined earlier in this paper. Thus, the following the following set of information was considered:

Grammar category (adj [adjective], adv [adverb], n [noun], tv [transitive verb], iv [intransitive verb], etc.)
Gender (nm [noun masculine], nf [noun feminine])
Technical commentary (when necessary)
Translation (the asterisk [*] means more than one translation option)
Collocations
Acronyms

Examples
Quotations, comments, and notes

4.2. TRANSLATION PROCESS AND EDITION OF ENGLISH-SPANISH SECTION

Once the contents of each letter had been determined (vocabulary base plus possible additions previously selected), the translation process was begun by the Spanish team, providing each entry with the necessary information. Each letter was then sent to London for revision by other member of the team. The process can be graphically described as shown in figure 7.

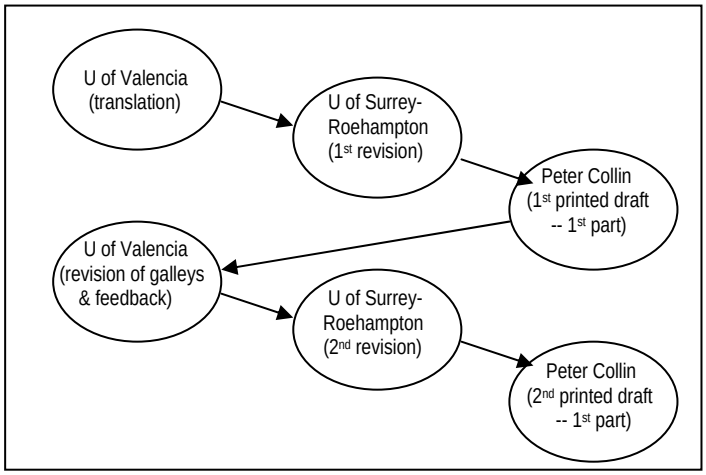

Figure 7. Translation process of first part (English-Spanish) of the dictionary.

When this initial process was finished it its sequential steps for each letter, as described in figure 7, a hard copy of the whole first part was issued and each member of the team revised it thoroughly for possible errors and their corrections, suggestions for new or better translations, and last-minute additions. This translation process was carried out in coded version of each letter. This coded version (which we have transcribed in order to safeguard copyright ownership) contained the following information, as exemplified through one of its entries in figure 8.

Once this entry and translation had been transmitted and revised, along with the rest of entries for each letter, Peter Collin Publishing produced a first hardcopy for final revision, both in London and in Valencia. Figure 9 shows an example of how this hardcopy looked like in print.

Code description	Dictionary information
entry	virus
grammatical category	n
technical information	(program which can corrupt data)
translation	virus
gender	m
example	if your PC is infected with a virus, your data is at risk =
translation	si su PC se infecta con un virus, los datos están en peligro
technical information	(removes a virus from a file)
collocation 1	anti-virus software =
translation	"software"
gender	m
alternative translation	* utilitario
gender	m
translation	antivirus
technical information	(checks files to see if they contain a virus)
collocation 2	virus detector =
translation	detector
gender	m
translation	de virus
commentary	COMMENT: viruses are spread by downloading unchecked files from a bulletin board system or via unregulated networks or by inserting an unchecked floppy disk into your PC - always use virus detector.

Figure 8. Coded version of entry in section English-Spanish of the dictionary.

virus *n* (program which can corrupt data) virus *m* **if your PC is infected with a virus, your data is at risk** = si su PC se infecta con un virus, los datos están en peligro (removes a virus from a file) **anti-virus software** = "software" *m* * utilitario *m* antivirus (checks files to see if they contain a virus) **virus detector** = detector *m* de virus

COMMENT: viruses are spread by downloading unchecked files from a bulletin board system or via unregulated networks or by inserting an unchecked floppy disk into your PC - always use virus detector.

Figure 9. Printed version of the entry "virus".

