COLLOCATION DICTIONARIES: A COMPARATIVE ANALYSIS

Miriam Buendía Castro

mbuendia@ugr.es
University of Granada

Pamela Faber

pfaber@ugr.es
University of Granada

Abstract

The importance of phraseological information in lexicographic resources is experiencing an exponential growth. This is evident in the publication in recent years of a wide variety of combinatorial or collocation dictionaries. This paper describes and compares the main monolingual collocation dictionaries for English and Spanish in regards to the following: (i) types of collocation encoded; (ii) kinds of collocational information offered; (iii) place for collocations in the micro or macrostructure of the dictionary. The objective of this analysis is to study the usefulness of these resources for translators.

Resumen

El creciente interés por la información fraseológica se constata en la gran cantidad de diccionarios combinatorios o de colocaciones que han surgido en los últimos años. En este artículo se describen y se comparan los principales diccionarios monolingües de colocaciones en inglés y en español teniendo en cuenta lo siguiente: (i) el tipo de colocaciones que codifican; (ii) el tipo de información colocacional que ofrecen; (iii) el lugar que ocupan las colocaciones en la micro y macroestructura del diccionario. El objetivo último de este análisis es estudiar la utilidad de estos recursos para los traductores.

Keywords: Phraseology. Collocations. Dictionaries. Lexicography. Translation.

Palabras clave: Fraseología. Colocaciones. Diccionarios. Lexicografía. Traducción.

MonTI 6 (2014: 203-235). ISSN 1889-4178
http://dx.doi.org/10.6035/MonTI.2014.6.7

1. Introduction

The Function Theory of Lexicography (Bergenholtz & Tarp 2010) highlights the need to take specific user needs into account in the design of a lexicographic resource. Also important is the fact the same user can perform different roles, depending on the context. For example, a translator can sometimes assume the role of a teacher or a linguist. In the words of Nuccorini (2003: 367):

> it must be borne in mind that often the dictionary-intended addressees do not coincide with the actual users and that, on the other hand, different roles are often performed by the same individual (for example a teacher and an advanced user, a linguist and a translator) who might adopt different perspectives.

There is a general consensus among translators that phraseological information in lexicographic resources is crucial, especially in the final production of the target language text. In this phase, the translator may need grammatical and syntactic information related to terms, including collocations in the target language. As such, the more collocations a dictionary contains, the better it can fulfil its function (Bergenholtz & Tarp 2010: 33). As Rundell states (2010: vii), collocations are as important as grammar, and they are what makes speakers sound natural and fluent:

> Why is collocation so important? Firstly, it is a central feature of language, and – whether you are speaking or writing – it is just as important as grammar. Getting the grammar right is an essential part of producing text which is free of errors. But selecting appropriate collocations is one of the keys to sounding natural and fluent. […] Secondly, collocation contributes to meaning. Most common words in English have more than one meaning, and we use the surrounding context to indicate (or work out) which meaning is intended. Collocations play a big part in this process.

The growing importance of phraseological information can be seen in the recent publication of combinatorial or collocation dictionaries. In this paper, the most representative English and Spanish collocation dictionaries for general language are described and compared in order to evaluate their potential usefulness for translators.

2. Phraseology in collocation or combinatorial dictionaries: description and comparison

To date, there are three general collocation dictionaries on the market for English,[1] namely, the *BBI Combinatory Dictionary of English* (BBI) (1986, 1997, 2009), *Oxford Collocations Dictionary for Students of English* (OCD) (2002, 2009), and *Macmillan Collocations Dictionary* (MCD) (2010). Collocation dictionaries for Spanish include the following: *Redes: Diccionario combinatorio del español contemporáneo* (REDES) (Bosque 2004), *Diccionario combinatorio práctico del español contemporáneo* (PRÁCTICO) (Bosque 2006), and the *Diccionario de colocaciones del Español* (DICE) (Vincze, Mosqueira & Alonso 2011). This section provides an overview of how collocational information is treated in these dictionaries, based on the headwords, 'bed' and 'cama', for English and Spanish resources, respectively. Since the DICE focuses on the domain of emotion, it does not contain the entry "bed". Therefore, 'indignación' [indignation] is the word used to describe the microstructure of an entry in the DICE.

As shall be seen, the way that each resource encodes, classifies, and displays collocations varies greatly. According to Nuccorini (2003: 367):

> the delimitation and description of contents, the theoretical principles adopted for the inclusion, selection, classification and presentation of headwords, the sources and the layout of phraseological dictionaries vary considerably both linguistically and lexicographically.

As such, there is a wide range in the number of headwords in the three English dictionaries (McGee 2012: 335). The BBI includes approximately double the number of entries as the OCD, and the OCD contains twice as many entries as the MCD. Therefore, collocations for a less common word are more likely to be found in the BBI. One thing in which collocation dictionaries normally coincide is that they are meant for text production (Nuccorini 2003: 367):

> English monolingual collocational dictionaries agree on one point: they are meant for encoding purposes and are consistently addressed to advanced learners and translators.

1. Although not described in this paper, other English collocation dictionaries worth mentioning are: (i) *A Dictionary of English Collocations* (Kjellmer 1994); (ii) *LTP Dictionary of Selected Collocations* (Hill & Lewis 1997). In the same line as McGee (2012: 327), the reason for not including Kjellmer's dictionary is that it was not conceived for learners of English and it focuses on adjacent collocations. The LTP is not included since it provides less coverage than the other three dictionaries described (e.g. it does not include the headword 'bed').

Although such dictionaries are generally used for production purposes, some of the most recent ones on the market can also be useful for decoding purposes (e.g. Bosque 2004). In our opinion, a collocational resource that is useful for translators should be focused on both objectives.

The *Explanatory Combinatorial Dictionary* (ECD) (Mel'čuk, Clas & Polguère 1995; Mel'čuk 1984-1999) deserves special mention in that it is the most influential combinatorial dictionary for any language. In fact, many of the dictionaries described in this section are based on the theoretical and methodological premises of the ECD. Its objective is to provide a systematic and formal description of the set of linguistic properties of lexical units, referred to as *lexemes* or *phrasemes,* which convey a specific meaning (L'Homme & Leroyer 2009: 271). The ECD proposes an inventory of *lexical functions* (LFs), a formal system for encoding collocations within the ECD. LFs are specified for each lexical unit (L) in the dictionary. They are part of a system designed to represent a large set of lexical relations. According to Mel'čuk (1996: 39), a lexical function *f* is a correspondence that associates a given lexical unit (L) (the argument or keyword, i.e. the base of the collocation) with a set of lexical items (L_1) (the collocate) which express a specific meaning associated with *f*. It can be represented by the following formula: *f* (L) = L_1. It should be highlighted that lexical functions are not lexical units of the language, but rather correspond to metalexies.

2.1. The BBI Dictionary of English Word Combinations

Of the six collocation dictionaries in this section, the *BBI Dictionary of English Word Combinations* (Benson, Benson & Ilson 1986, 1997, 2009) was the first to be compiled. Its third edition is evidence of its success. The 1997 edition included 18,000 entries and roughly 90,000 collocations. The most recent 2009 edition is an extensive revision with new collocations in the field of computing and Internet. It also includes new example sentences and more detailed descriptive usage notes (Benson, Benson & Ilson 2009: xi).

This dictionary is based on the *Meaning-Text Theory* of Mel'čuk (Mel'čuk et al. 1984-1999). Unlike the OCD and MCD, the BBI is the only dictionary that is not corpus-based, but rather 'corpus-refined' (McGee 2012: 330), which means that it is based on the authors' intuition. In the words of Benson, Benson & Ilson (2009: viii):

> Nowadays, our task is eased not only by the availability of corpuses of contemporary English (such as the British National Corpus) but also by the amazing resource of the Internet itself, which enables us to search in it for a word and find superb examples of that word in context. Nor should it be

> forgotten that an important source of new information in BBI 3 is, paradoxically, BBI 2, now that the computer allows material from an entry in BBI 2 to be added to other entries in BBI 3 when such material is appropriate.

Another striking difference in comparison to the OCD and MCD is that apart from including lexical collocations, the BBI is also a syntactic dictionary that supplies extensive information about the complementation patterns of verbs, nouns, and adjectives. The dictionary is thus referred to as a *combinatory dictionary,* rather than a *collocation dictionary* since it includes not only phraseology, but also valency (Benson, Benson & Ilson 2009: i):

> Traditionally, the combination of words into grammatical patterns has been called colligation or complementation or construction (though in BBI it is called collocation, too) and its result has been called valency. A dictionary that provides both phraseology and valency is a dictionary of word combinations; or, in the terminology of Igor Mel'čuk, whose work has inspired us, a combinatory or combinatorial dictionary. BBI is a combinatory dictionary.

In this sense, the BBI includes two types of collocations: (i) grammatical collocations also referred to as *colligations*[2] by other authors such as Hoey (2005); and (ii) lexical collocations. *Grammatical collocations* consist of a dominant word (a noun, adjective/participle, verb) and a preposition or a grammatical structure, such as an infinitive or clause (e.g. noun + *to* + infinitive). They are always listed under the dominant word; whereas *lexical collocations* are formed by constructions such as the following: verb + noun, adjective + noun, noun + verb, noun + noun, adverb + adjective, and adverb + verb (Benson, Benson & Ilson 2009: xiii).

