

Guiding principles for the elaboration of an English-Spanish dictionary of multi-word expressions

Eduardo de Gregorio-Godeo
Universidad de Castilla-La Mancha

So-called word combinations – also referred to as multi-word combinations or multi-word expressions – take shape when certain words regularly combine with certain other words or grammatical constructions. When exploring the word combinations of a language, both collocations and idiomatic expressions to a large extent examined. Collocations and idioms are usually taken to be multi-word expressions whose meaning is more than the sum of the meaning of their components.

Focusing on multi-word expressions in bilingual dictionaries, this contribution accounts for an ongoing research and editorial project guiding the elaboration of an English-Spanish dictionary of multi-word combinations. After presenting the lexicographic process leading the elaboration of the dictionary as such, this contribution will proceed to describe the principles determining the inclusion of entries and their presentation in the dictionary.

The rationale for this project is based on current lexicographic practices (Hartman 2001) having comprised four stages: (1) pre-lexicographic work, which consisted of a thorough examination of the market of English-Spanish dictionaries given the lack of specific dictionaries dealing with multi-word expressions in this area; (2) the research undertaken for the elaboration of the macrostructure of the dictionary and the use of various sources (e.g. existing English monolingual or multi-word dictionaries; bilingual dictionaries, and corpora), especially as far as usage examples, equivalents and their idiomaticity is concerned; (3) description issues, with a special emphasis on both the description of the multi-word expressions included in the dictionary, and the actual structure of dictionary entries; and (4) final formatting, which entails final presentation and revision prior to editing and publishing the dictionary.

Considering Spanish-speaking students of EFL and – to a lesser extent – translators as the potential users of this dictionary, this contribution will conclude with some final remarks of the educational implications of the project herein presented.

1. Introduction

So-called *word combinations* – also referred to as *multi-word combinations* or *multi-word expressions* – take shape when certain words regularly combine with certain other words or grammatical constructions (Benson, Benson and Ilson 1997: ix). When exploring the word combinations of a language, both collocations and idiomatic expressions are to a large extent examined. Collocations and idioms are usually taken to be multi-word expressions whose meaning is more than the sum of the meaning of their components (Ilson 2002: 333).

Focusing on multi-word expressions in bilingual dictionaries, this contribution accounts for a research and editorial project in progress guiding the elaboration of an English-Spanish dictionary of multi-word combinations. After presenting the lexicographic process leading the elaboration of the dictionary as such, this paper proceeds to describe the principles determining the inclusion of entries and their presentation in the dictionary.

2. Rationale for the lexicographic practice conducted

The rationale for this project is based on current lexicographic practices (Hartman 2001) having comprised four stages: (1) pre-lexicographic work, which consisted of a thorough examination of the market of English-Spanish dictionaries given the lack of specific dictionaries dealing with multi-word expressions in this area; (2) the research undertaken for the elaboration of the macrostructure of the dictionary and the use of various sources (e.g. existing English monolingual or multi-word dictionaries; bilingual dictionaries, and corpora), especially as far as usage examples, equivalents and their idiomaticity is concerned; (3)

description issues, with a special emphasis on both the description of the multi-word expressions included in the dictionary, and the actual structure of dictionary entries; and (4) final formatting, which entails final presentation and revision prior to editing and publishing the dictionary.

3. Pre-lexicographic work

According to Harman (2001), successful lexicography conceived of as dictionary making requires careful planning and implementation of the compilation process based on market research and the delineation of the potential users' reference needs to be met. After a detailed exploration of the existing market of English-Spanish dictionaries, a significant lack of specific dictionaries dealing with multi-word expressions was noticed. As a result, a bilingual dictionary of collocations and idioms seemed to be, without a doubt, a must for intermediate-to-advanced learners of EFL and, to a lesser extent, translators.

4. Research issues

Having completed the planning stage of the lexicographic process, the project commenced by producing a through database of multi-word expressions in English serving as a basis or macrostructure for the alphabetically-ordered dictionary entries in English including Spanish equivalents, usage examples and other information for the prospective English-Spanish dictionary of word combinations.

The macrostructure of lexical entries has been elaborated on the basis of existing English monolingual dictionaries of multi-word expressions, chiefly *The BBI Dictionary of English Word Combinations* (Benson, Benson and Ilson 1997) and, to a lesser extent, the *Oxford Collocations Dictionary* (2004). Apart from various other contemporary English-Spanish dictionaries, dictionaries of English idioms have likewise been examined.

The British National Corpus (BNC) has been employed as the most important source of information for the usage examples of the entries. Nonetheless, when usage examples were not found on the BNC, other English-language corpora were consulted, chiefly the *Collins COBUILD English Collocations on CD-ROM* (1995). When no examples were found within existing language corpora, the Internet was drawn upon as a resource for usage examples. Equivalents in Spanish for the word combinations that the dictionary consists of have been provided taking into account bilingual English-Spanish general-use dictionaries on the market. A wide range of such bilingual dictionaries have been accordingly utilized. The CREA (*Corpus of Reference of Contemporary Spanish*) corpus has been used in this respect to test the appropriateness and idiomaticity of the equivalents in Spanish.