4.3. SCREEN DUMP OF THE SPANISH-ENGLISH SECTION

Once the English-Spanish version was finished from A to Z, later printed and revised again, the screen dump was carried out in London by means a specially designed software. This became the most troublesome operation of the whole process, because this software cannot distinguish adequately semantic fields so that adequate collocations would appear under their proper entries. Misplaced collocations and exercises had to be located and placed under their correct entries. This operation was carried out in parallel by the research groups, in London and Valencia. Figure 10 shows a coded version of an entry the Spanish-English section of the dictionary.

Code description	Dictionary information
entry	cifrado, -a
category 1	1
grammatical category 1	adj
alternative uses	código cifrado * escritura cifrada =
translation	cipher
collocation 1	clave cifrada =
translation	cipher key
collocation 2	código cifrado de clave pública =
translation	public key cipher
collocation 3	mensaje cifrado =
translation	ciphertext
collocation 4	sistema cifrado =
translation	cipher system
example 1	utilice siempre un código cifrado seguro cuando envíe datos por línea telefónica =
translation	always use a secure cipher when sending data over a telephone line
example 2	finalmente descifró el sistema cifrado =
translation	he finally broke the cipher system
category 2	2
grammatical category 2	nm
translation	encryption
quotation	Los sistemas de cifrado, tradicionalmente constituidos por una rotación de caracteres, requieren que tanto remitente como receptor conozcan o posean las reglas de rotación.
source	PC Plus

Figure 10. Coded version of entry in section English-Spanish of the dictionary.

The printed representation of this entry will look as shown in figure 11.

cifrado, -a 1 *adj* código cifrado * escritura cifrada = cipher clave cifrada = cipher key código cifrado de clave pública = public key cipher mensaje cifrado = ciphertext sistema cifrado = cipher system utilice siempre un código cifrado seguro cuando envíe datos por línea telefónica = always use a secure cipher when sending data over a telephone line finalmente descifró el sistema cifrado = he finally broke the cipher system 2 *nm* encryption

---

Los sistemas de cifrado, tradicionalmente constituidos por una rotación de caracteres, requieren que tanto remitente como receptor conozcan o posean las reglas de rotación.

PC Plus

Figure 11. Printed version of a Spanish-English entry.

One final aspect in this Spanish-English version was the provision of exercises and quotations for a selection of entries from this part of the dictionary. As mentioned above in the corpus description, in the elaboration of the database (see figure 6 for a sample *FileMaker Pro* record) we included some of the entries we felt needed extra information, whether in the form of more examples and including quotations from up-to-date publications. In the first section, English-Spanish, this extra information had already been provided with in the proposed term selection provided with by Peter Collin Publishing. To this list, new terms were added and also more examples in some of the entries. However, in the second part of the dictionary, Spanish-English, we provided both some new terms and examples, but especially quotations from Spanish current periodicals or recently published books. Figure 12 shows the quotations used, 118 of them, and their origin.

5. CONCLUSION

The long debated argument over the usefulness of corpus linguistics, as far as our research project was concerned, seemed to have suddenly come to an end. Fortunately, we were able to use our collected data from both scanned texts, as well as downloaded texts, in files that we could utilize as reference material. Without our diversified corpus, our job would have been tremendously arid, perhaps even sterile. As we already pointed out, we understood our corpus as a valuable tool for our translation work through

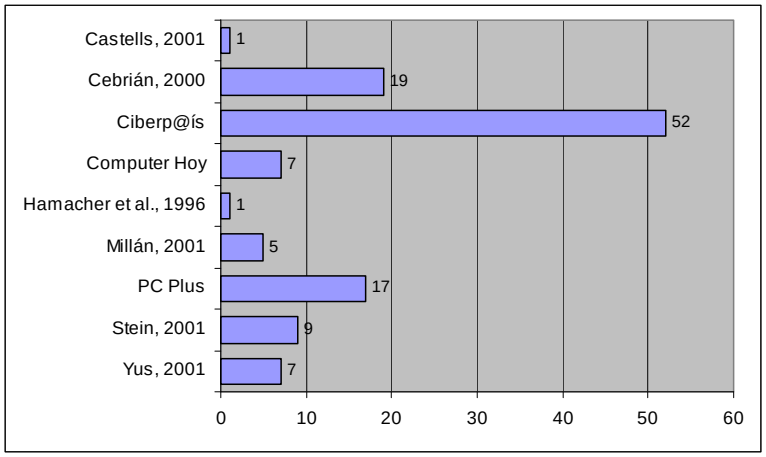

Figure 12. Number of quotations used in the Spanish-English section (n = 118).

