

## EXPERT KNOWLEDGE REPRESENTATION IN BILINGUAL E-DICTIONARIES: A CASE STUDY

Maria Teresa Ortego Antón\*  
Universidade de Valladolid

Purificación Fernández Nistal\*\*  
Universidade de Valladolid

**Abstract:** In recent decades, the number of new concepts and terms has risen rapidly due to scientific and technological development. Additionally, expert knowledge, which used to be exclusively for experts, also interests middlebrow language users. Compilers of e-dictionaries, aware of this change, are gathering in new editions specialised terms that have become part of our daily lives. In the current globalised world, the need to transfer scientific knowledge to other languages arises, so one of the main tools that translators and, especially, translation trainees employ to look up

---

\* Professora Associada na Universidad de Valladolid, Espanha (Faculdade de Tradução e Interpretação Soria). Doutorou-se em Tradução e Comunicação (cum laude) em novembro de 2012 pela Universidade de Valladolid (UVa). Fez doutorado sobre o tratamento de vocabulário especializado coletado nos dicionários gerais bilíngues. Sua pesquisa concentra-se em lexicografia bilíngue, terminologia e linguística de corpus aplicados aos estudos de tradução. Valladolid, Espanha. E-mail: [tortego@lesp.uva.es.x](mailto:tortego@lesp.uva.es.x)

\*\* Professora Titular em Tradução e Interpretação do Departamento de Língua Espanhola (Área de Tradução e Interpretação) na Universidad de Valladolid. Suas publicações incluem artigos e livros sobre a interferência nos estudos contrastantes, tradução audiovisual, terminologia, tradução especializada, e ferramentas para auxiliar de tradução. Tem experiência de mais de 25 anos na tradução técnico-científica de Inglês para Espanhol. Valladolid, Espanha. E-mail: [purifer@itbyte.uva.es](mailto:purifer@itbyte.uva.es).



an unknown term are bilingual dictionaries. Hence, we consider that the study of the treatment given to computing terms in bilingual dictionaries is a field that needs to be reviewed. From an *ad hoc* corpus composed of texts from the main journals published in the UK and the USA, the most frequent terms belonging to computing are extracted using *TermoStat Web 3.0* (Drouin, 2003). Then, we verify how terms are gathered in the dictionary wordlist, if they are labelled or not, which translation equivalents are given and if they are followed by contextual data. In addition, we check the use of the given equivalents in two Spanish reference corpora: *Corpus del Español* and *Corpus de Referencia de I Español Actual*. The results from the analysis might suggest a need to take into account new proposals in order to implement the data gathered in these reference works as well as inform new procedures in the design and use of these tools from the point of view of translators as main users.

**Keywords:** Computing terms. Corpora. Expert knowledge. General bilingual dictionary. Translation equivalents.

## LA REPRESENTACIÓN DEL CONOCIMIENTO ESPECIALIZADO EN LOS DICCIONARIOS ELECTRÓNICOS BILINGÜES: UN ESTUDIO DE CASO

**Resumen:** En las últimas décadas el número de conceptos y, por ende, de denominaciones ha aumentado como consecuencia del desarrollo científico y tecnológico. Además, el conocimiento especializado, que era exclusivo de los expertos, ha pasado a interesar también a los usuarios de una lengua con un nivel cultural medio. Los compiladores de los diccionarios electrónicos, conscientes de este cambio, incorporan paulatinamente en las nuevas ediciones de sus obras el léxico especializado que han pasado a formar parte de nuestra vida cotidiana. Por otro lado, en el contexto globalizado actual surge la necesidad de transferir el conocimiento especializado a otras lenguas, de manera que una de las herramientas más utilizadas por los traductores y, especialmente, por los traductores en formación, para realizar consultas cuando se desconocen los equivalentes se corresponde con los diccionarios electrónicos bilingües. No obstante, consideramos que el estudio del tratamiento conferido al léxico de la informática en los diccionarios electrónicos bilingües es un campo que necesita ser estudiado con mayor profundidad. Por tanto, de un corpus *ad hoc* compuesto por textos procedentes de los principales diarios publicados en Reino Unido y Estados Unidos, extraemos las unidades léxicas más frecuentes del campo de la informática utilizando *TemoStat Web 3.0*. (Drouin, 2003).

A continuación, comprobamos cómo se recogen dichas unidades léxicas en la nomenclatura, si se marcan diatómicamente, qué equivalentes de traducción se ofrecen y si dichos equivalentes van acompañados de información contextual. Asimismo, verificamos el uso de los equivalentes propuestos en dos corpus de referencia del español: Corpus del Español y Corpus de Referencia del Español Actual. Las conclusiones nos permitirán vislumbrar nuevas propuestas que implenten la información recogida en estas obras de referenci así como nuevos procedimientos en el diseño y uso de estas herramientas desde la perspectiva de los traductores como principales usuarios.

**Palabras clave:** Léxico de la informática. Corpus. Conocimiento especializado.

## **1. The importance of specialised vocabulary nowadays**

In recent decades, the number of new concepts and terms has risen rapidly due to scientific and technological development. Additionally, expert knowledge, which used to be exclusively for experts, has become of interest to middlebrow language users as a result of the democratisation of education and media broadcasting. Compilers of e-dictionaries are aware of this change, so in new editions, they are gathering specialised terms that have become part of our daily lives.