Lexical collocations are divided into seven types, namely, L1, L2, L3, etc. L1 collocations are mainly transitive verbs denoting Creation or Activation + an NP/PP (e.g. 'come to an agreement'). These combinations are referred to as CA collocations (Benson, Benson & Ilson 2009: xxxi). L2 collocations include verbs meaning Eradication and/or Nullification (e.g. 'reject an appeal'), which are referred to as EN collocations (Benson, Benson & Ilson 2009: xxxii). However, the BBI insists on the arbitrary character of EN and CA combinations, which makes it difficult for foreigners to produce them spontaneously. L3 collocations have the pattern of adjective + noun (e.g. 'strong tea'). L4 collocations are noun + verb combinations (e.g. 'alarms go off'). L5 collocations are noun + *of* + noun combinations (e.g. 'a bouquet of flowers'). L6 collocations are adverb + adjective combinations (e.g. 'deeply absorbed'); and

2. See Stubbs (2002) for a complete description regarding the differences between collocation and colligation.

finally, L7 collocations are verb + adverb combinations (e.g. 'affect deeply'). Regarding the listing of lexical collocations in the BBI when there is a noun in the collocation, the collocation is placed under the noun. If there are two nouns, it appears under the second noun; if there is no noun, then it appears under the adjective. When there is no noun or adjective, it is placed under the verb.

Generally speaking, the BBI does not include free combinations, which are defined as "elements that are joined in accordance with the general rules of English syntax and freely allow substitution" (Benson, Benson & Ilson 2009: xix). Figure 1 shows an example of an entry of the BBI corresponding to 'bed':

Bed *n.* ["article of furniture for sleeping"] 1. to make, make up a ~ (I'll make you up a ~ = I'll make a ~ up for you) USAGE NOTE: *to make up a bed* is to put bedclothes on it. *to make a bed* is to re-arrange the bedclothes on it neatly after someone has slept in it. 2. to strip; unmake a ~ 3. to go to ~ with (she went to ~ with a heating pad) USAGE NOTE: *to go to bed with smb.* means "to have sexual intercourse with smb." 4. to be, lie, stay in ~; to lie; sit on a ~; to wet the ~ ("to urinate in it accidentally") 5. to get out of ~ 6. to put smb. to ~ (to put the children to ~) 7. to take to one's ~ ("to remain in bed because of illness") 8. a double; king-size; queen-size; single; twin ~ 9. a bunk; camp (BE); folding; four-poster; hospital; rollaway; sofa; truckle (BE), trundle (AE) ~ 10. a water ~ 11.an unmade ~ 12. a feather ~ ("a feather mattress") (see also *bed and breakfast* at **breakfast**) ["ground at the bottom of a body of water"] 13. a river ~ ["plot of ground"] 14. a flower ~

Figure 1. Entry for 'bed' in the BBI

As shown in Figure 1, after the headword and its grammatical category, a definition is provided in square brackets and double quotation marks. After the definition, lexical collocations are listed. In the BBI, lexical collocations always precede grammatical collocations. Since the example does not contain grammatical collocations, all the combinations displayed correspond to lexical ones. For certain collocations, an additional explanation is given in double quotation marks and sample squares, e.g. 7 to take to one's ~ ("to remain in bed because of illness"). Other collocation senses also offer a usage example, e.g. 1. (I'll make you up a ~ = I'll make a ~ up for you). For some articles, usage notes are included (see sense 1), and these provide additional information concerning the headword. As shown in Figure 1, varieties of English are marked as AE, for American English (e.g. 'trundle bed'), and BE (for British English) (e.g. 'camp bed').

The order of presentation of lexical collocations inside entries are verb + noun (CA collocations) (e.g. 'make up a bed'), verb + noun (EN collocations)

(e.g. 'unmake a bed'), adjective + noun (e.g. 'a double bed'), noun + verb (none in this entry), noun + noun (e.g. 'a flower bed'). The BBI thus has a more systematic approach since all collocations are assigned to a category based on their pattern (L1, L2, L3, G1, etc.), and then patterns are consistently arranged in each dictionary entry in the order specified in the introduction to the dictionary. Despite this systematization, the BBI includes many set phrases that do not fit into any of the previously described types of grammatical and lexical collocations (Benson, Benson & Ilson 2009: xxxv). Such phrases are normally listed under the label *misc* (miscellaneous).

One of the positive aspects of the dictionary is that it does not exclusively rely on lexical collocations. As previously mentioned, it is also a syntactic dictionary which supplies information about the complementation patterns of verbs, nouns, and adjectives. Although Cowie (1998: 225) argues that complementation of verbs and nouns should be treated in a valency dictionary and not in a collocational one, in our opinion, constructional information for verbs should be included in a dictionary of collocations. In the words of Siepmann (2005: 416):

> collocation and verb complementation are intimately related, since many noun-verb collocations require a specific distribution of semantic roles.

The main drawback of the BBI is that it does not provide the semantic characterisation of collocations, which would enhance its value for translators focused on target text production. McGee (2012), however, claims that there is a kind of semantic classification since semantic classes of collocates are grouped together, though not explicitly labeled. Nevertheless, this classification only describes collocations that follow the pattern of verb + noun in CA (creation/activation) and EN (eradication/nullification). Other semantic areas are not explored.

2.2. Oxford Collocations Dictionary

The *Oxford Collocations Dictionary* was first published in 2002. A second edition was published in 2009 (McIntosh, Francis & Poole 2009), which contains 250,000 word combinations and about 9,000 headwords for nouns, verbs, and adjectives, along with 75,000 examples showing how collocations are used. Like the MCD, it is an English dictionary of collocations, aimed at upper-intermediate to advanced students of English for text production purposes.

Like the BBI, the OCD does not provide noun collocates for verb and adjective entries. For example for the verb 'study', the OCD provides adverb

collocates (e.g. 'carefully', 'closely', 'in depth'); phrases ('be easily'/'well studied'); and prepositions (e.g. 'for'). Nevertheless, no information is given regarding noun collocates.

The corpus used to extract the most salient combinations was the *British National Corpus,*[3] in the first edition, and the *Oxford English Corpus*[4] of 2 billion words in the most recent edition. When necessary, this edition also provides variants for British and American English. Figure 2 shows the entry in the OCD for 'bed'.

bed *noun*
1 piece of furniture for sleeping on
- ADJ. **double, king-size, single, twin | bunk, camp, feather, folding, four-poster, hospital, sofa | warm | unmade** *a messy room, with an unmade bed and clothes on the floor* | **marriage**
- VERB + BED **do** (*informal*), **make, make up | strip** *Please strip the beds and put the sheets in the washing machine.* | **climb into, crawl into, get into, go to, tumble into** *She crawled into bed exhausted.* | **climb out of, get out of, leap out of | lie (down) on, lie in, sit on** *He lay in bed, reading his book.* ◊ *Elisabeth was sitting on her bed writing a letter.* | **put sb to, tuck sb up in** *It's your turn to put the children to bed.* | **wet** *Don't punish a child who wets the bed.*
- BED + NOUN **clothes, linen**
- PREP. **in ~** *I like to be in bed before 11 o'clock.* **out of ~** *Are you out of bed yet?*
- PHRASES **bed and breakfast, the edge/side of the bed, the foot/head of the bead, get sb into bed** (=have sex with sb), **go to bed with sb** (=have sex with sb), **take to your bed** (=go to bed because you are ill), **time for bed** *Come on, children, it's time for bed.*

2. piece of ground for growing flowers, vegetables, etc.
- ADJ. **flower, rose, strawberry | ornamental | raised**
- PREP. **~of** *ornamental beds of roses*

Figure 2. Entry for 'bed' in OCD

In the OCD, when the headword is polysemic, as is the case for 'bed', a definition of each sense is provided. However, when the headword only has one sense, no definition is given. As shown in Figure 2, immediately after each definition, the associated collocations are displayed in bold in order to be easily identifiable from the rest of the text. Collocations are first organized in regards to their grammatical structure (e.g. Adj; Verb + *bed*; *bed* + Verb; Prep; Phrases).

3. < http://www.natcorp.ox.ac.uk/> [22/04/2013].
4. <http://oxforddictionaries.com/words/the-oxford-english-corpus> [22/04/2013].

The OCD is the only one of the three dictionaries described that devotes a special section to phrases. Authors of the MCD justify not including phrases by saying that the "OCD has a 'phrases' category where items are often included on grounds of non-transparency of meaning" (Kilgarriff et al. 2010: 373). Nevertheless, despite these minor inconsistencies, the OCD approach to phrase description is very useful. Phrases are frequent in general language and are difficult for non-native speakers to master. The idea of putting phrases in a separate section in a dictionary makes them easier for the user to understand.