which some of what Malinowski calls "untranslatable" words would have remained untranslated unless we paid attention to their use in context through several texts that our multi-corpus provided us with. Individual translation of words without contextualizing them would, therefore, constitute a futile effort in translation.

The major limitation of this type of research is, obviously, the limitations we imposed on our own accumulation of texts for our multi-corpus. As Pearson (1998) points out, the more generous a corpus is, the more believable it will be. However, we believe that the size of our multi-corpus was sufficiently generous to be helpful enough for our research and translation. We actually had to limit the number of entries and sub-entries, particularly in reference to what Kübler and Frérot (2003: 437) call "verb terms", such as *to telnet*, and also verb-preposition associations, such as *to compile* plus *against, as, from, in, in support, in support for, into, out of, with*, or *to configure* in connection with *as, for, in, in support for, on, to, under, with*, which would have made the dictionary endless. This actually calls for a more comprehensive investigation towards considering the incorporation of these and similar "verb terms" in subsequent and more complete editions of the dictionary.

Finally, we must emphasize that we have tried to produce a dictionary that may be categorized as a fully active specialized bilingual dictionary in which both grammatical and technical information play an important role in this type of dictionaries for both students and translation professionals, which was our main purpose to start with. We indeed feel that our job was accomplished successfully.

References [extract from the ***Bibliografía*** on pp. 256-281]

AARSETH, E. (1997). "The field of humanistic informatics and its relation to the humanities". *HumanIT* 4. Online journal: www.hb.se/bhs/ith/4-97/ea.htm [accessed November 10, 2005].

AHMAD, K, W. MARTIN, M. HÖLTER & M. ROGERS (1995). "Specialist Terms in General Language Dictionaries". University of Surrey Technical Report CS-95-14. Available online: www.mcs.surrey.ac.uk [accessed October 3, 2001].

ALCARAZ, E. (1990). *Tres paradigmas de la investigación lingüística*. Alcoy: Marfil.

—— (1993). "La lingüística y la metodología didáctica de las lenguas extranjeras" in V. García-Hoz (ed.), *Enseñanza y aprendizaje de las lenguas modernas*, 19-108. Madrid: RIALP.

—— & B. Hughes (2004): *Diccionario de términos económicos, financieros y comerciales (inglés-español, Spanish-English)*, 4th rev. ed. Barcelona: Ariel.

ARCHER, D., P. RAYSON, A. WILSON & T. MCENERY (eds.) (2003). *Proceedings of the Corpus Linguistics 2003 Conference. Lancaster University, 28-31 March 2003*. Lancaster: Lancaster University.

BARBADILLA, A. (2002). "La comunicación social de la ciencia e Internet". Available et http://bioinformatica.uab.es/divulgacio/css.htm [accessed 12/12/05].

BÉJOINT, H. & P. THOIRON (eds.) (1996). *Les dictionnaires bilingues*. Louvain-la-Neuve: Duculot. Universités Francophones.

BERGENHOLTZ, H. & S. TARP (Eds.) (1995). *Manual of Specialised Lexicography. The Preparations of Specialised Dictionaries*. Amsterdam & Philadelphia: John Benjamins.

BYGATE, M. & C. KRAMSCH (2000). "Editorial". *Applied Linguistics* 21,1: 1-2.

BIBER, D. (1993). "Representativeness in corpus design". *Literary and Linguistic Computing* 8: 243-257.

CABRÉ, M. T. (1993). *La terminología: teoría, metodología, aplicaciones*. Barcelona: Editorial Empúries.

