

Visual networks as a means of representing collocational information in electronic dictionaries

Sergi Torner: *Universitat Pompeu Fabra (sergi.torner@upf.edu)*

Blanca Arias-Badia: *Universitat Autònoma de Barcelona (Blanca.Arias@uab.cat)*

Abstract

The digital turn in lexicography has paved the way to new techniques for presenting information in dictionaries. This paper explores the potential of collocational networks (Phillips 1983, Williams 1998) as key lexicographical tools to represent information about lexical combinatorics. Collocational networks, which had initially been used as a means of identifying collocations in large corpora, must now be rethought in order to assist non-expert dictionary users in text production by allowing them to find the exact collocates they are looking for, as well as by offering the grammatical information needed to use collocations accurately. We advocate improving networks by incorporating information visualization techniques (Ware 2008, Pham 2012). Specifically, we suggest a number of measures which may be taken to both simplify access to the information provided by raw output from corpora—much of which may be noise for the dictionary user—and to enrich such collocational data by means of visually-explained relevant grammatical information.

Key words

collocation, lexical combinatorics, collocational network, online lexicography, information visualization

1. Introduction

Collocational networks were originally conceived mainly as a methodology for identifying collocations in corpora, since they are assumed to establish a network of relationships between lexical units. The main idea underlying this notion is that discourse, either specialized or general, is organized into collocational patterns, which can be represented as a network of words that collocate with one another. This approach was originally proposed by Phillips (1983) and later developed in several studies (Williams 1998, Alonso et al. 2011, Brezina et al. 2015, among many others). Some recent studies in visualization techniques have proposed that this kind of graphic representation may be a useful resource not only for information retrieval, but also in presenting lexical relations in dictionaries, in corpora or in other reference materials. Thus, for example, Alonso et al. (2011) suggest using collocational networks as a tool for navigating entries in a science dictionary; likewise, Masucci and Rogers (2006) use networks to represent lexical relations in Orwell's *1984*. McEnery (2006) uses collocational networks as a visual technique to show collocational relations in a general English corpus, and Jhang and Lee (2013) use a similar tool to represent collocational patterns in maritime English keywords; additionally, Mouri et al. (2014) use networks to explore common aspects of the learning process followed by learners of Japanese as a second language.

The main advantage of this kind of tool is that it makes an easily-readable representation of complex lexical relations possible, avoiding the use of metalinguistic apparatus which can be difficult to manage for non-expert users. In fact, visual representation techniques of complex information are a general device in information retrieval (Spence 2001, Ware 2008), i.e., they

are not exclusive to applied linguistics. In this article, we explore the potential of using collocational networks as a tool for representing collocational information in an online dictionary. In our proposal, we focus on Spanish examples, assuming that the very nature of collocational networks as visual tools makes the proposed model applicable to further languages. The kinds of challenges encountered when facing the problem of representing collocational information with networks are common to any language. Some particular aspects of our proposal might be especially relevant for Romance languages sharing grammatical features such as gender inflection, while others are fully transferable to a wide range of languages.

Information on collocations is crucial for language learning purposes: several studies have demonstrated that mastering lexical combinatorics fosters naturalness and fluency in language production (Meunier and Granger 2008, Wray 2002). In his account of productive priming, Hoey (2005: 11-17) also argued that awareness of the way in which a word is typically primed for native speakers is key for learners. Research on collocation learning has drawn attention to the fact that second language learners have difficulties in producing correct word combinations (Bahns and Eldaw 1993, Granger 1998, Howarth 1998, Laufer 2003, Nesselhauf 2003, 2005). Likewise, many studies have focused on the problems of using dictionaries to learn collocations in L2 learning contexts (Laufer 2011, Walker 2009), which has paved the way for dictionaries designed specifically to train collocational competence (cf. Spina 2010). Therefore, our proposal is mainly oriented towards learner dictionaries, although it may be useful for general or collocational dictionaries as well.

