

The usage of field labels in English-Spanish bilingual e-dictionaries from the perspective of translators

María Teresa Ortego

University of Valladolid

Campus Duques de Soria s/n, 42004 Soria (Spain)

E-mail: tortego@lesp.uva.es

Abstract

Translation is the vehicle to spread progress among societies that do not share the same language. However, the translation of specialized vocabulary is a problem that translators have to face. The first sources they look to when they do not know an equivalent are general bilingual e-dictionaries, according to previous research. In order to distinguish specialized vocabulary from general vocabulary, dictionaries use different mechanisms and the most prevalent one is field labelling. Our aim is to study how field labels are used to tag specialized vocabulary, so we analyse these field labels in the macrostructure and microstructure of a selection of general bilingual e-dictionaries from a translator's perspective. In the macrostructure of each dictionary, we look for references to the use and selection of field labels, we search for a list of field labels, then we check whether all of these are included in entries, and we extract the most representative fields, having counted the number of entries tagged with field labels. In the microstructure, we find where dictionaries place field labels in the structure of the entry, the typology used, and we test whether dictionaries label the same units by analysing five randomly selected units. Finally, we show the analysis results found in each dictionary, we compare them and we draw conclusions.

Keywords: bilingual e-dictionary; field; field label; specialized vocabulary; translation

1. Introduction

The priority for scientists is to spread progress among societies that do not share the same language, and the way to achieve this is through translation. However, the translation of specialized vocabulary is a problem that translators have to face. The first sources they look to when they do not know an equivalent are general bilingual e-dictionaries, according to previous research (Meyer, 1988; Roberts, 1990; Mackintosh, 1998; Varantola, 1998; Atkins & Varantola, 1998; Corpas et al., 2001). Although it can be thought that bilingual e-dictionaries do not contain specialized vocabulary because they are referred to as "general", they actually include a representative selection of lexical units belonging to different linguistic levels and subsets of language (Haensch, 1997) that a middle class user knows by his or her culture and the influence of the media.

However, translators are users who have some features which differ from other e-dictionary users: they master the source language and the target language, and they are trained in dictionary search. While they are not trained in the field for which they translate, they have the skills in order to learn quickly about the topic of translation.

Having justified the inclusion of specialized lexical units in bilingual e-dictionaries and described the profile of translators as users of general bilingual e-dictionaries, we focus on the most used mechanism to distinguish specialized vocabulary from general vocabulary, that is, field labels. Field labels are very helpful for translators because they show the field to which lexical units belong, especially in lemmas with polysemous meanings, helping translators to choose the correct meaning and hence, the correct equivalent for the context.

Our aim is to study how field labels are used to tag specialized vocabulary in the main English-Spanish general e-dictionaries, and we try to answer the following questions:

1. Do dictionaries explain the use and selection of field labels?
2. How many field labels are used in each dictionary?
3. How many fields are represented? Which are the fields with the highest number of tagged entries?
4. Where are field labels located in each entry of the dictionary?
5. Do dictionaries use the same typology of field labelling?
6. Do dictionaries tag the same specialized vocabulary?

2. Methods

To answer these questions and determinate how field labels are used to tag specialized vocabulary, first of all, we select two general bilingual e-dictionaries. Then, we explain how we are going to analyse information in the macrostructure and microstructure of the selected general bilingual e-dictionaries.

2.1 The selection of general bilingual e-dictionaries

We select two general bilingual e-dictionaries according to the following parameters extracted on previous research about analysis and assessment of dictionaries (Mary Haas, 1964; in Landau, 2001; Cabré & Gelpí Arroyo, 1996; Roberts, 1997; Landau, 2001; Santamaría Pérez, 2003; Gelpí Arroyo, 2003; Atkins & Rundell, 2008): to be an e-dictionary; to

mention translators among its users; to be an unabridged dictionary (Roberts: 1997); to distinguish between English and Spanish varieties; to have an intuitive structure so that learning how to use it does not take too much time for users; to be reliable, that is, to be based on corpus during its compilation; and, to be accessible, so that the user can find the dictionary in bookshops and libraries.

The dictionaries chosen which fulfil these criteria are:

-GALLIMBERTI, B. & RUSSELL, R. (eds.): *Gran diccionario Oxford: Español-Inglés*,

Inglés-Español. 4th edition. Oxford: Oxford University Press, 2008; from now on *GDO*.