Moreover, in the current globalised world, the need to transfer scientific knowledge to other languages arises, as it is produced or spread mainly in English. Within this framework, the transfer of specialised vocabulary is one of the obstacles that translators and, especially, translation trainees deal with. One of the main tools that they employ to look up an unknown term is the bilingual dictionary. Despite the fact that this type of dictionary is not the most suitable tool to search for specialised vocabulary because it often leads to mistakes when concepts are unknown, in previous research (Atkins & Varantola, 1998a, 1998b; Durán Muñoz, 2010; Bowker, 2012), these dictionaries were reported to have become one of the most widespread and frequently used tools among translators and interpreters.

In bilingual lexicography some reference works about different aspects that characterise specialised vocabulary gathered in English-French dictionaries were published. On the one hand, studies about the analysis of specialised vocabulary from a general approach were found. For instance, Jessen (1996) proved the presence of terms in three general English-French monolingual dictionaries and compared the results in four bilingual dictionaries. Other studies focused on a particular element, such as the research carried out by Boulanger (2001), who studied field labelling in general bilingual dictionaries. In addition, studies which emphasised a particular specialised domain also aroused the interest of researchers. Thoiron (1998) dealt with the place and role of medical vocabulary in general bilingual dictionaries, Roberts (2004) focused her interest on the study of wine terms, and Roberts and Josselin (2005) studied volcanology.

Taking into account the difficulties arising from scientific vocabulary transfer in interlingual communication as well as the importance of bilingual e-dictionaries as a search tool for users, we were surprised by the lack of studies in English-Spanish lexicography covering this area. Hence, given our experience in translator training and in the use of these tools, we consider that the representation of expert knowledge in English-Spanish e-dictionaries is a field that needs to be reviewed as long as it can offer resulting data that might improve the information gathered and aid translators and interpreters in their use of search procedures.

## **2. Setting the bounds to a domain: computing**

The study of specialised vocabulary gathered in the nomenclature of general bilingual e-dictionaries is an incredibly large task, and in general, it has been previously defined as difficult, if not impossible (Thoiron, 1998; Sierra Soriano, 2001; Rodríguez Reina, 2002). Consequently, our study is limited to a particular field of knowledge—that is, computing. On the one hand, major

advances in computing have changed the format of dictionaries as well as the lexicographical practice. In addition, this field has influence in every step involved in the compilation of a dictionary, such as in the organisation, structure, distribution of contents, reliability, accessibility, etc.

Furthermore, this field of knowledge is characterised by its great interdisciplinarity. Consequently, computing applications provide support to nearly all of the specialised domains in our society, so its changes have an impact on most fields of knowledge.

Thus, the aforementioned cross-sectional character will make the study of the treatment of computing terms gathered in general bilingual dictionaries capable of showing a global view of the treatment that compilers of these works give to expert knowledge in general. As a result, we consider that, once an analysis is made of a specialised domain which will be representative enough and cross-sectional to other specialised fields, the approach given to this field will tend to be reproduced in each of the fields covered in general bilingual dictionaries.

### **3. The limits of the field of study: English-Spanish bilingual e-dictionaries**

Today, we can find a wide range of English-Spanish e-dictionaries available on the market, so to be more exhaustive in our analysis, we need to set the bounds of our field of study. In order to carry out a comparison of the results in an effective way, the selected dictionaries must meet a number of scientific criteria. Accordingly, our study is limited to three dictionaries released by publishing companies that apply new lexicography techniques to their new editions. This proves that our selection of dictionaries comprises the most prestigious dictionaries in English-Spanish lexicography:

- *Collins Universal Español-Inglés, English-Spanish*; hereafter *CU*.

- *Gran Diccionario Oxford: Español-Inglés, Inglés-Español*; hereafter *GDO*.
- *Word Reference: English-Spanish*, hereafter *WR*.

As its designation shows, the electronic format is a feature which influences the data configuration—for example, in the way that the data are organised, in the inclusion of new data retrieval techniques, and in the addition of new information (Forget, 1999, 6). Some researchers (De Schryver, 2003; Sánchez Ramos, 2004, 177-211; Pastor & Alcina, 2010) have followed the trend started by Forget (1999) and have explored new features of electronic dictionaries. Owing to these features, e-dictionaries have become very useful tools for users.

Moreover, from the first approach to the selected dictionaries, they have met a set of requirements considered by experts (Roberts, 1997; Landau, 2001) as keys to support users in choosing a particular dictionary. Amongst the requirements, we would like to highlight the following properties: broad potential users, a college size, comprehensive coverage of the main regional English and Spanish, and a recognised publishing company.

In spite of the fact that lexicographers from the Aarhus School of Business (Denmark) claim to develop dictionaries taking into account different users' profile needs, publishers prefer to attach more weight to profitability. Consequently, when compiling a dictionary, lexicographers have in mind not only translators, but a broad variety of potential users, i.e. teachers, students or experts who need to transfer a communicative message from one language into another. Regarding size, a college dictionary includes between 35,000 and 70,000 entries and is probably used at higher levels in school and university undergraduate level. They are also used at home and at work by those whose duties have a considerable linguistic element (Svensen, 26).

Although it could be thought that English-Spanish dictionaries only cover British English, in order to satisfy the needs of a broad range of users, selected e-dictionaries include words from different

regional variants of English and Spanish. This last property guarantees that the input data is reliable, especially because lexicographers have based the compilations on preceding editions that have been revised and in which new terms extracted from corpora have been added.