5. Description issues (i): Selecting multi-word expressions

Bearing in mind the features of this word-combination dictionary, Benson, Benson and Ilson's (1986: 252-254) taxonomy of lexical combinations has been considered. Such a typology of multi-word combination includes: (a) *free combinations of words*, which are those whose components are the freest in regard to combining with other lexical items; (b) *idioms*, which are relatively frozen expressions whose meanings do not reflect the meanings of their component parts; and (c) *collocations*, which are fixed/recurrent word combinations, that is, loosely fixed combinations between free word combinations and idioms. Given that the meaning of free combinations of words may be found in a relatively easy way through an

examination of the meanings of the individual constituent words, this type of multi-word expressions has not been included in the dictionary herein presented.

When it comes to *collocations*, Benson, Benson and Ilson (1986) make a difference between *grammatical collocations* and *lexical collocations*. The former may be taken to be a phrase consisting of a dominant word (noun, adjective, verb) and a preposition or grammatical structure such as an infinitive or clause; for instance, *decide on* in *decide on a date*. The latter do not contain prepositions, infinitive or clauses. Typical lexical collocations consist of nouns, adjectives, verbs and adverbs; for example, *sincere apologies* in *accept my sincere apologies*. By and large, this dictionary focuses on lexical collocations. The main types of lexical collocations included in the dictionary are thus the following:

- Verb + Noun / Pronoun / Prepositional Phrase. The majority of these collocations consist of a verb of action or activation as well as a noun or a pronoun (e.g. *reach an agreement*).
- Adjective + Noun. An example of this category could be *dark/light blue*. If a number of adjectives may possibly collocate with a single noun, only the most frequent lexical collocations are included. Moreover, as in English many nouns have an adjectival function when they are placed before another noun (e.g. *application form*), such collocations appear in the dictionary entry of the second noun.
- Noun + Verb. These collocations take shape where the verb reflects characteristic actions of the noun (e.g. *wars break out*). Combinations which are easily predictable are not included in the dictionary (e.g. *painters paint*).
- Lexical collocations indicating the ‘unit’ commonly associated to a noun. Their structure in English is often *noun₁ of noun₂*. Such collocations convey the meaning that an individual belongs to a larger group (e.g. *a bunch of flowers*), or the specific and concrete character of a unit with regard to larger elements (e.g. *a piece of news*).
- Adverb + Adjective. Some examples of this type of collocation are *utterly different*.
- Verb + Adverb. Typical examples of this type of collocations could be *to appreciate deeply, greatly, keenly, sincerely*.

In addition to collocations, the dictionary incorporates a wide range of *idiomatic expressions* of contemporary British and American English falling within different categories: (i) traditional idioms (e.g. *Spill the beans*); (ii) new phrases (e.g. *it's all gone pear-shaped*); (iii) metaphorical phrases (e.g. *face the music*); (iv) two-word phrases (e.g. *wild card*); and (v) various other similes (e.g. *like two peas in a pod*). The dictionary does not incorporate phrasal verbs as a characteristic type of multi-word combination, as there are different bilingual dictionaries of this type.

6. Description issues (ii): Organisation of entries

This dictionary has been organised in an alphabetical order, which is based on the written form of the lexically relevant units rather than on their meaning, thereby adopting a semasiological approach in entries, that is, going from name to notion. The basic principles guiding the organisation of entries within the dictionary are as follows:

Determiners, prepositions and pronouns are not often headwords. Key headwords, compounds included, are alphabetically ordered in the dictionary. Single-word compounds precede those written as two words. Homographs follow this order: adjective, adverb, noun, verb. For instance, in the following entry, *cement* (n.) precedes *cement* (v.):

CEMENT I n.

1. to mix; pour cement *mezclar; verter cemento*. I haven't yet no, but I think he's outside mixing cement, we said we'd go and give him a hand to put his posts in [...]

CEMENT II v.

1. to cement smt. into place *consolidar, fortalecer, cementar algo*. When nest building, he will return to his nest with bits of weed to cement it into place, even when the diver is only a few inches away.

Entries contain at least one context of usage. The key headword is written in small capital letters. English collocations are written in bold and Spanish equivalents appear in italics. Regular font face has been adopted for usage examples, the word combination in question being underlined. Here is an example:

CORTEX n.

the cerebral cortex *corteza cerebral*. This deficiency is linked in some way to the death of nerve cells that arise in the forebrain and connect to many regions of the cerebral cortex.

Cross-references within the dictionary are highlighted in yellow. For example, in the following entry *way* is highlighted in yellow because the same idiom may be found in the entry for *way* in the dictionary:

CAT n.

1. to neuter a ; to spay a (female) cat *castrar, capar; esterilizar un gato*.

[...]

6. there's more than one way to skin a cat *cada maestrillo tiene su librillo*. There is a saying about there being more than one way to skin a cat: in the case of toxic wastes they could be contained in an indefinitely leakproof box.