-SCRIBEN, R. et al. (dirs.): *Collins Universal Español-Inglés, English-Spanish*. 9th edition. Barcelona: Random House Mondadori; Glasgow: Harper Collins Publishers, 2009, from now on *CU*.

Those dictionaries are offline e-dictionaries for PC:

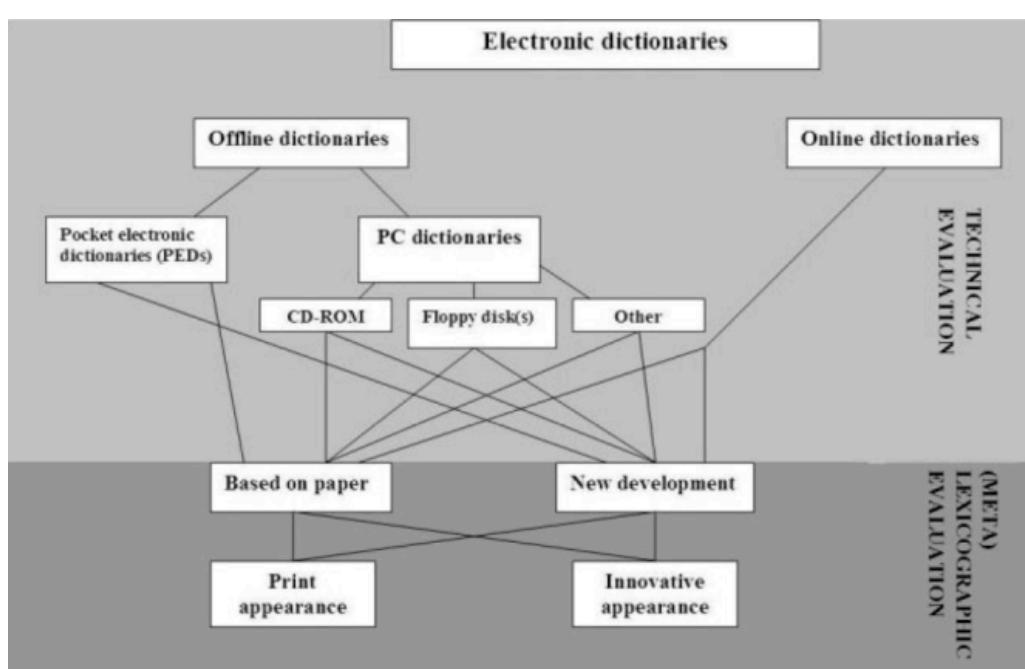


Figure 1: Classification of electronic dictionaries by Lehr (1996, translated in de Schryver, 2003)

CU and *GDO* use an easy interface, they include translators among the range of users to which the dictionary is aimed, and they are divided into two sections, one for understanding (English-Spanish) and the other for production (Spanish-English). In terms of their sizes, *GDO* includes almost 60,000 entries in each section. However, while we have not found any reference to the precise number of entries in *CU*, it is mentioned that more than 750,000 references and translations are included. Both dictionaries tag entries with usage labels to show the place where the unit is used.

	Spanish	English
CU	27	11
GDO	24	6

Table 1: Number of labels used to distinguish dialects.

Both dictionaries use different sizes of letters and different colours to distinguish different types of information, and so are well structured for new users.

Moreover, during the compilation of *CU* and *GDO*, lexicographers based their work on previous editions and on the use of corpora. *CU* is based on the results obtained in *Bank of English*¹ and *Banco de Español*, while *GDO* is based on *Oxford English Corpus*² and *Oxford Reading Programme*.

The last parameter, accessibility, is clearly satisfied - according to a study from Corpas Pastor et al. (2001), the

¹ Bank of English is the actual name for COBUILD, compiled by the University of Birmingham. Now it is composed by 550 million words of everyday English from different oral and written texts.

² Oxford English Corpus is composed by texts collected from 2000 until now. In 2010, it contained more than two billion words from different sources.

selected dictionaries are the most used dictionaries by translation students from the University of Malaga (Spain).

Having chosen and justified the e-dictionaries in which we are going to analyse field labels, the next step in our research is to describe how we are going to carry out the analysis.

2.2 The analysis of macrostructure and microstructure

In order to answer the proposed questions, we divide the analysis in to two parts. In the first part, we focus on the macrostructure in order to answer questions 1-3. We look for any references to field labelling in the Help section of the e-dictionaries. Then, we look for a list with the number of field labels used in each dictionary. Once we have the list, we test if each field label is used in each section of the selected dictionaries and we compare the results between sections and between dictionaries. Then, we try to find a list of domains and verify if the number of domains used in each section matches up with the list of field labels. Next, we find the domains with the highest number of tagged entries by using the searching options available in each dictionary.