Differences in meaning within the same grammatical relation are expressed in the OCD by means of a vertical line "|". As previously mentioned, McGee (2012:333) states that collocations within OCD entries are "semantically organized, but this organization is not actually labeled". However, in our opinion, the semantic organization in this dictionary is rather tenuous since the semantic relationship between the headword and the collocations is not stated, and thus must be deduced by users. For instance, regarding the collocates for the first sense in Figure 2, 'double', 'king-size', 'single', and 'twin' are separated from 'bunk', 'camp', 'feather', 'folding', 'four-poster', 'hospital' and 'sofa'. However, it is difficult for users to discern the relationship between 'feather bed' and 'sofa bed'. In addition, when necessary, the OCD provides explanations in brackets in regards to meaning and usage. Finally, it also includes usage examples in italics to illustrate how each collocation should be used in context. It goes without saying that in translation, contextual information is essential since the meaning of a word can vary considerably depending on its usage context. Evidently, because of space constraints, a lexicographic resource can only include a limited set of micro-contexts of a word but obviously, the usefulness of a dictionary for translators is in direct consonance with the amount of contextual information included.

2.3. MacMillan Collocations Dictionary

The *Macmillan Collocations Dictionary* (Rundell 2010) is the most recent dictionary of collocations in English though its compilation process started in the 1990s. Like the OCD, its objective is to help upper-intermediate to advanced English students improve their writing skills in order to pass the International English Language Testing System (IELTS). In the same way as the OCD and in contrast to the BBI, it is corpus-based. The corpus used is a two-billion word corpus, known as the *World English Corpus*.[5] Its novelty lies

5. For a complete description of the corpus, see <http://www.macmillandictionary.com/corpus.html> [12/04/2013].

in its methodology since it is the first dictionary to extract collocations automatically in the form of *word sketches* thanks to a new software tool, incorporated into the Sketch Engine corpus query system[6] (Kilgarriff et al. 2004). A *word sketch* is an automatic corpus-derived summary of a word's grammatical and collocational behavior. In fact, the OCD is a model for what was intended to be automatically produced by word sketches (Kilgarriff et al. 2010: 373):

> Our goal for what word sketches aim to do is provide a grammatically-organized list of collocates which would form a suitable entry for a collocations dictionary such as OCD.

As evidenced by Kilgarriff et al. (2010), word sketches offer reliable and valuable information, which, among other things, is extremely useful for text production in translation. These authors formally evaluate word sketches for four languages, namely, Dutch, English, Japanese, and Slovene. For this purpose, forty-two headwords were selected and twenty collocates for each headword were analysed in the four languages by asking users whether the collocation was suitable for inclusion in a published collocation dictionary. The evaluation showed that more than two thirds of the collocations were of publishable quality.

In line with the OCD, the MCD also has a very broad classification of collocations, as can be seen in the introduction of the dictionary:[7]

> Collocations are 'semi-preconstructed phrases' which allow language users to express their ideas with maximum clarity and economy. Not only that, there is a strong correlation between frequency in a corpus and typicality, which means that the use of common collocations contributes to the naturalness of a text.

The MCD contains about 4,500 entries. Its editors justify this relatively small number compared to other collocational resources such as the OCD by saying that they "prefer to give full coverage to a smaller number of words".[8] Even though the MCD contains fewer entries than the OCD, both dictionaries are similar in length (McGee 2012: 333):

> Although [the MCD] contains around half as many entries as the OCD – with around 4,500 keyword entries – the two dictionaries are of comparable length (MCD 911 pages, versus OCD 963 pages).

6. < http://www.sketchengine.co.uk/> [12/04/2013].
7. <http://www.macmillandictionaries.com/features/how-dictionaries-are-written/macmillan-collocations-dictionary/#1> [12/04/2013].
8. <http://www.macmillandictionaries.com/features/how-dictionaries-are-written/macmillan-collocations-dictionary/#1> [12/04/2013].

Figure 3 shows the entry for 'bed' as described in the MCD. For each headword, first the definition(s) is provided. In contrast to the OCD, in the MCD, the definition of each headword is always given, even when the headword only has one meaning. This means that the MCD can be used both for encoding as well as for decoding. In regards to translation, this allows the translator to find correspondences in the production phase as well as in the early reception stages.

bed N
a piece of furniture for sleeping on

adj +N types of bed **bunk, divan, double, folding/foldaway/fold-up/four-poster/4-poster, king-size/king-sized, single, sofa, twin** *If you sleep in a double bed up against a wall, ensure you sleep by the wall side.*
► ***describing*** *a bed* ***comfortable,*** **comfy** informal, **unmade** *The beds were comfortable and the kitchen had everything we could ever need.*

v+into+N **climb, collapse, crawl, fall, get, roll, sink, tumble** *We dumped our bags and just crawled into bed.*

v+*to*+N ***come, crawl, get, go, put sb, retire, send*** **sb** *He was glad to get to bed. Send your child to bed at a regular and reasonable time.*

Figure 3. Entry for 'bed' in the MCD

Collocations are subsequently provided after the definition of the headword, which is first categorized in terms of grammatical category. The MCD gives the grammatical relation between the headword and collocate by means of part of speech patterns (e.g. adj + N; v + *into* + N; v + *to* +N), but not explicitly like the OCD does (i.e. adj + *bed*). According to Fuertes-Olivera (2011: 59), it would have been preferable to have explicitly specified the grammatical labels for the sake of language students:

> My main contention is that the structural labels, i.e. the grammatical codes, are not explained, which hinders its usability in some teaching/learning situations, e.g. Spanish universities, in which students are unfamiliar with grammar information.

In our opinion, the absence of grammatical labels should not be a problem for advanced students of English. However, their explicit presentation makes lexicographic entries more transparent. As previously mentioned, unlike the OCD, the MCD does not put prepositions in a different section, but rather includes them in the constructions and examples.

The MCD expresses different ideas within the same grammatical category through semantic groupings with a definition headed by the symbol ►.

As shown in the grammatical category adj + n, two semantic groupings are provided: (1) *types of bed*, which include collocations such as 'bunk', 'divan', 'double', 'folding/foldaway/fold-up/four-poster/4-poster', 'king-size/king-sized', 'single', 'sofa', 'twin'; (2) *describing a bed*, with collocates such as 'comfortable', 'comfy', and 'unmade'. Its authors are very proud of the fact that they classify collocations according to meaning and describe the MCD as "the only fully corpus-based collocations dictionary which incorporates semantic groupings".[9]

Even though the underlying idea is very good, it is more viable from a grammatical or syntactic point of view than from a semantic perspective. This is evidenced, for instance, by the use of constructions described in terms of *v + into* or *v + to*, and the exclusion of basic collocations such as 'make' which could pose problems for students of English who must decide between 'do the bed' or 'make the bed'. Regarding the second meaning (describing a bed), only the collocations 'comfortable', 'comfy', and 'unmade' are given, even though a bed can also be 'hard', 'lumpy', or 'uncomfortable'. As shall be seen, a meaning-based classification of collocations is of paramount importance for certain user groups, such as translators.

The MCD also includes a section devoted to usage notes that provides the following information: (i) colligation, i.e. the tendency of a word to appear in a particular form (Hoey 2005) (e.g. a verb that is mainly used in its passive form; or a noun used primarily in its plural form); (ii) alternatives to collocations, which are common phrases that can be used instead of the collocation. Colligation information is displayed against a pink background, and possible alternatives to collocations against a grey one. The systematicity achieved by clearly stating what is included in each type of usage note and the color distinction gives users easy access to this information. Finally, like the OCD and the BBI, the MCD includes usage examples in italics to illustrate how each collocation should be used in context.

However, the most striking difference is the type of collocational information provided. In other words, it is the first dictionary to include noun collocates for adjectives and verb entries. Neither the BBI nor the OCD provides nouns for adjectives and verbs (McGee 2012: 334):

> The standard practice in collocation dictionary production until the publication of the MCD was to place the 'independent' or 'autonomous' base of

9. <http://www.macmillanenglish.com/courses/macmillan-collocations-dictionary/> [21/01/2013].

> the collocation (usually the noun) as an entry word, where one can find its 'dependent' collocates (verbs or adjectives).

This is indeed one of the most valuable and distinctive aspects of the MCD compared to the other English collocational dictionaries.

2.4. Redes. Diccionario combinatorio del español contemporáneo

Redes. Diccionario combinatorio del español contemporáneo, henceforth REDES (Bosque 2004), is the first combinatorial dictionary published for the Spanish language. It was developed by Ignacio Bosque and his team at the Complutense University of Madrid. The dictionary was elaborated from a corpus of 250 million words composed of texts from 68 Spanish and Latin American newspapers published from 1993 to 2003. This resource is mainly for linguists studying lexical restrictions in Spanish and, in general, for anyone interested in the use of the Spanish language. In 2006, another more practice-oriented dictionary appeared, which is based on the data contained in REDES: the *Diccionario combinatorio práctico del español contemporáneo,* henceforth PRÁCTICO (Bosque 2006) (see 2.5).

Apart from being the first dictionary to deal with word combinations in Spanish, the novelty of REDES lies in the fact that predicates are the target of this resource. This is in vivid contrast to the general tendency of traditional collocation dictionaries in other languages, where the noun (which normally corresponds to the base) is the primary focus of attention (Barrios 2007: 1):

> *Redes*, el primer diccionario combinatorio del español, obra innovadora en su enfoque pues en ella los predicados se convertían en núcleo y principal objeto de estudio.