When a word collocates with others, the dictionary identifies series of collocations alphabetically (e.g. **COST: to cut, reduce**). This helps to save space in the dictionary and to display synonyms and near-synonyms. So, in the series for *cost* below, *cut costs* and *reduce costs* are treated as synonyms. However, collocations which are not synonyms are separated by semicolon (;) (e.g. **to defray; drive up; to pay; spare no COST**). Synonyms are thus grouped together and separated by commas within the series of collocations:

COST n.

1. to defray; drive up; to pay; spare no; cut, reduce costs *sufragar; hacer subir; pagar; no ahorrar; recortar costes*. During 1983 the Inland Revenue received £27,000 million from wage-earners and small businesses, much of which was used to defray the cost of the nation's social welfare responsibilities.

Collocations whose meaning may be difficult to understand by the student are often defined between brackets on the left of the collocation:

COUNTER II n.

1. [surface, table over which business is conducted] a bargain; notions (IAm); check-in; check-out; ticket; lunch counter *mostrador de artículos de mercería; de facturación de equipajes; de caja; de control de billetes; donde se sirven comidas*. Its members occupied seats reserved for whites in a lunch counter at a Woolworth store in Greensboro, North Carolina, after they had been refused service.

Information about register, dialects and other social factors is also indicated between brackets, e.g. *formal* (form.), *American English* (IAm), *British English* (IBr); etc:

CURD n.

soybean (IAm), soya bean (IBr) curd *tofu/ queso de soja*. Widely used in Eastern cookery and in strict vegetarian diets in other parts of the world, tofu is bean curd and can be eaten cold or cooked in a number of ways.

As the following example of entry from the dictionary instantiates, idiomatic expressions are written in blue at the end of entries:

CURIOSITY n.

1. to arouse, excite, pique, whet; satisfy one's curiosity *producir, despertar, excitar, avivar; satisfacer la curiosidad de uno*. He claimed he had taken the measurements of Eclipse during life, and wished to satisfy his curiosity by trying them on the dead subject.

[...]

3. curiosity killed the cat *la curiosidad mató al gato*. The townspeople had learned the hard way that curiosity killed the cat - you stayed indoors if there was trouble.

Alternative translations of a collocation or idiom are separated by a slash (/):

CALAMITY n.

[...]

3. a calamity befalls smb; occurs *acaecerle a alguien un desastre / una catástrofe; suceder un desastre*. An even worse calamity occurred when the bore hole pump packed up.

Usage notes are highlighted in light blue, and range from pragmatic information to differences between British and American English, through other grammatical questions. They are also used to make reference to 'false friends'. Here is an example:

CAN n.

1. (IAm) an ash, garbage, trash can *cubo de la basura*. Then he seized the bag of cookies, and on his way back indoors, threw them where he considered they belonged, in the trash can with the rest of the garbage.

Nota de uso: El inglés británico prefiere dustbin.

The use of a long underscore (___) in an idiomatic phrase indicates that various nouns, adjectives or verbs may be inserted in the gap:

CAPITAL n.

[...]

13. ___ with a capital ___ (trouble with a capital T, fast with a capital F) *en mayúsculas, se usa con cualquier letra para enfatizar que se habla de algo en sentido extremo*. But it doesn't have to be exercise with a capital E. Think of it more as activity.

7. To conclude

At the time of publication of this contribution, the dictionary is close to completion, and the final stage of the project still needs to be undertaken. This will entail formatting, printing and proofreading the whole dictionary several times. A detailed revision of the dictionary will consequently have to be conducted before it is ready to be published.

All in all, we may conclude by saying that, concentrating on collocations and idioms as key multi-word-combination categories, this contribution has accounted for a project responding to the shortage of bilingual dictionaries by providing instant and user-friendly access to Spanish equivalents of English multi-word expressions in the form of an English-Spanish dictionary of multi-word combinations. Given that collocations and idioms are a fundamental area of concern for learners of EFL at an intermediate and advanced level, the dictionary will be a particularly valuable resource for EFL students.

References

- Benson, M.; Benson, E.; Ilson, R. (1986). *Lexicographic Description of English*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Benson, M.; Benson, E.; Ilson, R. (1997). *The BBI Dictionary of English Word Combinations*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- BNC (*British National Corpus*). Oxford: Oxford University Computing Services. 2001.
- Collins COBUILD English Collocations* [cd-rom]. London: HarperCollins. 1995.
- CREA (*Corpus de Referencia del Español Actual*) [on line]. [Madrid: Real Academia Española] <http://corpus.rae.es/creanet.html> [access date: 02 April 2010]
- Hartman, R. (2001). *Teaching and Researching Lexicography*. Harlow: Longman. 2001.
- Ilson, R. (2002). 'Lexicography'. In Malmkjær, K.; Anderson, J. (eds.). *The Linguistics Encyclopedia*. London/New York: Routledge. 331-339.
- Oxford Collocations Dictionary: for Students of English*. Oxford: Oxford University Press. 2004.