A NEW DICTIONARY OF COMPUTER SCIENCE: COMPLEMENTING GRAMMATICAL INFORMATION

JORDI PIQUE-ANGORDANS AND SANTIAGO POSTEGUILLO

Universitat de València and Universitat Jaume I de Castelló
jordi.pique@uv.es; postegui@fil.ubi.es

INTRODUCTION

Bilingual lexicography seems to have awakened in recent decades from a long period of almost oblivion, particularly in terms of the study of what a bilingual dictionary is, what should it contain and what are its main uses. Bilingual lexicography has indeed served a variety of purposes which have been determined by the development of society itself. However, the use of bilingual dictionaries in the classroom has been for long accepted as being a deterrent to language learning (Thompson, 1987), and their drawbacks have also been extensively treated, and monolingual dictionaries seem to avoid these deficiencies. However, Thompson further points out some major differences between monolingual and bilingual dictionaries, namely that headwords are defined in the foreign language and also that words are defined in the target language. These are two of what he calls "essential differences" (Thompson, 1987: 283) and makes a strong point towards the benefits of using bilingual dictionaries in class particularly for below-advanced level learners.

Our point here is not simply to indulge in Thompson’s plea for the use of bilingual dictionaries, but to take it one step further and plainly state that bilingual dictionaries, if adequately structured and built, are a necessary tool both in class as well as on every professional’s desk, and particularly the translation professionals. We want to bring into the open and dwell on the idea of a fully active dictionary, as opposed to the passive one. In teaching, Thompson (1987: 285) suggests a dictionary aimed in one direction, that is, either for Spanish learners of English or for English learners of Spanish. We would, however, promote a bilingual dictionary in both directions; secondly, a dictionary with full description of the foreign headwords, their syntactic behavior; and thirdly,
the use of many collocations and examples in the foreign language. The following words are based on our experience in translating a bilingual dictionary of computing (Collin et al., forthcoming) and the examples we propose are, therefore, derived from it.

THE CONCEPT OF ACTIVE VS. PASSIVE BILINGUAL DICTIONARY

A bilingual dictionary is usually understood as a dictionary for two national languages where the source language (SL) headwords of the dictionary are provided with the equivalent translation in the target language (TL). Bergenholtz and Tarp (1995: 49) point out that most specialized bilingual dictionaries published “have been designed to provide linguistic information, primarily in the form of equivalents, with a view to translation into and from the native language”. In most cases, however, the linguistic information provided is not enough, since it is usually partial and only present in one section of the dictionary; in other words, these dictionaries are usually designed to be unidirectional.

The typical bilingual dictionary has often been disregarded as a useful tool either for teaching or for translation purposes because it lacks information, both grammatical and syntactical, and in general how the foreign language behaves. Salerno (1999: 210-211) stresses the point that a passive dictionary is basically a bidirectional volume and its fundamental purpose is for decoding the TL, and it lacks at least some grammatical information which should appear close to the headword. Salerno (1999: 211), on the one hand, stresses the fact that a bilingual active dictionary, to be really active, must be fundamentally a support for encoding the TL. She further adds, quoting from Kromann et al. (1984: 212), that “it is first and foremost the equivalent which should be supplied with the morphological and syntactic information the user requires in the translation situation”. Gak (1992: 329), on the other, points out that “La spécificité du dictionnaire de type actif se manifeste tant dans sa microstructure que dans sa macrostructure”.

Rather than considering the macrostructure of a dictionary, that is, its whole architecture (including preface, tables, illustrations, annexes, etc.), we are going to concentrate on the microstructure. Although it has to do fundamentally with the selections of individual entries, we want to tackle the organization of each entry, what elements should it contain and what its functions are.

In addition, an active bilingual specialized dictionary should be undertaken having the “translation principle” as its basis. As Fuertes-Olivera and Velasco-Sacristán (2001: 37) point out, this means “giving a real translation and taking into consideration phraseology and meaning in context”. Ignoring these two ideas would produce not only inadequate information but it would also be an insufficient aid to professional translators and students alike.