The proposal is based on the results of the research project ‘La representación de la combinatoria léxica en los diccionarios de aprendizaje: nuevos métodos para nuevos diccionarios’ (FFI2012-37654; PI: Sergi Torner), which was implemented between 2013 and 2016 and funded by the Spanish Ministry of Economy and Competitiveness. This project’s aim was to develop a proposal for representing lexical combinatorics as collocational networks. Although the project was primarily oriented towards lexicography for foreign language learning purposes, we think that the project’s output could be of interest for both the lexicography and the visualisation community. The study focused on a limited portion of the Spanish lexicon, specifically, the lexical units from four conceptual domains in the Instituto Cervantes’ plan of studies, the *Plan curricular*, a key reference in the field of Spanish as a foreign language teaching and learning, namely (a) family relationships, including items such as *cuñado* (‘brother-in-law’), *suegro* (‘father-in-law’), *pareja de hecho* (‘common-law partner’); (b) social relationships, e.g. *colega* (‘workmate’), *amigo íntimo* (‘close friend’), *relacionarse* (‘to be in contact’); (c) information and communication, e.g., *noticia* (‘news’), *comentar* (‘to comment’), *declarar* (‘to declare’); and (d) verbal communication, e.g. *explicación* (‘explanation’), *sugerir* (‘suggest’), *confirmar* (‘to confirm’).

As will be seen, the examples in the present paper have been taken from previous lexicographic work on these four notional domains. In the framework of the aforementioned project, we built a database with productive lexical combinatorics for a learner of Spanish as a foreign language at the B2-C1 level. The Instituto Cervantes’ *Plan curricular* provided the initial vocabulary lists for the project, which were complemented with Spanish as a Foreign Language pedagogical materials, corpus analysis tools (Sketch Engine®), and lexicographical sources. As regards the latter, in particular, we consulted the units included in the ‘Catálogo de voces afines’ in the *Diccionario de uso del español* compiled by María Moliner, a list of words related in some way by meaning to the lemma, as well as *REDES* and *Práctico*, Spanish collocational dictionaries. In total, we compiled 707 lexicographic spreadsheets, each of which contains data to build one collocational network, for 534 different lemmas.

The proposal may be implemented in different types of electronic dictionaries (general monolingual dictionaries, collocational dictionaries or learner's dictionaries) although we largely focus on the benefits of our guidelines in works designed for L2 learners, as collocational competence has been argued to be a difficult area in language learning, as explained above.

Our proposal builds on previous accounts of network representations for lexical relations. However, we intend to make several improvements on the networks that have been previously used in the specialized literature. We start from the premise that, when used as an information representation tool in dictionaries designed for non-specialist users, networks can facilitate access to lexical combinatorics and enrich dictionary content, thus better meet potential user needs. On the one hand, a certain degree of simplification is needed to make information easier to find. As the potential number of collocations for most words of any given language may be very high, the resulting network that represents their collocational patterns may result in a large schema in which excess information hinders the user from finding a specific word in the network. Consequently, we propose reducing the number of boxes in the network, while preserving the number of collocations included. The proposed simplification is based on semantic criteria. To this end, grouping collocates with a close semantic content or labelling clusters of collocates may allow quicker navigation across networks. For instance, in Spanish we can use either *mostrar* ('show'), *expresar* ('express'), *demostrar* ('demonstrate') and *manifestar* ('exhibit') in combination with *amor* ('love') to express the idea of 'showing love'; as a consequence, the user might benefit from finding these four verbs grouped together, rather than scattered over unrelated nodes in the network.

On the other hand, we propose enriching networks with different kinds of data which provide relevant information about how to *use* the reported collocations in text production. For example, grammatical information (e.g., about syntactic functions of collocates, argument structure, or gender and number) may make the user's efforts to solve their language questions easier: it may not be necessary for them to turn to further tools, such as grammar books, or even to look at the whole dictionary entry to understand how words go together. Moreover, in an electronic dictionary, collocational networks constitute a secondary means of accessing information, because networks connect words to one another, and they can be used to navigate across dictionary entries. At the same time, it is also possible to provide pop-up windows with additional information, such as examples, short definitions or usage notes. In this way, not only are users told about which lemmas collocate, but they also gather knowledge about language in use.

2. Background

2.1. *Collocational networks*

The concept of collocational networks is based on the idea that texts are organized into lexical patterns. In texts, words are related to one another in syntagmatic relationships that constitute a network in which pairs of words showing mutual attraction are connected. As a consequence, collocational networks are conceived of as networks displaying interlinked pairs of words that constitute collocations, each word being a node connected to further words with which it collocates. This idea was originally postulated by Phillips (1983, 1985, 1989), who provided the theoretical and methodological basis for the concept. Phillips was mainly interested in the notion of *aboutness* of a text, that is, the general topic of the text or, as more recently defined by Oakes (2012: 126), the 'word types that typify a corpus'. His notion of aboutness relies on the regularities in the lexical organisation of texts. As he points out, the lexical component of any given text is strictly organized in terms of collocational patterns, so lexical relations can be

used to infer the topic of a text. In order to explore collocation relations, Phillips (1989) uses computational corpus linguistics methods to analyse the lexical structure of a text, and presents the results in the form of simple graphs.