CASTELLS, M. (2001). *La galaxia Internet*. Barcelona: Plaza & Janés.

CEBRIÁN, J. L. (2000). *La red: cómo cambiarán nuestras vidas los nuevos medios de comunicación*. Madrid: Taurus.

COLLIN, S. M. H. (1996). *Dictionary of Information Technology*, 2nd ed. Teddington: Peter Collin Publishing.

—— (1997). *Dictionary of Multimedia*, 2nd ed. Teddington: Peter Collin Publishing.

—— (1998). *Dictionary of Computing*, 3nd ed. Teddington: Peter Collin Publishing.

—— (2000). *Dictionary of Personal Computing and the Internet*, 3nd ed. Teddington: Peter Collin Publishing.

——, J. Piqué, S. Posteguillo & L. Melcion (2004). *Spanish Computing Dictionary Spanish-English/English-Spanish*. London: Bloomsbury.

DARWISH, A. (1989). "The translation process: A view of the mind". Available online: www.surf.net.au/writescope/translation/mindview.html [accessed January 9, 2003].

DRAE (2001). *Diccionario de la Real Academia Española*, 22nd ed. Madrid: Real Academia Española. Available online: www.rae.es [accessed October 2, 2005].

FRANCIS, G. (1993). "A corpus-driven approach to grammar", in M. BAKER, G. FRANCIS & E. TOGNINI-BONELLI (eds.), *Text and Technology*, 137-156. Amsterdam: John Benjamins.

FRANTZI, K. T. (2003). "Updating LSP dictionaries with collocational information" in ARCHER et al. (eds.), pp. 219-226.

FUERTES-OLIVERA, P. A. (2005). "Traditional methods in specialized lexicography". *Ibérica* 9: 149-157. Available online: www.aelfe.org.

—— & M. VELASCO-SACRISTÁN (2001). "A critical comparison of the macrostructure and microstructure of two bilingual English-Spanish dictionaries of Economics". *International Journal of Lexicography* 14,1: 31-55.

GAK, V. G. (1992). "De la langue système à la réalisation discursive dans un dictionnaire bilingue 'de type actif'", in H. TOMMOLA et al. (eds.), *Euralex '92 Proceedings*. (2 vols.) Tampere: Tampereen Yliopisto, vol. II, pp. 329-335.

HAMACHER, V. C., Z. G. VRANESIC & S. G. ZAKY (1996). *Computer Organization*, 4th ed. New York: McGraw-Hill (Spanish version: *Organización de computadoras*, translated by M. L. Fournier).

HENS CÓRDOBA, M. A. & M. VELLA RAMÍREZ (1992). "La traducción de textos informáticos: implicaciones para la didáctica de la traducción", in J. R. LOSADA, J. R. & M. MANSILLA [eds.], *Actas del VIII Congreso Nacional*

de Lingüística Aplicada. Vigo, 2, 3 y 4 de mayo de 1990, 361-370. Vigo: AESLA – Universidad de Vigo.

JAKOBSON, R. (1959). "On linguistic aspects of translation" in R. A. BROWER (ed.), *On Translation*, 232-239. Cambridge, MA: Harvard University Press.

KOCOUREK, R. (1982). *La langue française de la technique et de la science*. Wiesbaden: Brandstetter.

KROMANN, H.-P., RIIBER, T., & P. ROSBACH (1984). "«Active» and «Passive» bilingual dictionaries: the Scerba concept revisited", in R. R. K. HARTMANN (ed.) *LEXeter '83 Proceedings*. Tübingen: Max Niemeyer Verlag, 207-215.

—— (1991). "Principles of bilingual lexicography", in F. J. Hausmann et al. (eds) (1989-1991). *Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie. An International Encyclopedia of Lexicography. Encyclopédie Internationale de Lexicographie*. (3 vols.) Berlin and New York: De Gruyter, vol. III, pp. 2711-2728.