ENTRY STRUCTURE

In planning a dictionary, one of the main efforts should concentrate on providing all the necessary and useful information for its prospective user. As pointed out by Mudget (1989: 129), dictionaries are usually consulted for two possible functions in mind: (a) text reception, that is, the analysis of a given word or collocation, and its decoding; and (b) text production, that is, for the purpose of synthesis and, especially, encoding the text. He also adds that, in some tasks like translation or reading aloud, both functions are involved. According to these premises, a bilingual dictionary understood as a tool for both decoding and encoding a given TL, we would certainly call for and foster a truly bidirectional dictionary which would cater to the needs of professionals and students alike. This would, in some way, oppose Thompson’s (1987) view of a dictionary aimed in only one direction.

This entry structure, however, is being constrained when we come to the distinction between a bilingual dictionary for language in general as opposed to a specialized bilingual dictionary, or a dictionary of Language for Specific Purposes (LSP). The grammar content in the first case is preeminent, while in the second its lexi-co-semantic specialization is preeminent. This is the reason why Mudget (1989: 230) sustains that in LSP dictionaries, parts-of-speech labels seem to be rare. In our computer dictionary we intended to follow a completely different principle and all parts of speech were labeled adequately. A similar pattern was followed in terms of gender. Not so much emphasis was placed on marking word formation, although new words resulting from derivation are also included. In addition, unlike medical sciences which take many words either from Latin or Greek and many technical terms are made up of these roots often combined with prefixes and suffixes, computer science technology is distinguished by the constant use of semi-technical words, many of which are taken from everyday life, along with newly-created words which lexicographers
have to come to grips with day after day. Content and specialization then become the main concern when the translation process begins and, as a result of this consideration, sufficient collocations and examples are included in the appropriate entries. Gak (1992: 331) puts it this way:

La spécialisation de sens s'impose dans tous les cas où les volumes sémantiques de deux langues ne coïncident pas. La spécialisation est montrée dans le dictionnaire à l'aide des traductions, des explications et des exemples.

While in some languages the main concern might be the verbal system or the gender of words, as is the case in Spanish, a bilingual LSP dictionary will be more concerned with collocations, explanations, examples and their translation from and into the TL.

Taking this into consideration, our ideal entry should be as complete as possible and should include at least the following elements next to the headword: word class, gender, inflection and syntax information when necessary, collocations, and examples. These elements should be sufficient for the professional or the language learner so that a general language dictionary would not be necessary to solve language problems. In the case of a technical dictionary such as ours, information is also necessary in the case of some difficult concepts which can only be understood by an expert in the subject matter. As Gak (1992: 335) says, an active dictionary should be more than just a simple dictionary, "il doit être une sorte d'encyclopédie de la langue dans laquelle on traduit".

HEADWORD SELECTION AND ENTRY CONTENT

Bergenholz and Tarp (1995: 71) insist on the fact that specialized bilingual dictionaries, particularly in science and technology, should be based on an "exhaustive lemma selection". They further add that this selection usually encounters different approaches in the process:

- (a) LSP terms only,
- (b) 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.

(Bergenholz & Tarp, 1995: 103)

The first concern in the development of our dictionary, therefore, was the selection of terms. In this respect this task was supplied through a corpus drawn from the already published dictionaries in computing and the Internet by Peter Collin Publishing. To this corpus, a number of extra materials was added, particularly extra examples and encyclopedic information, extracted from popular journals, published in Spanish, such as PC World, Cyberg@ts. Recently published books, like Cebrían (2000), Castells (2001), or Yus (2001), and including widely accepted textbooks, such as Hammacher (1984) and Tanenbaum (1989), and their translation into Spanish, were also consulted for additional examples.

The next step was to decide what sort of information, particularly grammatical, should be added, as well as collocations and examples, and including encyclopedic information and quotations from relevant publications. The initial corpus had already been established with the required planning to provide the necessary systematic classification of each entry in the subject area of the dictionary.

In the following figure, we will show the treatment of the headword "address", signaled by a number in square brackets [1]. The different items added to the entry are represented the same way, as follows: [2] denoting different word class (in Figure 1, "address" used as noun, and later as verb); [3] word class; [4] order of semantic meaning; [5] semantic description of word and encyclopedic information; [6] translation; [7] gender; [8] nominal group and collocation; [9] contextualization of the headword in a sentence; [10] quotation from specific literature; [11] source of quotation; [12] cross reference.