Phillips' approach is concerned with the lexical structure of texts. In a seminal paper, Williams (1998) extends Phillips' approach to characterize not individual texts but the lexical component of specialized sublanguages. Focusing on the sublanguage of science, Williams (1998: 152-153) states the following:

The conceptual framework of scientific sublanguage can be visualized through closed set collocational networks that are prototypical of a given sublanguage. These networks are proposed as a tool for the extraction of the lexis of scientific discourse communities and the management of that lexis in specialized dictionaries.

In order to identify collocational relations between words, Williams uses statistical measures. Specifically, Williams (1998) advocates the use of Mutual Information (MI) statistics; more recently, he has suggested using MI for scientific texts and z-score for general texts (Williams 2008: 162). Starting from the nuclear word list of the most frequent lexical forms in a corpus - the initial nodes in the network - the probability that a given pair of lexical forms is a collocation is computed. Thus, a list of potential collocations is obtained for each node in which the mutual attraction between the pair of words is statistically significant; that is, for each pair of lexical forms, their co-occurrence is more frequent than expected from the type and the length of corpus. The list of potential collocations must be manually reviewed by human annotators in order to remove unexpected noise. Then, the process is repeated recursively for collocates, which are then treated as nodes in order to detect the collocations in which they take part. Thus, lexical forms are interconnected by virtue of their collocational patterns, making up a complex network.

In Williams (1998), the collocational patterns of a given lexical form are represented via a graph where every lexical form is a node in the graph, the hub of a set of connections to other nodes in the graph. The same theoretical and methodological principles have been used in several subsequent studies (e.g. Williams 2001, Williams and Millon 2010, Alonso et al. 2011, and Williams et al. 2012). Figure 1, for example, shows the collocational network for the verb *to treat* in scientific texts as it is presented in Alonso et al. (2011). In the graph, the verb is the central node¹ (in blue, and in a larger text font), which is connected to its collocates (in red). These, in turn, are central nodes to a second-order cluster of collocates (in green), forming a partial collocational network.