The second part of the analysis is to answer questions 4-6 by looking at the microstructure of the dictionary. We observe how each dictionary uses field labels within the entry: place, typology, abbreviations, etc.

Then, we select five words at random belonging to specialized vocabulary from the most represented fields. We test if they are tagged in each section of the selected dictionaries following the principle of reversibility³ (Svensen, 2009).

Finally, we compare the results obtained in *CU* and *GDO*.

3. Results

In this section we describe the results obtained from the search in the selected dictionaries in order to find answers to the proposed questions.

3.1 Do dictionaries explain the use and selection of field labels?

As we have mentioned before, references to the use and selection of field labels should be described in the Help section. *CU* explains that field labels are used when the meaning of a word is technical. It also offers a list with every kind of abbreviation and label. However, *GDO* is less explicit than *CU* and does not offer advice on the use of field labels. In spite of the lack of data about field

³ The principle of reversibility means that the equivalent must be included in the other section of the dictionary. For example, if we look up *exit* in English-Spanish section, the equivalent *salida* should be included as an entry in Spanish-English section and the equivalent proposed should be *exit*.

labels, it offers a list with the different labels used. From both lists, we have extracted only field labels.

3.2 How many field labels are used in each dictionary?

From the total list of labels used in each of the selected e-dictionaries, we chose only those which refer to fields. In *CU*, we found 99 abbreviations related to a field which are used to tag specialized vocabulary. Then, we tested if all of them appeared in both sections of the dictionary and the result is that 24 of them were duplicated.

	English-Spanish	Spanish-English
Architecture	Archit	Arquit
Biology	Bio	Biol
Commerce	Comm	Com
Sewing	Sew	Cost
School	Scol	Escol
Pharmacy	Pharm	Farm
Railways	Rail	Ferro
Philosophy	Philos	Fil
Physics	Phys	Fís
Physiology	Physiol	Fisiol
Photography	Phot	Fot
Computers	Comput	Inform
Mathematics	Math	Mat
Mechanics	Mech	Mec
Meteorology	Meteo	Met
Mythology	Myth	Mit
Music	Mus	Mús
Nautical	Naut	Náut
Optics	Opt	Ópt
Psychology	Psych	Psic
Chemistry	Chem	Quím
Theatre	Theat	Teat
Technical	Tech	Téc
Typography	Typ	Tip

Table 2: Fields with two labels in *CU*.

For example, to tag a specialized lexical unit from Computers, *CU* uses the label (Comput) in the English-Spanish section and the label (Inform) in the Spanish-English section.

GDO offers a list of abbreviations from which we extracted field labels. However, the microstructure of the dictionary does not use any abbreviations. Instead, it uses the name of the domain in English for the English-Spanish section (Ex: Medicine) and in Spanish for the Spanish-English section (Ex: Medicina). In all, *GDO* uses 188 field labels, 94 labels in each section.

3.3 How many fields are represented? Which are the fields with the highest number of tagged entries?

Although one might think that the number of field labels and fields represented in each section of the dictionary would match up, we found that the number of fields

represented in *CU* rises to 75 whereas the number of labels is 94.

We found that if all the fields tag any entry in both sections of *CU*, the result is that the following fields are not included:

ENGLISH-SPANISH		SPANISH-ENGLISH	
FIELD	LABEL	FIELD	LABEL
Biology	Bio	Stock Exchange	St Ex
Science	Sci	Science	Sci
Sport	Dep	Ecology	Ecol
Bullfighting	Taur	Skiing	Ski
		Government	Govt
		Industry	Ind
		Radio	Rad

Table 3: Fields which are not represented by sections.

So according to the figures, *CU* includes 71 fields in the English-Spanish section and 66 fields in the Spanish-English section.

In *GDO*, only 89 fields are used to tag entries in the English-Spanish section: Arms, Entertainment, Printing and Publishing, Bullfighting and Wine are not included. The Spanish-English section only contains 81 fields. We were unable to find lexical units tagged with Anthropology, Post, Railways, Nuclear Physics, Printing, Publishing, Civil Engineering, Electric Engineering, Chemist Engineering, Mechanics, Occultism, Labour Relations and Tourism.