---

[1] address
[2] 1
[3] noun
[4] (a) Europe
[5] (street, town, etc.)
[6] dirección
[7] f
[8] address list =
[9] we keep an address list
[10] [5] (in central processing unit)
[11] m
[12] * lista
[13] [7] f
[14] [8] address list =
[16] [9] each separate memory address has its own unique address
[17] [10] in dirección
[18] [11] m
[19] [12] address =
[20] [13] m
[21] [14] address =
[22] [15] list of node addresses;
[23] [16] list of the network addresses

[1] we keep an address list de la memoria tiene su dirección propia

[2] 1

[3] noun
two thousand businesses in Europa

[4] (a) Europe

[5] (street, town, etc.)

[6] dirección
de direcciones con dos mil sociedades comerciales en Europa

[7] f

[8] address list =

[9] we keep an address list de la memoria tiene su dirección propia

[10] [5] (in central processing unit)

[11] m

[12] * lista
each separate memory address has its own unique address

[13] [7] f

[14] [8] address list =


[16] [9] each separate memory address has its own unique address

[17] [10] in dirección
de direcciones con dos mil sociedades comerciales en Europa

[18] [11] m

[19] [12] address =

[20] [13] m

[21] [14] address =

[22] [15] list of node addresses;

[23] [16] list of the network addresses

[6] * velocidad
del tiempo
address 1 noun (a) (street, town, etc.); dirección f o domicilio m; address list = repertorio m o lista f de direcciones; we keep an address list of two thousand businesses in Europe = mantenemos un listado de direcciones con dos mil sociedades comerciales en Europa; (b) (in central processing unit) dirección; each separate memory word has its own unique address = cada palabra individual de la memoria tiene su dirección propia; this is the address at which the data starts = ésta es la dirección donde comienzan los datos; absolute address or actual address or direct address = dirección absoluta o dirección real o dirección directa o dirección de máquina; address access time = tiempo m o velocidad f de acceso a la dirección; address base = base f de la dirección; (list of node addresses; list of the network addresses) address book = libro m de direcciones; address bus = enlace m común o bus m de direcciones; (special code that identifies an address) address code = código m de direcciones; address computation = cálculo m de una dirección; address counter = contador m de direcciones; address decoder = decodificador m o descodificador m de direcciones; address format = formato m de direcciones; address mapping = tabla f de correspondencia de direcciones o topografía f de las direcciones; address mark = marca f de direcciones; address modification = modificación f de direcciones; address register = registro m de direcciones; (used to link one IP address to a low-level physical address) address resolution protocol (ARP) = protocolo m de resolución de direcciones o protocolo ARP; see also TCP/IP; address space = espacio m de direcciones o espacio direccional; address strobe = señal f o impulso m de validez de la dirección; address track = pista f de direcciones; (address produced by calculating an expression) address translation = traducción f de direcciones; address word = palabra f de direcciones; base address = dirección base; (address of the node to which data is being sent) destination address = dirección de destino; initial address = dirección de origen o inicial; machine address = dirección (de) máquina; network address = dirección de red; relative address = dirección relativa; 2 verb (a) (letter) dirigir o enviar; to address a letter or a parcel = dirigir una carta o un paquete; (b) (in CPU) direccionar o definir una dirección; a larger address word increases the amount of memory a computer can address = una palabra de dirección más larga aumenta la cantidad de memoria que un ordenador puede definir.

The world's largest open data network, the Internet, links more than 10,000 local networks and 3 million workstations in 50 countries. It has grown so fast that its address space is 'bust' and is being redesigned to allow further expansion.

Computing

Figure 2. Entry "address" formatted in dictionary form (adapted from Collins et al. [forthcoming]).
As can be observed in Figure 1, the headword was usually accompanied with an indication of word category, indicating with a digit if two or more word classes were present in the same entry, and their corresponding semantic rendering and description. Gender was also incorporated next to the Spanish translation of nouns, or indicating with -a when the double possibility of an adjective existed; for instance, the rendering of the English adjective ‘analytical’ would be represented as analítico, -a. Numerous collocations and examples have been added, along with technical or encyclopedic information and quotations from specialized journals.