For example, according to data gathered in the ‘Help’ section of the *CU*, lexicographers consult the *Bank of English* for the English language, which corresponds with the current version of the corpus built in the *COBUILD* project. When *CU* was compiled, the aforesaid corpus included about 550 million words of current English, from oral and written texts, which covered at least eight regional English varieties.

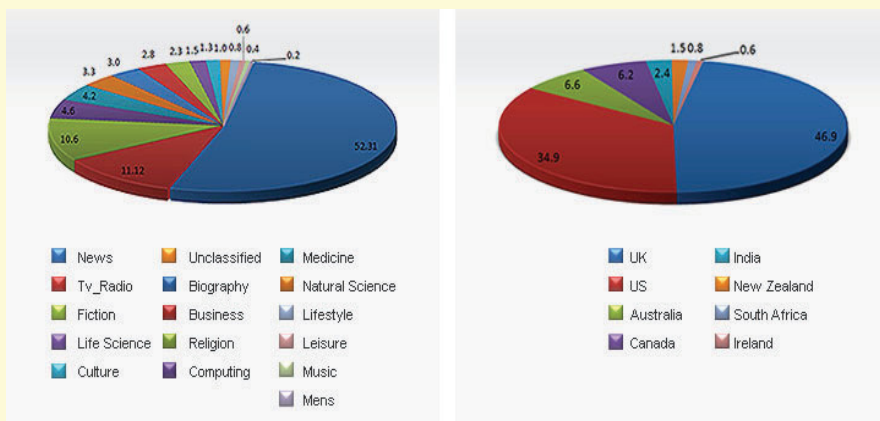


Figure 1. Distribution of domains and regional varieties in *Bank of English*.

For the Spanish language, *CU* compilers used the *Banco de Español*. From this corpus an example of the KWIC is shown in the ‘Help’ section, which is reproduced in Figure 2.



que jugarse la existencia. Que yo estoy	decidido	a sacar el asunto adelante y las
estadounidense, Bill Clinton, está	decidido	a seguir adelante con un proyecto de
que debe sufrir un artista cuando está	decidido	a seguir sin concesiones su verdadera
su segunda mujer. Les anunció que estaba	decidido	a suicidarse y todos los que le
de que me reconcilie, no lo haría. Estoy	decidido	a terminar con el matrimonio". El
crisis místicas, pero, sobre todo,	decidido	a convertirse en pintor; hizo progresos
que le reafirme en su posición de líder.	decidido	a luchar por el título, su trabajo ha
en esquí de fondo, un joven triatleta	decidido	a ser bombero o un empleado de artes
la categoría social de la dama.	decidido	a seguir el juego, el dueño del hotel
la oportuna aclimatación. Me levanté	decidido	a comenzar el suplicio de la escalada
asomó la cabeza por la ventanilla	decidido	a echar una mano y, cuando se quiso dar
en el caso, acudió ayer a la Audiencia	decidido	a declarar. Pero esta mañana (por
exploré el terreno diplomático,	decidido	a intentar por mi cuenta y riesgo lo
sesteando, pero el Zaragoza salió	decidido	a resolver el partido y supo aprovechar
de los encuestados afirma no estar	decidido	del todo. Conviene saber que la
pero él está ya prácticamente	decidido.	Se presentará como candidato
hablando dos o tres minutos y se dirigió	decidido	hacia mí. Cuando pasó al lado de Celia
refleja sentimentalmente quien cabalga	decidido	hacia el poder. Chirac tiene la lista
de asalto. Salió de su casa, se encaminó	decidido	y enfiló la mirada hacia la oficina de
me dirigi a la terminal. La crucé	decidido	y salí al exterior. Tomé un taxi. Pasé

Figure 2. Concordances of *decidido* in *Banco de Español*.

Regarding *GDO*, lexicographers rely on the *Oxford English Corpus* and *Oxford English Programme* for the English language.

The *Oxford English Corpus* is based mainly on material collected from pages on the World Wide Web (some printed texts, such as academic journals, have been used to supplement certain subject areas). It represents all types of English, from novels and specialist journals to everyday newspapers and magazines and from Hansards to the language of blogs, emails, and social media. It contains samples of language from all parts of the world, 80% from British English and American English, and the remaining 20%, about 400 million words, from other English varieties. In 2010 it included more than two billion words from 2000 onwards. The structure of the corpus is shown in Figure 3:



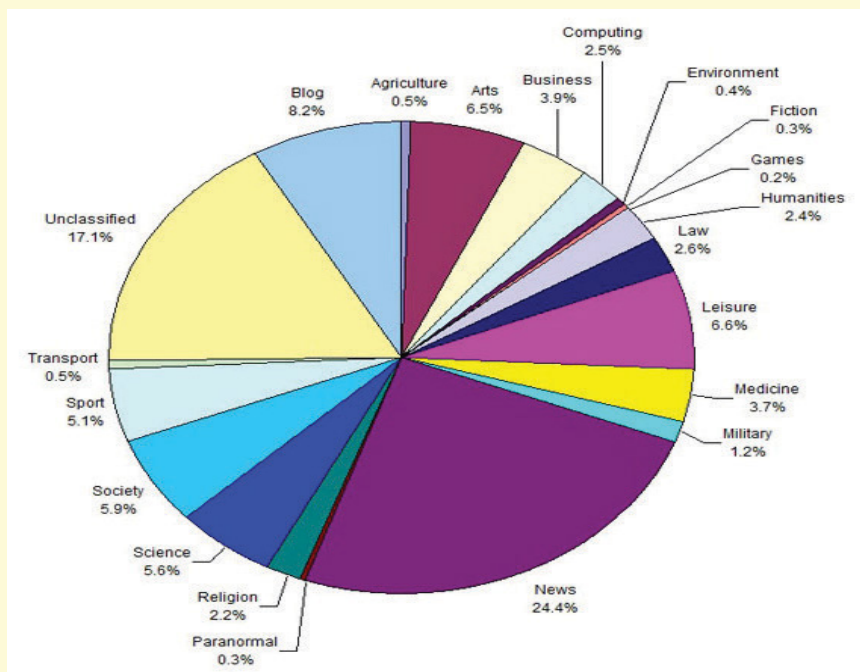


Figure 3. Structure of *OEC*.

The *Oxford English Programme*, whose origins date back to 1857, when the *Oxford English Dictionary* was launched, provides data about the use of words in English and about the emergence of new terms. This programme comprises a network of readers who submit quotes in which the use of a particular word is recorded, following a procedure similar to the compilation of the *Oxford English Dictionary*. Until the 1990s, quotes were stored on printed cards. Since then, they are included in a database called *Incomings*, which contains more than 62 million words, lots of them from readers, who send an average of 17,000 quotes per month. However, *GDO* lexicographers do not offer data about the use of corpora for compiling the Spanish section.

As for *WR*, in the 'About' section, no use of corpora is stated. However, the editors point out that entries have been supplemented with information that users have contributed on the forum. In

addition, the selected works rely on a long tradition that supports their inclusion as the sample of analysis in this study. In truth, the first edition of *CU* was printed in 1971, as stated by Hastings (206), and since then, nine editions have been published. Although the first edition of *GDO* was issued later, in 1994, since then, four other editions have been published. These figures prove the publisher's impact as well as the popularity of the aforementioned dictionaries in Spanish-speaking settings, which was, amongst other reasons, driven by the establishment of English as a foreign language in Spanish-speaking countries.

On the other hand, the boundaries of our sample are also based on the results of prior research, such as the study led by Corpas et al. (2001, 248), in which it is stated that our selection of dictionaries is the most frequently used by translation students: “*encuanto a los bilingües con el español, los alumnos de inglés (B) suelen tener el Oxford, el Collins o ambos*”<sup>1</sup>. This conclusion was reached after analysing the answers from questionnaires filled out by translation students from the University of Málaga (Spain). Moreover, Fernandez Quesada (149-162) carried out another study that came to the same conclusion. While studying typical translation mistakes in a sample of language students, she insisted that the students' more frequently used tools were “*el Gran Diccionario Oxford (2005) y el Collins Universal (2008)*”.

In addition, Muñoz Sánchez (2006) and Perrino (2009), amongst others, show that *WR* is one of the most frequently used dictionaries amongst professional translators.

#### 4. Methodology

First of all, we compiled an *ad hoc* corpus composed of texts from the main news publications published in the UK and in the USA between 2000 and 2008. These texts were available in electronic format and appeared on the Internet under the heading of ‘Technology’ or ‘Computing’.

In Table 1, the sources are shown, as well as the number of files and the number of tokens found in each text. These figures are the result of analysing the selected texts using *AntConc*, a freeware concordance program developed by Anthony (2012).

SOURCE	FILES	TOKENS
CNN (AmEn)	2	2942
New York Times (AmEn)	12	11505
TheGuardian (BrEn)	14	13262
TheTelegraph(BrEn)	4	2856

Table 1. *Ad hoc* corpus composition.

Then, texts were joined using a free piece of software, *Biz Bee File Merge*, in order to be processed as one file by *TermoStat Web 3.0*, a term extractor tool (Drouin, 2003) that uses a hybrid (i.e. statistical plus linguistic) method to identify candidate terms. It takes into account not only the structure of potential term candidates (using a part-of-speech tagger to identify nouns and adjectives and complex structures that contain these items) but also the relative frequencies of these potential candidates in the text being processed (called the analysis corpus) and a very large collection of newspaper articles (called the reference corpus). This method allows *TermoStat Web 3.0* to find not only multi-word but also single-word candidate terms in a single extraction process.

Once logged, we uploaded the file generated by *Biz Bee File Merge*, and we activated the option of simple terms and all the categories available (e.g. adjective, adverbs, nouns and verbs) and complex terms, and the analysis was launched. From the results provided by *TermoStat Web 3.0*, we excluded proper nouns, e.g., Facebook, and common nouns.

Candidat de regroupement	Fréquence	Score (Spécificité)	Variantes orthographiques	Matrice
pc	138	138.35	pc pcs	Nom
browser	55	110.21	browser browsers	Nom
internet	47	101.56	internet	Nom
com	49	99.83	com	Nom
laptop	37	86.22	laptop laptops	Nom
online	33	79.73	online	Adjectif
program	80	78.33	program programs	Nom
software	77	75.84	software	Nom
hard drive	26	75.68	hard drive hard drives	Adjectif Nom
user	84	69.44	user users	Nom
search engine	22	69.36	search engine search engines	Nom Nom
drive	91	66.44	drive drives	Nom
motherboard	19	64.22	motherboard	Nom
s	24	62.77	s	Adjectif
deskton	24	60.87	desktop	Nom

Figure 4. Results provided by *TermoStat Web 3.0*.