Bosque prefers to refer to his dictionaries as *diccionarios combinatorios* [combinatorial dictionaries] in the same way as the BBI or the ECD. The reasons are the following (Barrios 2007: 2): (i) Bosque's dictionaries include something more than collocations, i.e. "fenómenos de combinatoria en el sentido amplio" [combinatory phenomena in the broadest sense]; (ii) the notion of dictionary of collocations lends itself easily to multiple interpretations (e.g. the entry for 'book' in the OCD includes 'good' and 'great', whereas in PRÁCTICO, neither 'buen libro' [good book] nor 'libro genial' [great book] is included). This is evidence that certain resources, commonly referred to as *dictionaries of collocations,* also include frequent combinations with the subsequent risk of draining the notion of *collocation* of its substance.

REDES contains 7,115 lemmas composed of nouns, adjectives, adverbs, verbs as well as noun, adjective, adverbial, and verb phrases. It has two types

of entry: (i) *entradas analíticas* [analytical entries] or *entradas largas* [long entries]; (ii) *entradas abreviadas* [abbreviated entries] or *entradas cortas* [short entries]. Generally speaking, long entries are for *selecting* base words, whereas short entries are for *selected* words or *collocates*, as stated by Bosque (2004: xxxviii):

> De manera muy simplificada, puede decirse que las palabras que aparecen en las entradas analíticas son, en la mayor parte de los casos, PALABRAS SELECCIONADORAS, mientras que las voces que aparecen las referencias cruzadas son PALABRAS SELECCIONADAS.

Since in REDES 'cama' [bed] is only included in the form of a short entry and not as a long one, it is thus not useful for our purposes. Consequently, to illustrate the type and organization of collocational information in REDES, we have randomly chosen the phrase 'a bombo y platillo' (Figure 4), which appears in REDES as a long entry. 'A bombo y platillo' means to spread far and wide, in reference to a piece of news or an event.

a bombo y platillo *loc.adv.* ■ Admite la variante, menos usada, *a bombo y platillos*. Se combina con…
A VERBOS QUE DENOTAN DIFUSIÓN O TRANSMISIÓN DE INFORMACIÓN: 1 anunciar++: …un congreso en el que se anunciaría a *bombo y platillo* que la familia nacionalista se iba a vivir a una misma casa. CAN080101 **2 proclamar ++:** …algo que la Comisión proclamó *a bombo y platillo* hace diez días pero que ahora no resulta tan claro. EME190895. **3 difundir ++**: Esta noticia fue difundida *a bombo y platillo* en una rueda de prensa…EME 200694 **4 propagar +:** …la prensa falangista lo habría propagado *a bombo y platillo*. LVE190395. **5 pregonar +:** …ha sido pregonado *a bombo y platillo* su rotundo éxito…LVE161095. **6 presentar:** …fue presentado *a bombo y platillo* a los medios de comunicación… ABC190595 **7 transmitir:** …produce rubor el recuerdo de la infame entrevista (…) transmitida *a bombo y platillo* por los telediarios…EPE131099 **8 publicar:** … para que fuesen publicadas *a bombo y platillo*. LVE020395 **9 airear:** …aireadas *a bombo y platillo* por asociaciones estadounidenses…EME240796 **10 promulgar:** …unos carteles enormes promulgan *a bombo y platillo* el «gran» esfuerzo del Ayuntamiento…EPE261201

B VERBOS QUE DESIGNAN OTRAS ACCIONES VERBALES, MÁS FRECUENTEMENTE SI EXPRESAN LA DE PONER ALGO DE MANIFIESTO: 11 afirmar +: Se afirma *«a bombo y platillo»* que el precio del dinero está en los niveles más bajos de su historia…EME120194

Figure 4. Extract from a long entry in REDES: 'a bombo y platillo'

In REDES, long entries can have a length of various pages. The most important aspect of long entries is that they are semantically divided into lexical classes, identified with capital letters (A, B, C, D, etc.). The descriptor of the

lexical class is the text that defines or characterizes it. This description is in capital letters so that users can easily identify the sense that they are looking for. In this regard, in Figure 4, two lexical classes are displayed: (i) A. *VERBOS QUE DENOTAN DIFUSIÓN O TRANSMISIÓN DE INFORMACIÓN* [verbs that denote diffusion or transmission of information]; (ii) B. *VERBOS QUE DESIGNAN OTRAS ACCIONES VERBALES*, MÁS FRECUENTEMENTE SI EXPRESAN LA DE PONER ALGO DE MANIFIESTO [verbs that designate other verbal actions, more specifically if they express the idea of making something evident].

A meaning-based classification and description of combinations is one of the most positive features of REDES. A semantic classification for collocations is necessary since as stated by Siepmann (2005: 424): "[…] dependencies exist not merely between lexical units, but also between semantic features".

Subsequently, the combinations activated within each lexical class are given in bold typeface headed by a number. However, the numbering does not start over each time a new lexical class is encountered since these numbers are subsequently used to automatically create short entries. The frequency of each combination is represented by means of the following symbols: "++" is a very frequent combination; "+" signifies a fairly frequent one; no symbol means that it is an acceptable combination for native speakers; "-" means that the combination is not very frequent though possible.

Documented examples showing the combination in context are provided, along with the reference that identifies the source date. An explanation of the labels used for documenting the examples is given in the introduction to the dictionary. For example, as shown in Figure 4, the example for the verb 'presentar', is …*fue presentado a bombo y platillo a los medios de comunicación*…, which is identified as ABC190595. This means that the example comes from the Spanish newspaper *ABC,* dated 19 May 1995. However, as stated by Bosque (2004: xlix), sometimes it is very difficult to retrieve a certain combination in a referenced source because it is not very frequent, even though it sounds natural to a native speaker. Should this happen, the authors create the examples themselves, labeling them as *undocumented* (INDOC) (Bosque 2004: xlix):

> Por muy amplio que sea el corpus con el que se trabaje, no es posible encontrar en él todas las combinaciones que correspondan a una clase léxica determinada.

At the end of long entries, two more specifications are provided: (i) a grey square with the phrase *se combina con* [combines with], which offers word

combinations that do not fit in the other lexical classes; (ii) cross-references to other entries by means of *véase también* [see also].

At this point, it is important to highlight that a lemma is either listed as a long entry or as a short entry, but not both. In short entries, the lexical classes of long entries disappear since these short entries were not written one by one, but rather automatically obtained from long entries. In other words, the computer extracted nouns, phrases, verbs, adjectives, and adverbs from the body of analytical entries, and converted them into lemmas. Neither grammatical information nor definitions (even when the lemma has various senses) are provided for short entries. The various senses within a lemma are distinguished by means of this symbol: ♦. In REDES, there are five types of short entries:

(i) *Referencias cruzadas a las voces* [cross-references to terms], which are lemmas for selected words, i.e. collocates or words that can combine with many other words;
(ii) *Referencias cruzadas a los conceptos* [cross-references to concepts], which refer to concepts rather than words;
(iii) *Entradas del índice conceptual* [entries of the conceptual index] that also designate concepts. In contrast to cross-references to concepts, which are indexes of the lexical classes of the corresponding analytical entries, these entries are indexes of the lemmas or words that have an entry in REDES. They are identifiable since they are displayed in small caps and italics;
(iv) *Series abreviadas* [abbreviated series], to which 'cama' [bed] belongs. These entries do not contain any superscript. An analytical entry is not proposed for these entries since, generally speaking, some of the most frequent combinations for these entries do not fit into the analytical entry. As shown in the example for the lemma 'cama' [bed] (Figure 5), only verbal collocations are listed. This entry does not include any kind of semantic information, examples, or more nouns or adjective phrases collocating with the headword. The collocational information provided for lemmas, however, is greatly enhanced in the dictionary PRÁCTICO.

cama ◆ acostar(se) (en), deshacer, guardar, hacer, levantar(se) (de), meter(se) (en), postrar(se) (en), quedar(se) (en), recostar(se) (en), sacar (de), salir (de), tender(se) (en)

Figure 5. Entry for 'cama' [bed] in REDES

(v) *Remisiones* [cross-references], which refer to both suggestions (*véase también*) [see also] or cross-references (*véase*) [see].

2.5. Diccionario combinatorio práctico del español contemporáneo

The *Diccionario combinatorio práctico del español contemporáneo* (Bosque 2006), henceforth PRÁCTICO, is a combinatory dictionary for Spanish derived from REDES. It is the practice-oriented version of REDES, conceived for text production (Barrios 2007: 1–2):

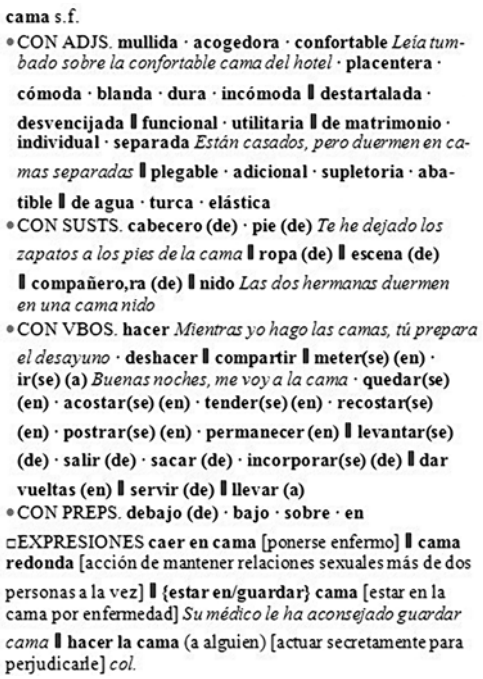
> Si *Redes* era un intento de reflexión acerca de la restricción léxica, *Práctico* pretende ayudar a encontrar la palabra adecuada, sumándose a la lista de trabajos onomasiológicos orientados a la producción y no a la comprensión […].