KÜBLER, N. & C. FRÉROT (2003). "Verbs in specialised corpora: from manual corpus-based description to automatic extraction in an English-French parallel corpus" in ARCHER et al. (eds.), pp. 429-438.

LEHRBERGER, J. (1986). "Sublanguage Analysis", in R. GRISHMAN & R. KITTREDGE (eds.), *Analyzing Language in Restricted Domains*, 19-38. Hillsdale, NJ: Lawrence Erlbaum Associates.

MALINOWSKI, B. (1935). *Coral Gardens and their Magic: A Study of the Methods of Tilling the Soil and of Agricultural Rites in the Trobriand Islands.* Two vols. London: George Allen and Unwin (recent facs. ed., 1991).

MARELLO, C. (2004). "Lexicography in Italy: specific themes and trends". *International Journal of Lexicography* 17,4: 349-356.

MILLÁN, J. A. (2001). *Internet y el español*. Madrid: Fundación Retevisión.

MUGDAN, J. (1989). "Grammar in dictionaries of languages for special purposes (LSP)". *Hermes* 3: 125-142.

NEWMARK, P. (1988). *A Textbook of Translation*. New York: Prentice Hall.

PALMER, F. R. (ed.) (1968). *Selected Papers of J. R. Firth*. Harlow: Longman.

PEARSON, J. (1998). *Terms in Context.* Amsterdam/Philadelphia: John Benjamins.

PÉREZ HERNÁNDEZ, M. C. (2002). "Explotación de los córpora textuales informatizados para la creación de bases de datos terminológicas basadas en

el conocimiento". *Estudios de Lingüística Española* (*ELiEs*) 18. Available at http://elies.rediris.es/elies18/ [accessed October 10, 2005].

PIQUÉ-ANGORDANS, J., S. POSTEGUILLO & L. MELCION (2006). "The development of a computer science dictionary, or how to help translating the untranslatable" in E. ARNÓ MACIÀ, A. SOLER CERVERA & C. RUEDA RAMOS (eds.), *Information Technology in Languages for Specific Purposes: Issues and Prospects*, 213-230. New York: Springer International.

POSTEGUILLO, S. (2003). *Netlinguistics. Language, An Analytical Framework to Discourse and Ideology in Internet*. Castelló de la Plana: Publicacions de la Universitat Jaime I.

RODRÍGUEZ GONZÁLEZ, F. (1999). "Anglicisms in contemporary Spanish. An overview." *Atlantis* 21: 103-139.

SAGER, J., D. DUNGWORTH & P. F. MCDONALD (1980). *English Special Languages: Principles and Practice in Science and Technology*. Wiesbaden: Oscar Brandstetter.

SALERNO, L. (1999). "Grammatical information in the bilingual dictionary. A study of five Italian-French dictionaries." *International Journal of Lexicography* 12,3: 209-222.

SINCLAIR, J. M. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

STEIN, G. (2001). *Éxito y fracaso en la nueva economía. Reglas para dirigir en la era de Internet*. Barcelona: Ediciones Gestión.

STUBBS, M. (2001). "Texts, corpora, and problems of interpretation: A response to Widdowson". *Applied Linguistics* 22,2: 149-172.

TANENBAUM, A. S. & A. S. WOODHULL (1997). *Operating Systems: Design and Implementation*, 2ª ed. Englewood Cliffs, NJ: Prentice-Hall International (Spanish versión: *Sistemas operativos: Diseño e implementación*, translated by F. GÓMEZ MOLINERO).

WIDDOWSON, H. G. (2000). "On the limitations of linguistics applied". *Applied Linguistics* 21,1: 3-25.

WOOLDRIDGE, R. (2005). "La web comme corpus d'usages linguistiques". *Cahiers de Lexicologie* 85,2: 209-225.

YUS, F. (2001). Ciberpragmática: El uso del lenguaje en Internet. Barcelona: Ariel.

[In: P. A. FUERTES OLIVERA (coord.) (2007). *Prolemas lingüísticos de la traducción especializada*. Valladolid: Universidad de Valladolid, pp. 87-107.]