Hence, the ten terms which will constitute the sample of analysis are ‘PC’, ‘browser’, ‘Internet’, ‘laptop’, ‘online’, ‘program’, ‘software’, ‘hard drive’, ‘user’ and ‘search engine’. From a morphosyntactic standpoint, computing discourse, following a similar pattern to other domains, is characterised by the predominance of nouns, provided that the concepts are linguistically represented through nominalisations (Sager et al., 234).

Once the sample was limited, each term was looked up in each of the selected dictionaries, and the following criteria used in the previous research (Roberts, 2004; Josselin, 2005; Ortego Antón, 2012) were proved. First of all, we determined whether the term was included in the wordlist and, second, which procedure was used by the compilers—that is, if it was given the status of entry, nested entry or example in another entry. Then, we checked if the term was labelled and which tag was used for it, e.g., an abbreviation, the name of the domain or if several labels could coexist to refer to the computing field. Next, we extracted equivalents, whose uses were proved in two Spanish reference corpora. First,

we tested their use in *Corpus del Español* (Davies, 2002), hereafter *CdE*, limiting the search to the 20<sup>th</sup> century.

*CdE* is a 100 million word corpus of Spanish texts funded by the National Endowment for Humanities: 20 million words from the 1200s-1400s, 40 million from the 1500s-1700s, 40 million from the 1800s-1900s. The 20 million words from the 1900s are divided equally among literature, oral texts, newspapers and encyclopaedias.

Then, we validated whether equivalents were registered in *Corpus de Referencia del Español Actual*, hereafter *CREA*. *CREA* is a corpus of standard varieties of Spanish and currently contains 133 million words sampled from a wide range of written (90%) and spoken (10%) text categories produced in all Spanish-speaking countries between 1975 and 1999 (divided into 5-year periods). The texts in the corpus are distributed evenly between Spain and Latin America. The domains covered in the corpus include science and technology, social sciences, religion and thought, politics and economics, arts, leisure and ordinary life, health, and fiction. This corpus was designed as a monitor corpus which is continually updated so that it always represents the last 25 years of the history of Spanish. New data is added proportionally to maintain the corpus balance and to ensure that the various trends in current Spanish are represented. Texts for 2000-2004 are currently being incorporated (Sanchez, 2002; in McEnery et al., 2006).

Finally, we checked if the equivalents were accompanied by contextual data, such as examples, actants or the most common collocations.

## **5. Analysis and results**

### **5.1 Term inclusion**

The first element to be analysed is the inclusion of the sample of terms in the selected dictionaries. In the *CU*, one of the terms, ‘hard

drive’, is not included in the nomenclature, and ‘search engine’ is registered as a nested entry in the entry devoted to ‘search’. The other eight terms are included as entries. The percentage-based distribution is calculated as shown in Chart 1.



Chart 1. Term inclusion in the *CU*.

However, in the *GDO* and on *WR*, all of the terms are registered as entries in the nomenclature.

## 5.2 Field labelling

Referring to field labelling, in the *CU*, four of the terms (‘PC’, ‘Internet’, ‘laptop’ and ‘user’) were not labelled. Three of them (‘online’, ‘program’ and ‘software’) are labelled with the abbreviation of the domain in parentheses—that is, (Comput)—and two of them are labelled with the tag ‘Internet’ in parentheses. Hence, we differentiated between two different tags to label computing terms in this dictionary.

Concerning the *GDO*, only four of the terms, ‘browser’, ‘online’, ‘program’ and ‘search engine’, were labelled with the tag (Computing), which is the name of the domain between parentheses.

However, on *WR*, we observed that three terms were not labelled, which corresponded to 'PC', 'Internet' and 'laptop'. In addition, two of the terms, 'browser' and 'online', were labelled with '*informática*'. The rest of the terms were labelled using multiple tags, e.g., computer software, computer program, computers, etc.

The percentage-based distribution is calculated as shown in the following charts.

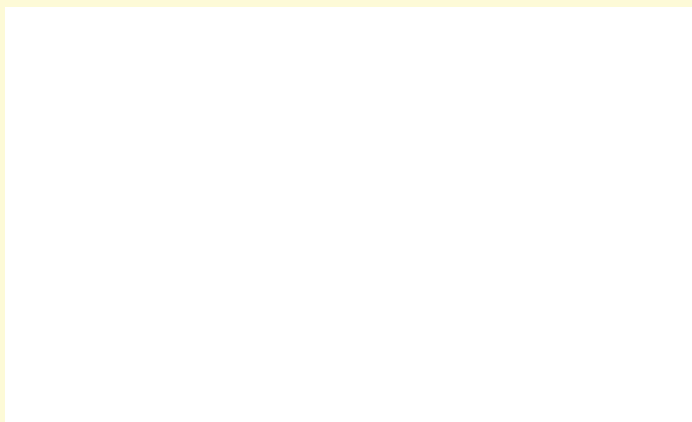


Chart 2. Field labelling in the *CU*.