The second part of the question is to find the most represented fields. This question is complicated because the search engine in *CU* does not accept brackets in the search options nor recognizes the difference between upper and lowercase letters. So, we had to count the lexical units tagged with field labels individually.

ENGLISH-SPANISH		SPANISH-ENGLISH	
FIELD	ENTRIES	FIELD	ENTRIES
Medicine	588	Medicine	861
Military	571	Military	688
Computers	465	Commerce	621
Commerce	416	Law	595
Law	406	Sport	568
Automobiles	405	Politics	565
Music	362	Religion	537
Politics	355	Nautics	535
Nautics	332	Technical	519
Economy	320	Botany	469

Table 4: The most representative fields by sections in *CU*.

The same procedure was carried out in *GDO*. However, it was easier than in *CU* because the search engine accepts brackets and capital letters.

In the following table we offer a synthesis of the most represented fields in *GDO*:

	ENGLISH-SPANISH	SPANISH-ENGLISH	
Field	Entries	Field	Entries
1 Sport	534	Medicine	417
2 Law	445	Law	394
3 Computing	437	Sport	387
4 Medicine	424	Zoology	343
5 Military	411	Cookery	336
6 Music	367	History	331
7 Linguistics	329	Religion	302
8 Cookery	320	Military	280
9 Religion	301	Music	271
10 Finance	281	Botany	265
		Computing	

Table 5: The most represented fields by sections in *GDO*.

3.4 Where are field labels located in each entry of the dictionary?

The location of field labels depends on the meaning of the lexical units they label. For monosemous lexical units, field labels are placed after the spelling, pronunciation and grammar category in both dictionaries. For polysemous lexical units, field labels are placed after the number of letters which indicates which sense is specialized and before the equivalent.

3.5 Do dictionaries use the same typology of field labelling?

CU uses abbreviations between brackets, in italics and a blue colour. *GDO* uses the name of the field in English in the English-Spanish section and in Spanish in the Spanish-English section, between brackets and the first letter in uppercase. It also uses a blue colour.

3.6 Do dictionaries tag the same specialized vocabulary?

From the list of the most represented fields, we have randomly selected five lexical units:

FIELD	LEXICAL UNIT
Medicine	pacemaker
Military	tank
Computers	defender
Law	bailiff
Music	flash memory

Table 6: Randomly selected units.

We searched those lexical units in the English-Spanish section of each selected e-dictionary and then we tested if the selected dictionaries followed the principle of reversibility. The results of searching lexical units in English-Spanish sections and the equivalents proposed as entries in the Spanish-English section of both dictionaries are shown in the following tables:

ENGLISH-SPANISH		SPANISH-ENGLISH	
Lemma	Equivalent	Lemma	Equivalent
Pacemaker	(Med) marcapasos	Marcapasos	Pacemaker (Sin arca)
Tank	(Mil) tanque, carro (de combate)	Tanque	(Mil) tank
		carro	(Mil) tank
defender	(Sport) defensa	Defensa	(Dep) la defensa (= jugadores) the defence, the defense (EEUU)
bailiff	(Jur) alguacil	Alguacil	(Jur) bailiff, constable
Flash memory	Memoria flash	Memoria flash	

Table 7: Results in CU.

ENGLISH-SPANISH		SPANISH-ENGLISH	
pacemaker	(Medicine) marcapasos	Marcapasos	Pacemaker (sin marca)
tank	(Military) tanque, carro de combate	Tanque	(Armas) (carro) tank
		Carro de combate	Tank (sin marca)
defender	(Sport) defensa	defensa	(Deporte) (conjunto) defense* Defensa (jugador) defender
Bailiff	(Law) (in UK) alguacil (in US) funcionario que custodia al acusado en un juzgado	Alguacil	(oficial) bailiff (sin marca)
Flash memory	(Computing) memoria flash	Memoria flash	Flash memory

Table 8: Results in GDO.

From the previous table, we observe that only *tank-tanque* and *defender-defensa* in CU and GDO are tagged with field labels in both sections. Moreover, only *bailiff-alguacil* are tagged with field labels in GDO.

Then, we have found two phenomena which affect other units. Firstly, units are described with field labels in one section and without field labels in the other section. For example in the entry of *pacemaker*, the equivalent *marcapasos* is tagged with (Med) and (Medicine) in both dictionaries, whereas the entry *marcapasos* in the

Spanish-English section doesn't have any field labels. Secondly, some equivalents in English-Spanish are not included as entries in the nomenclature of the Spanish-English section. For example, the equivalent of *flash memory*, *memoria flash*, is not included in the nomenclature of the Spanish-English section.