It thus targets teachers and students of Spanish as a foreign language, translators, and generally anyone that aims at speaking Spanish with native fluency. Although it is shorter than REDES (1,305 pages instead of 1,839), it contains almost twice the number of entries (14,000 in contrast to the nearly 8,000 of REDES), and about twice the number of word combinations (400,000 in contrast to 200,000). Conceptual entries are not included in PRÁCTICO because it is a less conceptual and more practical dictionary. This is indeed the reason why it has more word combinations and less descriptive information. PRÁCTICO has three types of entry:

(i) *simple entries*, which merge analytical and short entries in REDES (see Figure 6). The lemmas are displayed in bold font;
(ii) *generic entries*, which are not present in REDES. They group the words with the same combinatorial potential in the same semantic field in one entry so as to avoid repetition. For instance, since all the months of the year combine with a similar set of words, there is a generic entry 'MES' [month] that includes all the months of the year. The lemmas of generic entries appear in capital letters.
(iii) *cross-references entries*, which are enclosed in square brackets and are supplementary aids designed to facilitate information retrieval. For example, the lemma 'julio' [July] corresponds to a cross reference entry that refers the user to the entry for 'MES' [month], as previously above.

The content of simple entries in PRÁCTICO is very different from that in the short entries of REDES. This is the case of 'cama' [bed] as shown in Table 6, which corresponds to the entry for 'cama' [bed] in PRÁCTICO. When compared to the same lemma in REDES, most of the information in PRÁCTICO

is new. The microstructure of an entry in PRÁCTICO is similar to that of the OCD (see Table 6).



cama s.f.
• CON ADJS. **mullida · acogedora · confortable** *Leía tumbado sobre la confortable cama del hotel* **· placentera · cómoda · blanda · dura · incómoda ▮ destartalada · desvencijada ▮ funcional · utilitaria ▮ de matrimonio · individual · separada** *Están casados, pero duermen en camas separadas* **▮ plegable · adicional · supletoria · abatible ▮ de agua · turca · elástica**
• CON SUSTS. **cabecero (de) · pie (de)** *Te he dejado los zapatos a los pies de la cama* **▮ ropa (de) ▮ escena (de) ▮ compañero,ra (de) ▮ nido** *Las dos hermanas duermen en una cama nido*
• CON VBOS. **hacer** *Mientras yo hago las camas, tú prepara el desayuno* **· deshacer ▮ compartir ▮ meter(se) (en) · ir(se) (a)** *Buenas noches, me voy a la cama* **· quedar(se) (en) · acostar(se) (en) · tender(se) (en) · recostar(se) (en) · postrar(se) (en) · permanecer (en) ▮ levantar(se) (de) · salir (de) · sacar (de) · incorporar(se) (de) ▮ dar vueltas (en) ▮ servir (de) ▮ llevar (a)**
• CON PREPS. **debajo (de) · bajo · sobre · en**
□EXPRESIONES **caer en cama** [ponerse enfermo] **▮ cama redonda** [acción de mantener relaciones sexuales más de dos personas a la vez] **▮ {estar en/guardar} cama** [estar en la cama por enfermedad] *Su médico le ha aconsejado guardar cama* **▮ hacer la cama** (a alguien) [actuar secretamente para perjudicarle] *col.*

Figure 6. Entry for 'cama' in PRÁCTICO

In contrast to the analytical entries in REDES, entries in PRÁCTICO do not contain an *entradilla*, [an introductory description]. Like the OCD, in PRÁCTICO, when the headword has more than one sense, a brief definition of each sense is first provided in square brackets preceded by the symbol "▮". For 'cama' [bed] (Table 6), no description is provided since it only has one meaning.

As can be seen in Table 6, collocations in PRÁCTICO are first organized according to grammatical category. In this regard, PRÁCTICO distinguishes between combinations *CON ADJS.* [with adjectives], *CON SUSTS.* [with nouns], *CON VBOS.* [with verbs], *CON PREPS.* [with prepositions], and *EXPRESIONES* [phrases]. Like the OCD, it also devotes a special section to phrases, which is a completely new feature, especially considering that this type of information is not provided in REDES. However, this dictionary does not include a section for non-verb combinations. Users must thus deduce this

type of information, which might be a problem for non-native speakers of Spanish (Model 2008: 196):

> el diccionario carece de indicaciones explícitas, por lo cual el usuario tiene que detectar la función de un sustantivo a partir de otras indicaciones menos obvias. Pese a que se ha elaborado un sistema para revelar si un sustantivo es sujeto u objeto del verbo indicado, este sistema resulta algo enrevesado y no se aplica en todos los casos.

Like the OCD, collocations are grouped in terms of semantic proximity. The resulting groups are separated by "||". As shown in Table 6, 'meter(se) (en)', 'ir(se) (a)', 'quedar(se) (en)', 'acostar(se) (en)', 'tender(se) (en)', 'recostar(se) (en)', 'postrar(se) (en)', and 'permanecer (en)' are similar in meaning. However, only a native speaker of Spanish would realize that 'acostarse' is somewhat different from 'quedarse en'.

In our opinion, again, this type of classification is often confusing since the semantic relation between members is not explicitly labeled and is often difficult to infer. This is indeed one of its main drawbacks when compared to REDES. In PRÁCTICO, the semantic classes of REDES are not included. There is only an attempt to implicitly group collocates by semantic proximity as stressed by Barrios (2007: 7):

> Si el lector de Práctico es un lingüista y busca paradigmas de clases léxicas, no encontrará un análisis de los rasgos que permiten definir los subgrupos pero aunque los rasgos semánticos no se hagan explícitos permanecen implícitos en la forma de agrupar.

In addition, PRÁCTICO provides explanations in square brackets, when needed, concerning meaning and usage. Finally, it provides usage examples in italics to illustrate how the collocations should be used in context. It is important to emphasize here that the documented examples in REDES have been replaced by examples created by the authors themselves in order to help students of Spanish and also not to create an excessively large dictionary.

2.6. Diccionario de Colocaciones del Español

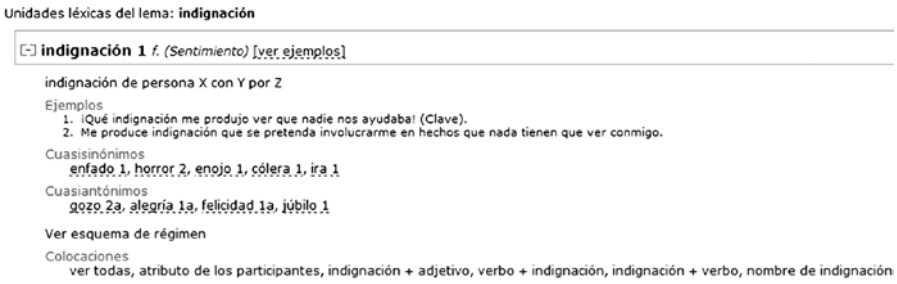
The *Diccionario de Colocaciones del Español,*[10] henceforth DICE (Vincze, Mosqueira & Alonso 2011), is an online collocations dictionary of Spanish, currently being developed at the University of A Coruña (Spain). To date its contents are restricted to the domain of 'feeling' and it only contains 211 lemmas, which are each associated with one or various lexical units (LUs). Therefore, the number of LUs described in the dictionary is larger. The website

10. <http://www.dicesp.com> [16/05/2013].

provides two access points: (1) to the dictionary itself; (2) to an advanced search option consultation. An additional didactic model is currently being implemented. This last option verifies whether a certain combination is correct and permits the following two kinds of consultation: (i) users can ascertain whether a specific base combines with a specific collocate; (ii) users can obtain the collocates for a given base when wishing to express a specific meaning.

As in most of the collocation dictionaries in this section, the DICE focuses on encoding and targets learners of Spanish (Alonso 2010). Like the BBI, REDES, and PRÁCTICO, it is based on the premises of the ECD. However, it differs from them in that it is the only one that describes the semantic content of collocations in terms of lexical functions.

As previously mentioned, the DICE only focuses on nouns of feeling, and thus does not include 'cama' [bed]. Therefore, the headword 'indignación' [indignation] is used as an example to describe the microstructure of an entry in the DICE (see Figure 7).


Unidades léxicas del lema: **indignación**

indignación 1 *f. (Sentimiento)* [ver ejemplos]

indignación de persona X con Y por Z

Ejemplos
1. ¡Qué indignación me produjo ver que nadie nos ayudaba! (Clave).
2. Me produce indignación que se pretenda involucrarme en hechos que nada tienen que ver conmigo.