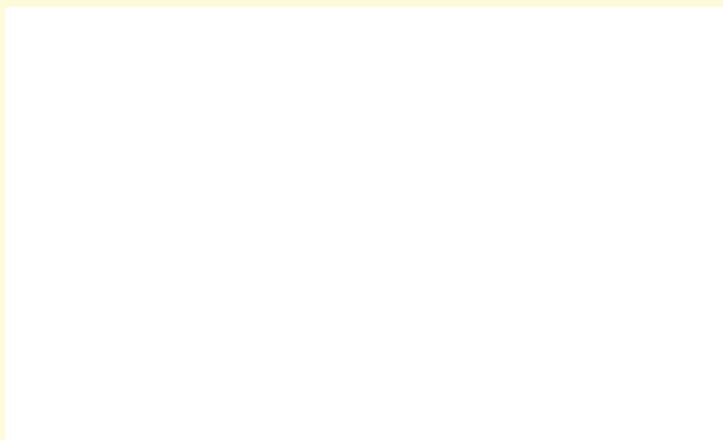


Chart 3. Field labelling in the *GDO*.



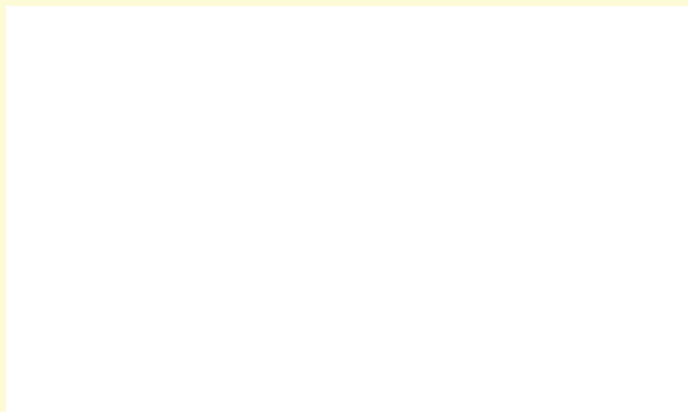


Chart 4. Field labelling on *WR*.

From the results shown in the preceding charts, *WR* is the dictionary that labels more terms, but at the same time, this work is characterised by the use of different tags.

### 5.3 Equivalents

Concerning equivalents, we found a great range of them, as the following table demonstrates, in which the equivalents gathered in the three dictionaries are highlighted in bold:

TERMS	EQUIVALENTS
PC	PC (CU, WR) OP (CU)
browser	navegador
Internet	<b>Internet</b> <b>la red (WR)</b>

laptop	computador/a portátil (CU) laptop (GDO, WR) ordenador (CU) portátil (WR)
online	<i>en línea</i> <i>conectado</i> (CU, GDO) <i>en internet</i> (WR) <i>on-line</i> (CU) <i>por internet</i> (WR)
program	<i>programa</i> <i>aplicación</i> (WR) <i>software</i> (WR)
software	<i>software</i> <i>aplicación</i> (WR) <i>programa</i> (WR)
Hard drive	disco duro (WR)
user	usuario
search engine	<i>buscador</i> <i>motor de búsqueda</i> (CU)

Table 2. Equivalents gathered in the selected dictionaries.

Regarding the use of the proposed equivalents, it must be pointed out that the use of almost all of them is registered in the *CdE* and in *CREA*. Only three of them, ‘*OP*’, ‘*on-line*’ and ‘*motor de búsqueda*’, gathered in the *CU*, are not registered in the *CdE*. Nevertheless, we must assess these data cautiously because some of the equivalents are registered only in a document, as can be observed in the following table:

TERMS	EQUIVALENTS	CdE	FREQ.	CREA	FREQ.
PC	<i>PC (CU, WR)</i>	X	4.03	X	15.45
	<i>OP (CU)</i>	-		X	5.74
browser	<i>navegador</i>	X	0.31	X	1.62
Internet	<i>Internet</i>	X	13.50	X	54.40
	<i>la red (WR)</i>	X	24.10	X	86.54
laptop	<i>computador/a portátil (CU)</i>	X	0.22	X	10.27
	<i>laptop (GDO, WR)</i>	X	0.35	X	0.20
	<i>ordenador (CU)</i>	X	17.61	X	24.81
	<i>portátil (WR)</i>	X	2.85	X	4.43
online	<i>en línea</i>	X	5.13	X	11.05
	<i>conectado (CU, GDO)</i>	X	4.43	X	4.73
	<i>en internet (WR)</i>	X	2.19	X	1.12
	<i>on-line (CU)</i>	-		X	3.08
	<i>por internet (WR)</i>	X	1.05	X	0.56
program	<i>programa</i>	X	142.40	X	220.37
	<i>aplicación (WR)</i>	X	40.66	X	84.86
	<i>software (WR)</i>	X	5.26	X	14.04
software	<i>software</i>	X	5.26	X	14.04
	<i>aplicación (WR)</i>	X	40.66	X	84.86
	<i>programa (WR)</i>	X	142.40	X	220.37
harddrive	<i>disco duro (WR)</i>	X	0.57	X	3.48
user	<i>usuario</i>	X	10.08	X	36.72
search engine	<i>buscador</i>	X	0.96	X	2.13
	<i>motor de búsqueda (CU)</i>	-		X	0.13

Table 3. Frequency of equivalents in *CdE* and *CREA*.