4. Conclusions

In conclusion, we find that labelling specialized vocabulary in general bilingual e-dictionaries is not systematic.

In the macrostructure, GDO does not describe the procedure followed by compilers to select and tag specialized vocabulary and to insert it into the nomenclature. Furthermore, neither of the selected dictionaries offers a list composed by field labels used. The lists consulted include all type of labels, so translators and other users find hard to get used to labels, especially when they have to use two or more dictionaries.

The number of field labels used in the selected dictionaries does not match up. This phenomenon also occurs with the fields listed in the dictionaries. In addition, the number of field labels does not fit between sections in the same dictionary.

So we reach to the conclusion that dictionaries from the same size and category do not include the same proportion of specialized vocabulary, and the results reached by a translator will depend on the dictionary used to aid translation.

In the microstructure, the place that field labels take in the entry is the same in both dictionaries, although the typography changes. CU prefers the use of abbreviations while GDO uses the name of the field between brackets. It would be useful to normalize the typography of field labels and the fields used in dictionaries of same typology in order to make searches easier for translators. Moreover, labelling of specialized vocabulary is not systematic. If we look at the results of the comparison of five randomly selected lexical units, we observe that they are labelled in one section of the dictionary, but the equivalents are not labelled in the other section or even equivalents are not included in the nomenclature.

To sum up, we will continue with some more studies into specialized lexical units in the framework of our PhD project. Meanwhile, we would like to ask compilers and editors to normalize the use of field labels in general bilingual e-dictionaries, to be more systematic in the labelling of equivalents in other sections of the dictionary, and to be more careful in field labelling in order to save users time and to save translators time in their searching tasks. In addition, dictionaries should be more systematic and mark the same units equally in both sections.

5. Acknowledgements

I would like to thank the financial funds from University of Valladolid (Spain) to assist to e-LEX2011.

6. References

- Atkins, B.T.S., Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Atkins, B.T.S., Varantola, K. (1998). Monitoring dictionary use. In B.T.S. Atkins (ed.) *Using dictionaries: Studies of dictionary use by language learners and translators*. Tübingen: Max Niemeyer, pp. 83-122.
- Cabré, M.T. & Gelpí Arroyo, C. (1996). La lexicographie bilingue catalane contemporaine: analyse et évaluation. In H. Béjoint, P. Thoiron (eds.), *Les Dictionnaires bilingues*. Louvain-la Neuve: Duculot, pp. 213-230.
- 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. In M.C. Alaya Castro (ed.) *Diccionarios y enseñanza*. Alcalá de Henares: Universidad de Alcalá, pp. 239-273.
- De Schryver, G.-M. (2003). Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography*, 16(2), pp. 143-199.
- Gelpí Arroyo, C. (2003). Mesures d'avaluació lexicogràfica de diccionaris bilingües. PhD thesis. Barcelona: Universitat Pompeu Fabra.
- Haensch, G. (1997). *Los diccionarios del español en el umbral del siglo XXI: Problemas actuales de la lexicografía*. Salamanca: Universidad de Salamanca.
- Landau, S.I. (2001). *Dictionaries: The art and craft of lexicography*. Cambridge: Cambridge University Press.
- Mackintosh, K. (1998). An empirical study of dictionary use in L2-L1 translation. In B.T.S. Atkins (ed.) *Using dictionaries: Studies of dictionary use by language learners and translators*. Tübingen: Max Niemeyer, pp. 123-149.
- Meyer, I. (1988). The general bilingual dictionary as a working tool in "thème". *Meta*, 33(3), pp. 368-376.
- Roberts, R. (1990). Translation and the bilingual dictionary. *Meta*, 35(1), pp. 74-81.
- Roberts, R. (1997). Using dictionaries efficiently. *38th Annual Conference of the American Translators Association*, San Francisco, California.
- Santamaría Pérez, M.I. (2003). *La fraseología española en el diccionario bilingüe español-catalán: Aplicaciones y contrastes*. Alicante: Universidad de Alicante.
- Svensen, B. (2009). *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Varantola, K. (1998). Translators and their use of dictionaries: User needs and user habits. In B.T.S. Atkins (ed.) *Using dictionaries: Studies of dictionary use by language learners and translators*. Tübingen: Max Niemeyer, pp. 179-192.