Cuasisinónimos
enfado 1, horror 2, enojo 1, cólera 1, ira 1

Cuasiantónimos
gozo 2a, alegría 1a, felicidad 1a, júbilo 1

Ver esquema de régimen

Colocaciones
ver todas, atributo de los participantes, indignación + adjetivo, verbo + indignación, indignación + verbo, nombre de indignación

Figure 7. Entry for the LU 'indignación 1' [indignation] in the DICE

As shown in Figure 7, the information provided for each LU in the DICE is the following: (1) a semantic tag, which is the general meaning of the LU (in this case, the semantic tag is *sentimiento* [feeling]); (2) its actantial structure, which lists the participants in the situation designated by the LU (i.e. *indignación de persona X con Y por Z* [indignation of person X with Y because of Z]); (3) usage examples, mainly derived from the *Corpus de Referencia del Español Actual* (CREA)[11] as well as from the corpus *LexEsp* (Sebastián et al. 2000), the *Corpus del Español*,[12] directly from the web, and from other lexicographical

11. <http://corpus.rae.es/creanet.html> [16/05/2013].
12. <http://www.corpusdelespanol.org> [16/05/2013].

works; (4) quasi-synonyms and quasi-antonyms of the LU; (5) the syntactic schema (*esquema de régimen*), which is displayed in another window with syntactic information regarding the linguistic realizations of actants (e.g. the participant Y with 'indignación' can be linguistically headed by the prepositions 'con', 'contra', 'ante' and 'hacia' (e.g. 'La indignación con/contra/ante/hacia el gobierno'); (6) collocations.

As shown in Figure 1, the collocations for each LU are classified in five groups: (i) *atributo de los participantes* [attributes of the participants], (e.g. 'en un momento de indignación'); (ii) *indignación + adjetivo* [LU + adjective] (e.g. 'indignación popular'); (iii) *verbo + indignación* [verb + LU] (e.g. 'aumentar', 'acrecentar', 'alimentar'); (iv) *indignación + verbo* [LU + verb] (e.g. 'aplacarse'); (v) *nombre de indignación* [noun *de* + LU], which lists all collocations headed by the combination of a noun + *de* (e.g. 'sentimiento de indignación').

By clicking on each collocation group, the system displays a list of glosses or brief linguistic descriptions that are characteristic of a specific meaning (Alonso, Nishikawa & Vincze 2010: 370-371) (see Table 1). In other words, they are linguistic adaptations of LFs. Table 1 displays an extract of the glosses for the construction *verb + indignación*. As shown, the LF may also appear optionally. In addition, the number of collocates that are part of the meaning of each gloss is specified. For example, as shown in Table 1, for the gloss 'causar que la ~ sea mayor' [to cause the indignation to become greater], the associated LF is *Caus Pred Plus*. Also given are three collocates in consonance with this semantic description.

[+] **sentir ~** Oper1 (3 valores en total)
[+] **sentir una ~ intensa** Magn + Oper1 (3 valores en total)
[+] **continuar sintiendo ~** Cont Oper1 (1 valor en total)
[+] **hacer a algo objeto de ~** Labor12 (1 valor en total)
[+] **causar que la ~ sea intensa** Caus Pred Magn (1 valor en total)
[+] **causar que la ~ sea mayor** Caus Pred Plus (3 valores en total)

Table 1. Extract of the glosses for the group of collocations *verb + indignación* in the DICE

By clicking on the plus symbol at the left of each gloss, the system displays the collocates that convey the meaning of the gloss. Accordingly, as shown in Figure 8, for 'causar que la ~ sea mayor' [to cause the indignation to become greater], the verbs 'aumentar', 'acrecentar' and 'alimentar' are encountered. This does not mean, however, that the verbs are synonyms, but rather that they have a similar meaning.

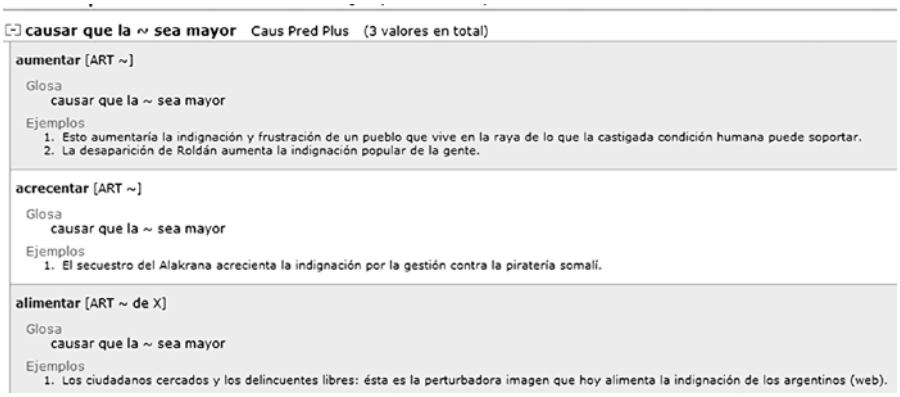


Figure 8. Collocations of the gloss 'causar que la indignación sea mayor' [to cause the indignation to become greater], for the LU 'indignación' in the DICE

As can be observed, the DICE includes the following information for each collocate: (i) the syntactic schema of the collocation in square brackets (e.g. for the verb 'alimentar' the schema is [ART ~ *de X*], which means that 'alimentar' comes in a construction headed by an article and followed by the preposition 'de' and an actant X); (ii) the gloss; (iii) usage examples extracted from the corpus (e.g. 'Los ciudadanos cercados y los delincuentes libres: ésta es la perturbadora imagen que hoy alimenta la indignación de los argentinos'). At times, certain entries also include expressions that despite not being collocations, express the same meaning as the collocation.

As previously mentioned, the DICE also has an advanced search component that permits users to find answers to specific questions. The system allows three types of searches: (1) *directas* [direct searches]; (2) *inversas* [indirect searches]; (3) *ayuda a la redacción* [help in writing]. Direct searches permit users to find the collocates of a base by means of an LF. Nevertheless, the system only allows users to search for one LF at a time. Additionally, they can specify the lemma and the LU.

For example, as shown in Figure 9, a direct search could provide an answer to the question of which verb can combine with 'indignación' [indignation] to express the meaning of the lexical function *Caus*. For this query, the system displays 30 collocations organized in terms of the LFs underlying their meaning. 'Indignación'[indignation] is found within the LF *Caus Pred Plus*. In our opinion, the idea of providing a query system is very helpful for final users. Nonetheless, the problem of direct searches in the DICE lies in the fact that meaning can only be expressed in terms of LFs. This rather difficult

metalanguage makes it unsuitable for the majority of users, except for linguists. Paradoxically, however, the intended users of the dictionary are learners of Spanish: "tomamos como punto de referencia al usuario de diccionario, especialmente al aprendiz de español" (Alonso 2010: 55).

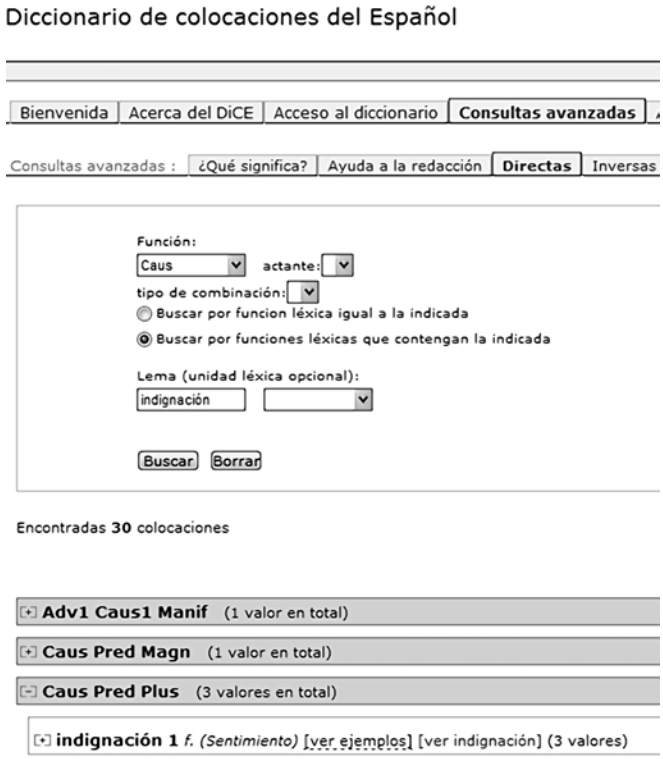


Figure 9. Direct search option in the DICE

The indirect search option (Figure 10), allows for two types of queries. The first one permits users to find the base of a collocation from the collocate (referred to as *valor* [value] in the DICE). Additionally, the LF can also be specified. For example, as shown in Figure 10, it can be used to find out which lemma is selected by the collocate 'alimentar'[to feed] to express the meaning of the LF *Caus*. Once again, the problem is the metalanguage of the LFs. It is true that users can launch a query without having to specify the LF, but in that case, an excessively large number of records is displayed.

Figure 10. Indirect search option in the DICE

The second query that can be launched with the indirect search option permits users to find the bases associated with a specific LF. This option is "oriented towards comprehension" (Alonso, Nishikawa & Vincze 2010: 372), and thus is for decoding purposes. Once again, the queries in the DICE do not seem to respond to the initial linguistic production aims of the authors of the dictionary, which according to Alonso (2010: 65), are oriented towards linguistic production.