Moreover, most of the equivalents are not followed by information that shows how to use them in a given context:



Chart 5. Type of context following equivalents in the *CU*.

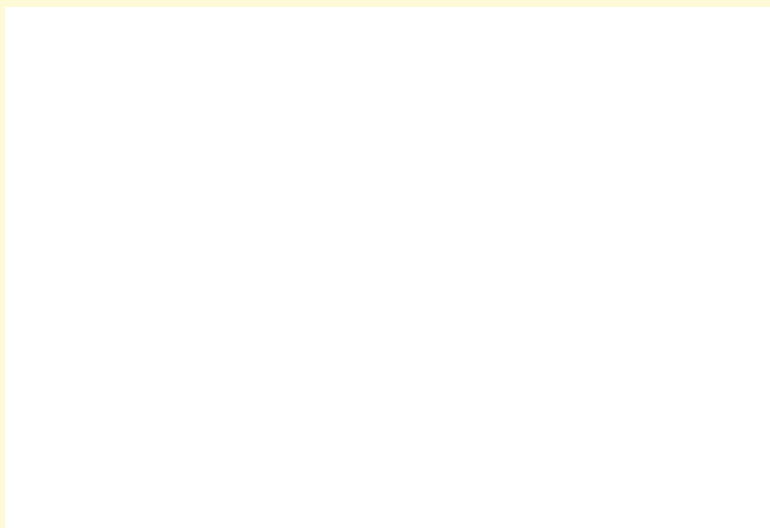


Chart 6. Type of context following equivalents in the *GDO*.



Chart 7. Type of context following equivalents on *WR*.

Hence, the variety of equivalents proposed could become an obstacle for dictionary users because they do not have enough data to decide which equivalent fits better into a given context.

## Conclusions

The volume of specialised vocabulary used in general discourse and gathered in the selected dictionaries differs in each of them. In fact, in two of the dictionaries, *GDO* and *WR*, the whole sample is included, but in the *CU*, some of the terms are not covered. In addition, the methods used for field labelling differ in each of the dictionaries; for example, in the *CU*, abbreviations are used, but in the *GDO*, the compilers decided to label using the name of the field. In addition, on *WR*, a great range of different resources can be found to label specialised terms.

Moreover, the percentage of labelled terms differs in each of the selected dictionaries, a fact which shows that this procedure has not been carried out with the desired accuracy. This finding has direct implications for users and, especially, for translators and interpreters, who will find different information depending on

the dictionary they utilise. Furthermore, it would be desirable to include some contextual data in the next editions, as the equivalents are isolated, and no indications about their use can be found in most of the analysed entries.

Finally, we can confirm that the equivalents proposed in the selected dictionaries are used in Spanish. However, as they appear without context of their use, we should train translators and interpreters in the use of dictionaries and complement their training with the employment of other tools, such as parallel corpora, until editors overcome the identified barriers. Now the need arises for testing other fields or areas of knowledge that may or may not confirm these tendencies.

### Note

1. TN: Regarding bilingual dictionaries, our English (B) students usually consult *Oxford*, *Collins* or both of them.

### References

Anthony, L. (2012). *AntConc (3.2.4. version)* [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from <http://www.antlab.sci.waseda.ac.jp/>

Atkins, B. T. S. ,& Rundell, M. (2008).*The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Atkins, B. T. S., & Varantola, K. (1998a). Language Learners Using Dictionaries: The Final Report on the EURALEX/AILA Research Project on Dictionary Use. In B. T. S. Atkins (Ed.), *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators* (pp. 21–81). Tübingen: Max Niemeyer Verlag.

Atkins, B. T. S., & Varantola, K. (1998b). Monitoring Dictionary Use. In B. T. S. Atkins (Ed.), *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators* (pp. 83–122). Tübingen: Max Niemeyer Verlag.

Boulanger, J. C. (2001). L'aménagement des marques d'usage technoléciales dans les dictionnaires généraux bilingues. In J. Pruvost (Ed.), *Les dictionnaires de la langue française* (pp. 247–271). Paris: Honoré Champion.

Bowker, L. (2012). Meeting the needs of translators in the age of e-lexicography. In S. Granger & M. Paquot (Eds.), *Electronic Lexicography* (pp. 379–397). Oxford: Oxford University Press.

Corpas Pastor, G., Leiva Rojo, J., & Varela Salinas, M. J. (2001). El papel del diccionario en la formación de traductores e intérpretes: análisis de necesidades y encuestas de uso. En M. C. Alaya Castro (Ed.), *Diccionarios y enseñanza* (pp. 239–273). Alcalá de Henares: Universidad de Alcalá.

Davies, M. (2002). *Corpus del Español: 100 million words, 1200s-1900s*. Retrieved from <http://www.corpusdelespanol.org>.

De Schryver, G. M. (2003). Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography*, 16 (2), 143–199.

Drouin, P. (2003). Term extraction using non-technical corpora as a point of leverage. *Terminology*, 9 (1), 99–117.

Durán Muñoz, I. (2010). Specialised Lexicographical Resources: a Survey of Translators' Needs. In S. Granger & M. Paquot (Eds.), *e Lexicography in the 21st century: New Challenges, New applications. Proceedings of ELEX 2009, Cahiers du Cendal* (pp. 55–66). Louvain-la-Neuve: Presses Universitaires de Louvain-La-Neuve.