Grammatical information would be one of the main additions to the typical bilingual dictionary. In our entries, both word class and gender would be indicated from the onset, as in the following example:

**mirror** 1 noun espejo m; mirror disk = disco m espejo; mirror image = imagen f inversa o de espejo; 2 verb (a) (make an identical copy) hacer una copia idéntica; (b) (duplicate all disk operations onto a second disk) hacer una copia espejo; there's less chance of losing your data now that you have mirrored the server's disk drive = hay menos posibilidades de perder los datos ahora que ha hecho una copia idéntica de la disquetera del servidor

Mirroring of the database is handled automatically by systems software

**Computer News**

Figure 3. Headword “mirror” with two word categories.

Word formation has also room in an active bilingual dictionary. Words derived from a generally used prefix are also incorporated in this dictionary. See for instance the words derived from the prefix mini- in the following entry:

**mini**- prefix; (small) mini-; (usually a 3.5 inch disk) minidisk = (mini)disquete m; (slang) miniwinnix = minidisco m duro (de tipo Winchester)

Figure 4. Prefix “mini-” and uses.

And among some of the words derived, we can find entries such as for the following words: minicomputer, minifloppy, minimax, among others.

Often, synonyms and antonyms are also incorporated into the entry, as well as information comments. In addition, notice the indication of the variety of English which may affect the spelling of a given word, as in the following two entries:

**maximise** or **US maximize** verb maximizar or llevar al máximo; (expand an application icon) ampliar la pantalla or una ventana or un icono (del sistema Windows); compare MINIMISE

**COMMENT:** you maximise a window by clicking once on the up arrow in the top right hand corner

**minimise** or **US minimise** verb (a) minimizar or reducir; we minimised costs by cutting down the number of components = redujimos los costes al recortar el número de componentes (b) (shrink an application window to an icon) minimizar or transformar una ventana en un icono; compare MAXIMISE

**COMMENT:** the application can continue to run in the background; you minimise a window by clicking once on the down arrow in the top right hand corner

Figure 5. Two opposing terms with indication of two varieties of English and comments.

One of the main difficulties encountered by translators is the abundance of acronyms in computer science. This has caused the inclusion of many cross references (in capital letters) within each section of the dictionary through which sufficient information is provided. See, for instance, the following entry for such common acronym as PC, colloquially and academically used in both languages. In it, information is also included which makes reference to computers already superseded by modern computer generations, data which is also included as additional historical information:

1. Complete words in capital letters would indicate a cross-reference entry and redirect the user to that headword or expression.
PC (a) PERSONAL COMPUTER ordenador m personal or ordenador individual or PC m; (computer compatible with the IBM PC) PC compatible = ordenador personal compatible; (IBM PC compatible computer) PC/ AT = PC/AT; (keyboard with twelve function keys) PC/ AT keyboard = teclado m PC/AT; (keyboard with ten function keys) PC/XT keyboard = teclado PC/XT; (b) PRINTED CIRCUIT (BOARD); (c) PROGRAM COUNTER

In the UK, the company is known not for PCs but for PC printers

Which PC?

Figure 6. Entry with acronyms and cross-references.

Finally, encyclopedic information, or technical information, would also be included in some entries where information of this kind may be needed to clarify some concepts. Such is the case of the following entries:

analytical adjective analítico, -a; analytical engine = máquina analítica (de Babbage)

COMMENT: mechanical calculating machine developed by Charles Babbage in 1833 that is generally considered the first general purpose digital computer

DMA = DIRECT MEMORY ACCESS; (interface IC that controls high-speed data transfer) acceso m directo a la memoria or DMA; DMA controller = controlador m DMA or de acceso directo a la memoria; DMA cycle stealing = acceso directo a la memoria mediante utilización en robo de ciclo

A 12-bit DMA controller, 16-bit video I/O ports and I/O filters complete the chip.

Figure 7. Entries with acronyms, comment, and quotation.