In our opinion, this kind of search option is not very useful. Evidently, if users have doubts about the meaning of a word, they would consult a monolingual dictionary, which is easier, and less time-consuming than obtaining the meaning of a word through its LF.

Finally, the *help in writing* option verifies whether a specific combination is correct. It allows the following two kinds of consultations: (i) users can ascertain whether a specific base combines with a specific collocate; (ii) users

can obtain the collocates for a given base when wishing to express a specific meaning, by means of the gloss. For example, as shown in Figure 11, it permits users to launch the query and obtain all the verbs that can combine with 'indignación' [indignation], when it is used as a direct object with the meaning 'causar que la indignación sea mayor' [to cause the indignation to be greater]. For this consultation, the system displays the three verbs: 'acrecentar', 'alimentar', and 'aumentar'.

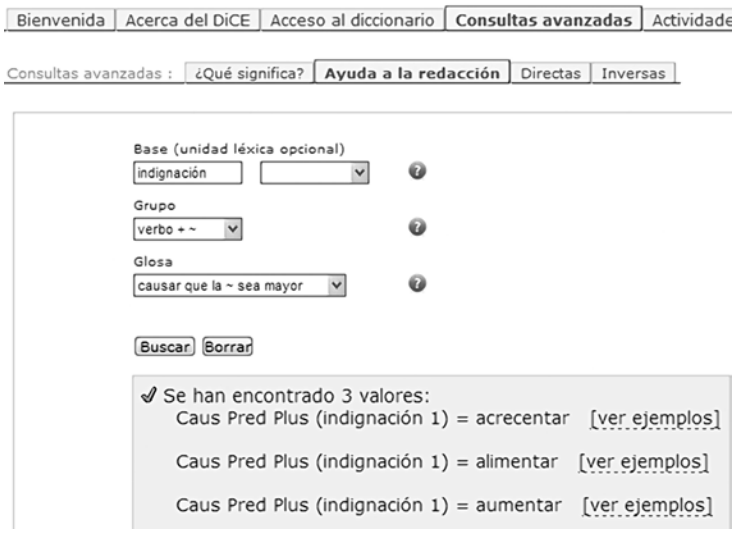


Figure 11. *Ayudas a la redacción* search option in the DICE.

Of the three types of search in the DICE, the option of *ayudas a la redacción* seems to be the only one that is accessible to all users, and which is of great help for encoding.

Succinctly put, the DICE is the only one of the six dictionaries with free online access. The fact that it is available online means that collocations can be easily accessed by users. It also allows for various types of queries to access collocations, which is another of its major advantages. However, as already mentioned, the metalanguage used makes it unsuitable for a layperson since too much linguistic knowledge is assumed. Another of its problems is its limited number of lemmas in comparison to the other resources.

3. Comparative analysis of the collocation dictionaries

In this section, we offer a comparative analysis of all the resources described in Section 2, regarding (i) the types of collocation to be encoded; (ii) the kinds of collocational information offered; (iii) the place for collocations in the micro or macrostructure of the dictionary.

With the exception of the DICE and the MCD, the dictionaries analyzed only provide one way of accessing collocations. This is done either by the base or the collocate itself. In contrast, the DICE allows users to access collocations in various ways. The search engine allows users to enter one of the following: (i) base term; (ii), collocate; (iii) lexical function. A very positive aspect of the MCD is that it is the only dictionary that includes nouns and adjectives in verb entries. Consequently, users can access collocations either by the base term or the collocate.

Regarding the classification of collocations within an entry, only the short entries within REDES do not offer any type of classification, whereas the other resources do. All of the dictionaries, with the exception of REDES and PRÁCTICO, classify collocations by syntactic schema though the organization and formulation of this information is far from homogeneous. The BBI systematically classifies collocations within the microstructure of each entry, but does not explicitly specify this organization. However, the order followed is described in the introduction of the dictionary: verb + noun (CA collocations); verb + noun (EN collocations); adj + noun; noun + verb; and noun + noun.

In contrast, the OCD explicitly specifies the grammatical category and order of the two components (i.e., for the lemma *bed,* adj, verb + *bed*; *bed* + verb; prep; phrases). The MCD also provides this information but without indicating the headword. This means that instead of verb + *bed*, it says verb + noun. The order followed is adj + n; v + *into* + n; v + *to* + n. The DICE first includes attribute combinations, followed by n + adjective; verb + noun; noun + verb; and noun *of* + noun.

PRÁCTICO also classifies collocations by part of speech. However, the syntactic schema between the noun and verb is not expressed. The only resource that gives a meaning-based classification of collocation is REDES in its long entries. It is true that the OCD and MCD claim to differentiate meaning within the same grammatical category. However, their attempt is somewhat less than successful since the relationship between the various components in the group is not expressed, and there are inconsistencies in the members of each group.

	Access	Classification	Description	Advantages	Disadvantages
BBI	Base term	Syntactic (not explicit)	- Explanation, when needed - Examples of use - Usage notes - Distinction between British and American English	- Theoretically based - It is also a kind of syntactic dictionary	- Only access by the base term - Paper format - Not classified according to meaning [only CA collocations (creation/activation) and EN collocations (eradication/nullification)]
OCD	Base term	- Syntactic - Semantic	- Explanations when needed - Examples of use	- Separate section for prepositions and phrases. - Attempt to classify collocations by meaning	- Only access by the base term - Paper format - Not theoretically based
MCD	- Base term - Collocate	- Syntactic - Semantic	- Usage notes - Examples of use	- Encoding and decoding purposes - It is the only dictionary that include noun collocates for adjective and verb entries. - Attempt to classify collocations by meaning - Fine-grained description of combinations	- Only access by the base term - Paper format - Less number of entries
REDES	Collocate	- Long entries: meaning - Short entries: no classification	- Long entries: Frequency of use Documented examples - Short entries: no description provided	- Theoretically based - Classification according to meaning (long entries) - Fine-grained description of long entries.	- Paper format - Some entries excessively long
PRÁCTICO	Collocate	Part of speech	- Explanations, when needed - Examples of use	- Theoretically based - Special section for phrases	- Paper format - Only access by the base term - Not syntactic relation expressed between noun and verb
DICE	- Base term - Collocate	Syntactic	- Syntactic behaviour - The gloss: i.e. meaning - Examples of use	- Free online format - Theoretically based - Various ways of accessing collocations (through the base and the collocate) - Meaning provided - Multiple ways of accessing collocations and searching	- Difficult access due to metalanguage - Lack of user friendliness for general users - Limited number of hits: only focuses on the domain of feeling

Table 2. Comparative analysis of all the collocation dictionaries

Concerning the description of collocations, the vast majority of resources provide explanations of the meaning of collocations. When necessary, usage notes and usage examples are often provided. However, the BBI is the only resource that differentiates American English from British English, whereas REDES is the only one that indicates the frequency of each combination. Finally the DICE has the advantage of offering a gloss for each combination. Table 2 summarizes the analysis in this paper, and highlights the advantages and limitations of each resource.

5. Conclusions

Thanks to the availability of large corpora and lexical analysis tools, a wide range of new lexicographic resources has emerged. Such resources include an important number of combinatory or collocation dictionaries, which target different user groups, such as language learners, linguists, teachers, translators, *inter alia*. However, as shown in our analysis, these resources differ in the way that phraseological information is represented and organized in their entries. This type of variation reflects the fact that there is still no general agreement as to which kinds of word combination should be included in dictionaries and how they should be described and classified.

The comparative analysis of the most representative monolingual collocation dictionaries in English and Spanish undertaken in this paper lead to the following conclusions:

- A collocation dictionary should provide various ways of accessing collocations in order to enhance the effective retrieval of relevant meaning-related information. This is crucial for translators who must find a way to produce a target text with the same meaning as the source text.
- A collocation dictionary should give some kind of classification of collocations within an entry. This way, translators, as well as other user groups, will be able to find the information that they are looking for quickly. This classification can be done in one of the following ways: (i) by part of speech; (ii) by the syntactic schema between the noun and the verb; (iii) by meaning. Evidently, for translation purposes, the most interesting classification would be a meaning-based classification similar to the one used in the long entries in REDES.
- A collocation dictionary should provide a description of collocations so that users are better able to understand the meaning of a specific collocation. This is imperative for translators who must work within the context of strict time constraints.

– Usage notes and examples of use should also be included in dictionary entries because they are a reflection of context. Contextual information is extremely valuable because it shows how the word behaves in a specific communicative setting and also exemplifies how a collocation is used in real language.

Acknowledgements

This research was carried out within the framework of the project RECORD: Representación del Conocimiento en Redes Dinámicas [Knowledge Representation in Dynamic Networks, FFI2011-22397], funded by the Spanish Ministry of Science and Innovation.

References

ALONSO, Margarita (2010) "No importa si la llamas o no colocación, descríbela." In: Mellado, Carmen; Patricia Buján; Claudia Herrero; Nely Iglesias & Ana Mansilla (eds.) 2010. *La fraseología del s. XXI: nuevas propuestas para el español y el alemán*. Berlin: Frank & Timme, pp. 55-80. Electronic version available at: <http://www.dicesp.com/app/webroot/files/file/Alonso 2010(1).pdf>. [05/07/2011].