Fernández Quesada, N. (2009). El diccionario bilingüe y los errores de traducción en los estudiantes de inglés como lengua 'c'. In J. M. García Platero & M. A. Castillo Carballo (Eds.), *Investigación lexicográfica para la enseñanza de lenguas* (pp. 149–163). Málaga: Servicio de Publicaciones de la Universidad de Málaga.

Forget, N. (1999). *Les dictionnaires électroniques dans l'optique de la traduction*. Thèse de maîtrise. École de traduction et d'interprétation. Université d'Ottawa. Retrieved from <http://aix1.uottawa.ca/~etithese/nforg/>

Gallimberti, B., & Russel, R. (Eds.) (2008). *Gran Diccionario Oxford: Español-Inglés, Inglés-Español*. 4<sup>a</sup> ed. Oxford: Oxford University Press.

Hastings, A. (2000). El diccionario inglés-español, español-inglés como herramienta para el traductor. In R. C. Gonzalo García & V. García Yebra (Eds.), *Documentación, terminología y traducción* (pp. 109–122). Madrid: Síntesis.

Jessen, A. (1996). *The Presence and Treatment of Terms in General Dictionaries*. M. A. Thesis. Ottawa: University of Ottawa.

Josselin, A. (2005). *Place et rôle des terminologies dans les dictionnaires généraux unilingues et bilingues. Étude d'un domaine de spécialité: Volcanologie*. PhD Thesis. Lyon: Université Lumière Lyon II, Centre de Recherche en Terminologie et Traduction. Retrieved from [http://theses.univ-lyon2.fr/documents/lyon2/2005/josselin\\_a#p=0&a=top](http://theses.univ-lyon2.fr/documents/lyon2/2005/josselin_a#p=0&a=top)

Landau, S. I. (2001). *Dictionaries: The Art and Craft of Lexicography*. 2nd Edition. Cambridge: Cambridge University Press.

McEnery, A., Xiao, R., & Tono, Y. (2006). *Corpus Based Language Studies: An Advanced Resource Book*. London: Routledge.

Muñoz Sánchez, P. (2006). Electronic Tools for Translations in the 21st Century. *Translation Journal*, 10 (4). Retrieved from <http://www.bokorlang.com/journal/38tools.htm>.

Ortego Antón, M. T. (2012). *Estudio contrastivo inglés/español del tratamiento del léxico especializado recogido en los diccionarios generales bilingües: Collins*

*Universal y Gran Diccionario Oxford*. Tesis doctoral. Soria: Universidad de Valladolid. Retrieved from <http://uvadoc.uva.es/handle/10324/2022>.

Pastor, V., & Alcina, A. (2010). "Search Techniques in Electronic Dictionaries. A Classification for Translators, *International Journal of Lexicography*, 23(3), 307-354.

Perrino, S. (2009). User-Generated Translations: the Future of Translation in a Web 2.0 Environment, *The Journal of Specialised Translation* 12, 55-78.

Real Academia Española. (2001). *Corpus de referencia del español actual (CREA)*. Retrieved from <http://corpus.rae.es/creanet.html>

Roberts, R. P. (1997). Using Dictionaries Efficiently. *38th Annual Conference of the American Translators Association*. San Francisco, California.

Roberts, R. P. (2004). "Terms in General Dictionaries" In J. M. Bravo Gozalo (Ed.), *A New Spectrum of Translation Studies* (pp. 121-140). Valladolid: Universidad de Valladolid.

Roberts, R. P., & Josselin, A. (2005). Le traitement des termes dans les dictionnaires généraux. In H. Bejoint & F. Maniez (Eds.), *De la mesure dans les termes. Hommage à Philippe Thoiron* (pp. 324-348). Lyon: Presses Universitaires de Lyon.

Rodríguez Reina, M. P. (2002). Las marcas de especialidad: Una cuenta pendiente de difícil solución. In G. Guerrero Ramos & M. F. Pérez Lagos (Eds.), *Panorama actual de la terminología* (pp. 327-357). Granada: Comares.

Sager, J. C., Dungworth, D., & McDonald, P. F. (1980). *English Special Language. Principles and Practice in Science and Technology*. Wiesbaden: Oscar Brandstetter.

Sánchez Ramos, M. M. (2004). *El uso de los diccionarios electrónicos y otros recursos de Internet como herramientas para la formación del traductor (inglés-español)*. Tesis doctoral. Castellón de la Plana: Universitat Jaume I.

Schriren, R., Love, C., & Amiot-Cadley, G. (Eds.) (2009). *Collins Universal Español-Inglés, English-Spanish*, 9th edition. Barcelona: Random House Mondadori/Glasgow: Harper Collins Publishers.

Sierra Soriano, A. (2001). *El diccionario bilingüe. Estructura y nomenclatura*. Alicante: Editorial Club Universitario.

Svensen, B. (2009). *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.

Thoiron, P. (1998). Place et rôle de la terminologie dans les dictionnaires bilingues non spécialisés. Le cas de la terminologie médicale dans le Dictionnaire Hachette-Oxford (français-anglais). In S. Mellet & M. Vuillaume (Eds.), *Mots chiffrés et déchiffrés* (pp. 621-650). Paris: Honoré Champion.

WordReference, (2014). Retrieved from <http://www.wordreference.com/es/>

Recebido em: 19/10/2014

Aceito em: 14/01/2015