We understand that providing the information contained in the different entries, both the students' and the translators' task will be greatly facilitated. The doubts professional translators verbalize through the existing computer information for over possible translations of controversial terms, made us consider the need to widely address them. However, even with this decision in mind, one comes to the realization that not all the difficult translations will be solved satisfactorily. Consider the following nominal group: “address access time”. The obvious initial translation could be “tiempo de acceso a la/una dirección”. However, when this cluster is contextualized it may also mean “velocidad de acceso a la/una dirección”. This is why we opted for providing both possibilities, although a more complete dictionary would also include examples with such a contextualization. In addition, we have tried to comply, with the numerous examples given, to provide pragmatic information both on “register” and also on what Widdowson (1978: 57) calls “usage” through which the mode in which the language is manifested is emphasized.

AN ONLINE VERSION OF THE DICTIONARY

Online dictionaries are an underdeveloped area of study in lexicography (Campoy, 2002). This is a gap that should be offset with both further research and more practical lexicographical undertakings. In this respect, although our dictionary is to be published first in traditional printout form in 2004, we are already working in the development of an online version of the dictionary. In order to achieve this, first the initial publisher’s reluctance has to be overcome. Many publishers feel uncomfortable with editing their material for online versions since they feel this will diminish their profits from printout sales. This may be true and opens an important debate which needs to be tackled by authors, publishers and users together. If we want more and better online dictionaries we may have to accept to pay a price for them (by subscriptions) or else always struggle with publishers’ reluctance for the development of such online resources. This debate, however, is not the focus of our paper. Instead, we briefly describe in the following lines how an online version of our dictionary could be organized.

An online specialized dictionary is quite different in terms of use to a traditional printout version. Searches are carried out through a Boolean engine after keying in the specific term we are looking for. Links allow for a much faster cross-reference system. In this respect, we would activate links in each entry highlighting in blue any terms used in an example that are also defined in the dictionary. This would allow users to quickly access complementary entries to obtain further linguistic input.
Another important addition in an online version would be the incorporation of pronunciation information in each entry. This input would be not only in written form (i.e. with the appropriate phonetic transcription of the term in the international phonetic alphabet), but also including the possibility to hear the actual pronunciation of the term by clicking on the phonetic transcription. The lack of phonetic information is a major gap in specialized lexicography which online versions could certainly fill.

Finally, an online version would also allow not only for internal links (i.e. links for cross-referencing within the text or links with sound files), but could and should also incorporate external links. For instance, an online version could connect the examples drawn from specific quotations from various journals and publications to their online sites when available. This would be the case for most of the quotations and examples drawn from specialized computer publications which have online sites. Finally, the technical comments could incorporate links to sites where further technical information may be found along with visual diagrams to complement the explanations.

CONCLUSION

As mentioned in the literature, the interest in the construction of a dictionary and how much grammatical information should be incorporated has increased over the last few decades, and as pointed out by Salerno (1999: 219), particularly in the field of the English monolingual learner’s dictionaries. However, the area of specialized bilingual dictionaries has also received a lot of attention, and this has been the aim of their publishers. They have encouraged the addition of many more examples and collocations, along with the necessary grammatical information. This has been our aim in the development of our dictionary and it is hoped that we have accomplished it. In fact, this is what has been carried out in a set of dictionaries which are currently being developed by the University Institute of Applied Modern Languages (Instituto Universitario de Lenguas Modernas Aplicadas, IULMA) in Alacant and Castelló (Spain). This Institute is now working on the development of specific dictionaries related to important industrial sectors in the area, namely the shoe industry, the stone and marble industry, or the ceramics industry, among others. A final step in the development of these new set of specialized dictionaries will include the generation of their respective online versions. These versions will include internal links for cross-referencing, and external links to connect quotations and technical comments with relevant sites and, finally, links to sound files for phonetic input on the right pronunciation of terms.

REFERENCES


Housmann, F. J., Reichmann, O., Wiefand, H. E., and Zgusta, L. (eds.)
Mugdan, J. 1989. “Grammar in dictionaries of Languages for Special Purposes
(LSP)”.
Hermes 3: 125-142.
Salerno, L. 1999. “Grammatical information in the bilingual dictionary. A study
of five Italian-French dictionaries”. International Journal of
Lexicography 12,2: 211-222.
Englewood Cliffs, NJ: Prentice-Hall International (Spanish version by
F. Gómez Molinero: Sistemas operativos: Diseño e implementación.
286.
University Press.