ALONSO, Margarita; Alfonso Nishikawa & Orsolya Vincze. (2010) "DICE in the Web: An Online Spanish Collocation Dictionary." In: Granger, Sylvane & Magali Paquot (eds.) 2010. *eLexicograpy in the 21st Century: New Challenges, New Applications. Proceedings of eLex 2009*. Louvain-la-Neuve: Cahiers du Cental 7, Presses universitaires de Louvain, pp. 369-374. Electronic version available at: <http://www.dicesp.com/app/webroot/files/file/eLex 2009 DiCE in the Web.pdf>. [25/02/2011].

BARRIOS, María A. (2007) "Diccionarios combinatorios del español: diferencias y semejanzas entre Redes y Práctico." *redELE. Revista Electrónica de Didáctica del Español como Lengua Extranjera* 11, pp. 1-14. Electronic version available at: <http://www.mecd.gob.es/dctm/redele/Material-RedEle/Revista/2007_11/2007_redELE_11_01Barrios.pdf?documentId=0901e72b80df2cb7>. [05/01/2011].

BENSON, Morton; Evelyn Benson & Robert Ilson. (2009) *The BBI Combinatory Dictionary of English*. 3rd edition. Amsterdam/Philadelphia: John Benjamins.

BERGENHOLTZ, Henning & Sven Tarp. (2010) "LSP Lexicography or Terminography? The Lexicographer's Point of View." In: Fuertes-Olivera, Pedro A. (ed.) 2010. *Specialized Dictionaries for Learners*. Berlin/New York: De Gruyter, pp. 27-37.

BOSQUE, Ignacio. (2004) *Redes. Diccionario combinatorio del español contemporáneo*. Madrid: S.M.

BOSQUE, Ignacio. (2006) *Diccionario combinatorio práctico del español contemporáneo*. Madrid: S.M.

COWIE, Anthony Paul (1998) "Phraseological Dictionaries: Some East-West Comparisons." In: Cowie, Anthony Paul (ed.) 1998. *Phraseology. Theory, Analysis, and Applications*. Oxford: Clarendon Press, pp. 209-228.

KILGARRIFF, Adam; Pavel Rychly; Pavel Smrz & David Tugwell. (2004) "The Sketch Engine." In: *Proceedings of the 11th EURALEX International Congress*. Lorient: EURALEX, pp. 105-116.

KILGARRIFF, Adam; Vojtěch Kovář; Simon Krek; Irena Srdanovic & Carole Tiberius. (2010) "A Quantitative Evaluation of Word Sketches." In: *Proceedings of the XIV Euralex International Congress*. Leeuwarden: Fryske Academy, pp. 372-379. Electronic version available at: <http://www.euralex.org/elx_proceedings/Euralex2010/019_Euralex_2010_1_KILGARRIFF KOVAR KREK SRDANOVIC TIBERIUS_A Quantitative Evaluation of Word Sketches.pdf>. [03/07/2012].

L'HOMME, Marie Claude & Patrick Leroyer. (2009) "Combining the Semantics of Collocations with Situation-driven Search Paths in Specialized Dictionaries." *Terminology* 15:2, pp. 258-283.

MCGEE, Iain. (2012) "Collocation Dictionaries as Inductive Learning Resources in Data-driven Learning: an Analysis and Evaluation." *International Journal of Lexicography* 25:3, pp. 319-361.

MCINTOSH, C.; B. Francis & R. Poole. (2009) *Oxford Collocations Dictionary for Students of English*. 2nd edition. Oxford: Oxford University Press.

MEL'ČUK Igor A. et al. (1984-1999) *Dictionnaire explicatif et combinatoire du français contemporain. Recherche lexico-sémantiques I, II, III, IV*. Montréal: Les Presses de l'Université de Montréal.

MEL'ČUK, Igor A. (1996) "A Tool for the Description of Lexical Relations in the Lexicon." In: Wanne, L. (ed.) 1996. *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam/ Philadelphia: John Benjamins, pp. 37-71.

MEL'ČUK, Igor A.; André Clas & Alain Polguère. (1995) *Introduction à la lexicologie explicative et combinatoire*. Lovain-la-Neueve: Duculot/Aupelf-UREF.

MODEL, Benedikt A. (2008) "Review of the Book BOSQUE, Ignacio. (2006) *Diccionario combinatorio práctico del español contemporáneo*. Madrid: S.M." *Revista de Lexicografia* 12, pp. 196. Electronic version available at: <http://ruc.udc.es/dspace/bitstream/2183/5520/2/RL_13_REC-2.pdf.txt>. [24/01/2013].

NUCCORINI, Stefania. (2003) "Towards an Ideal Dictionary of Collocations." In: Sterkenburg, Piet Van (ed.) 2003. *A Practical Guide to Lexicography*. Amsterdam/ Philadelphia: John Benjamins, pp. 366-387.

RUNDELL, Michael. (2010) *Macmillan Collocations Dictionary for Learners of English*. Oxford: Macmillan Publishers Ltd.

SEBASTIÁN, Nuria; Mª Antonia Martí; Manuel Carreiras & Fernando Cuetos. (2000) *LEXESP. Léxico informatizado del español*. Barcelona: Ediciones de la Universitat de Barcelona.

SIEPMANN, Dirk. (2005) "Collocation, Colligation and Encoding Dictionaries. Part I: Lexicological Aspects." *International Journal of Lexicography* 18:4, pp. 409-443.

STUBBS, Michael. (2002) *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell Publishers.

VINCZE, Orsolya; Estela Mosqueira & Margarita Alonso. (2011) "An Online Collocation Dictionary of Spanish." In: Boguslavsky, Igor & Leo Wanner (eds.) 2011. *Proceedings of the 5th International Conference on Meaning-Text Theory*. Barcelona, pp. 275-286. Electronic version available at: <http://www.dicesp.com/app/webroot/files/file/MTT 2011 Vincze et al_.pdf>. [22/02/2013].

BIONOTES / NOTAS BIOGRÁFICAS

Miriam Buendía Castro holds a PhD in Translation and Interpreting from the University of Granada. Currently, she is a researcher at the Department of Translation and Interpreting at the University of Granada. She has published in prestigious international journals, such as *Terminology*. She has presented her work in international conferences, such as EuraLex or LREc. She has done two research stays in the Centrum voor Vaktaal en Communicatie, at the Erasmushogeschool (Brussels), under the supervision of Professor Rita Temmerman. Her main research interests are terminology, phraseology, and corpus linguistics.

Miriam Buendía Castro es doctora en Traducción e Interpretación por la Universidad de Granada. En la actualidad es investigadora en el Departamento de Traducción e Interpretación en la Universidad de Granada. Tiene publicaciones en revistas de impacto como *Terminology*. Ha presentado su trabajo en conferencias internacionales como son EuraLex o LREC. Ha realizado dos estancias de investigación en el Centrum voor Vaktaal en Communicatie, en el Erasmushogeschool (Bruselas), bajo la supervisión de la profesora Rita Temmerman. Sus principales áreas de interés son la terminología, la fraseología y la lingüística de corpus.

Pamela Faber lectures and works in terminology, translation, lexical semantics, and cognitive linguistics. She holds degrees from the University of North

Carolina at Chapel Hill, the University of Paris IV, and the University of Granada, where she has been a full professor in Translation and Interpreting since 2001. She is the director of the LexiCon research group, with whom she has carried out various research projects on terminological knowledge bases, ontologies, and cognitive semantics. One of the results of these projects and the practical application of her Frame-based Terminology Theory is EcoLexicon (ecolexicon.ugr.es), a terminological knowledge base on environmental science. She has published close to 100 articles, book chapters, and books, and has been invited to present her research in universities in Madrid, Barcelona, Leipzig, Brussels, Zagreb, Mexico D.F., Lodz, and Strasbourg, among other places. She serves on the editorial and scientific boards of several journals, such *as Fachsprache, Language Design, Terminology,* and the *International Journal of Lexicography.* She is also a member of the AENOR standardization committee.

Pamela Faber es docente y trabaja en terminología, traducción, semántica léxica y lingüística cognitiva. Es titulada por la Universidad de Carolina del Norte, la Universidad de París IV, y la Universidad de Granada, donde es catedrática en Traducción e Interpretación desde 2001. Es directora del grupo de investigación LexiCon, en el que ha dirigido varios proyectos de investigación acerca de bases del conocimiento, ontologías y semántica cognitiva. Uno de los resultados de estos proyectos y de la aplicación práctica de su teoría de la Terminología basada en Marcos es *EcoLexicon* (ecolexicon.ugr.es), una base de conocimiento terminológica del medio ambiente. Ha publicado cerca de 100 artículos, capítulos de libro y libros, y ha sido invitada a presentar su investigación en universidades de Madrid, Barcelona, Leipzig, Bruselas, Zagreb, México D.F, Lodz o Estrasburgo, entre otros. Forma parte del comité de revisión de revistas como *Fachsprache, Language Design, Terminology,* y de *International Journal of Lexicography.* Además es miembro del comité de estandarización de AENOR.