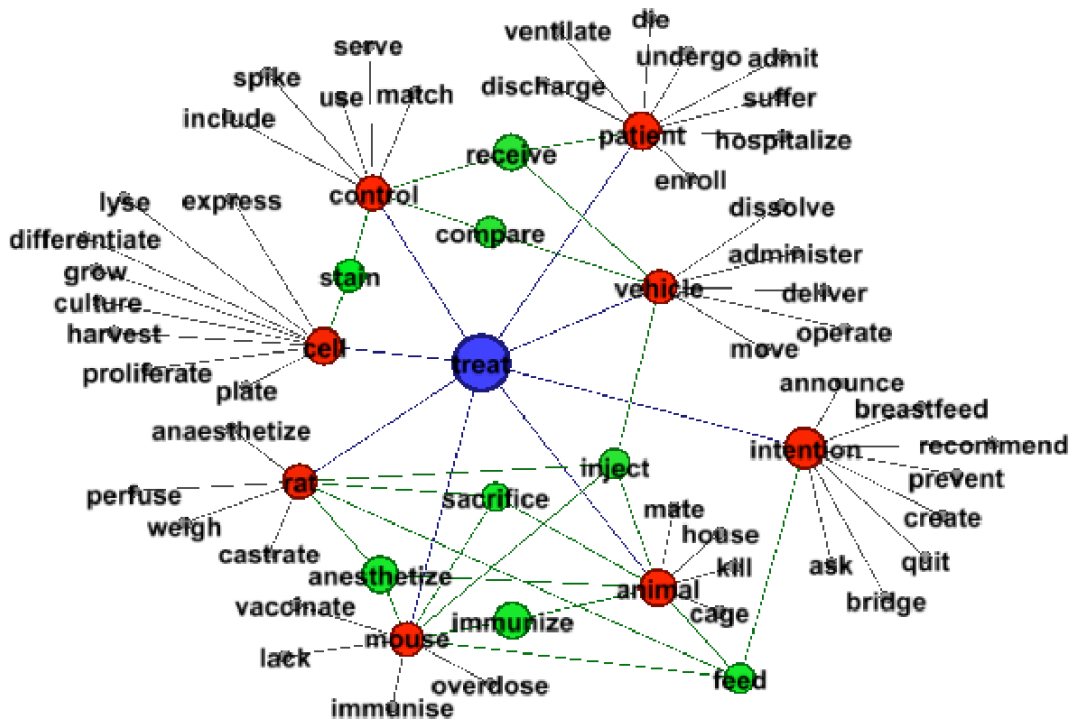


Figure 1. Collocational network for the verb *to treat* in Alonso et al. (2011).

Following Williams (1998), extensive literature has been published which discusses a number of aspects of his proposal, or applies similar ideas to treat collocations in corpus linguistics (cf. Gries 2013 for an updated review of research on collocations). The lion’s share of the discussion relates to the criteria used to define and identify collocations, as well as to discuss which statistical measures are optimal for extracting collocations from corpora (cf. Evert and Krenn 2005, McEnery 2006, Millon 2011, Brezina et al. 2015, among others). While the importance of this discussion is undeniable, this article is not the place to address the question of how to identify collocations; rather, we shall focus on the visualization of collocational networks irrespective of the methodology used to derive them. In the same way, the suggestions below are equally applicable for collocational patterns of either general or specific language.

Existing proposals for collocational network design share a vast number of basic features. Generally, the proposed networks relate pairs of words, each word being a node in the graph; related words (or inflected forms of words) are usually the only information available, as shown in Figure 1. For example, Figure 2 shows the collocational network for the near-synonyms *maritime* and *marine* in a corpus of maritime English provided in Jhang and Lee (2013).

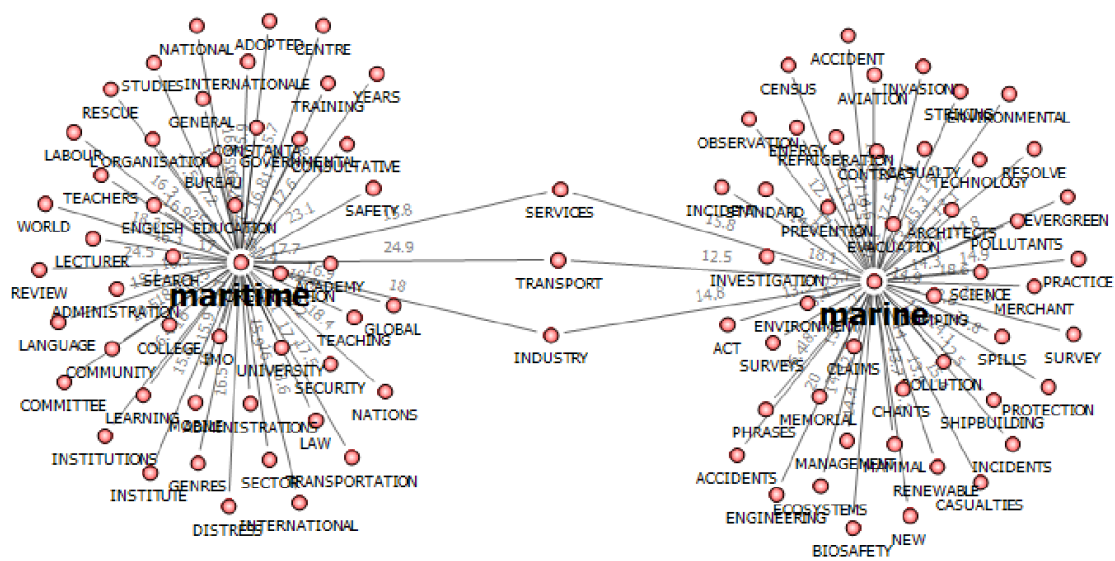


Figure 2. Collocational network for the near-synonyms *maritime* and *marine* in Jhang and Lee (2013).

Some authors have proposed improving this kind of representation by adding different kinds of information. McEnery (2006) and Brezina et al. (2015) suggest adding information on the directionality of lexical selection (cf. Gries 2013), so that collocations are assumed to be pairs of words in which one of the members has a greater preference to co-occur with the other. They use arrows to represent directionality, as shown in Figure 3 (extracted from Brezina et al. 2015: 151):

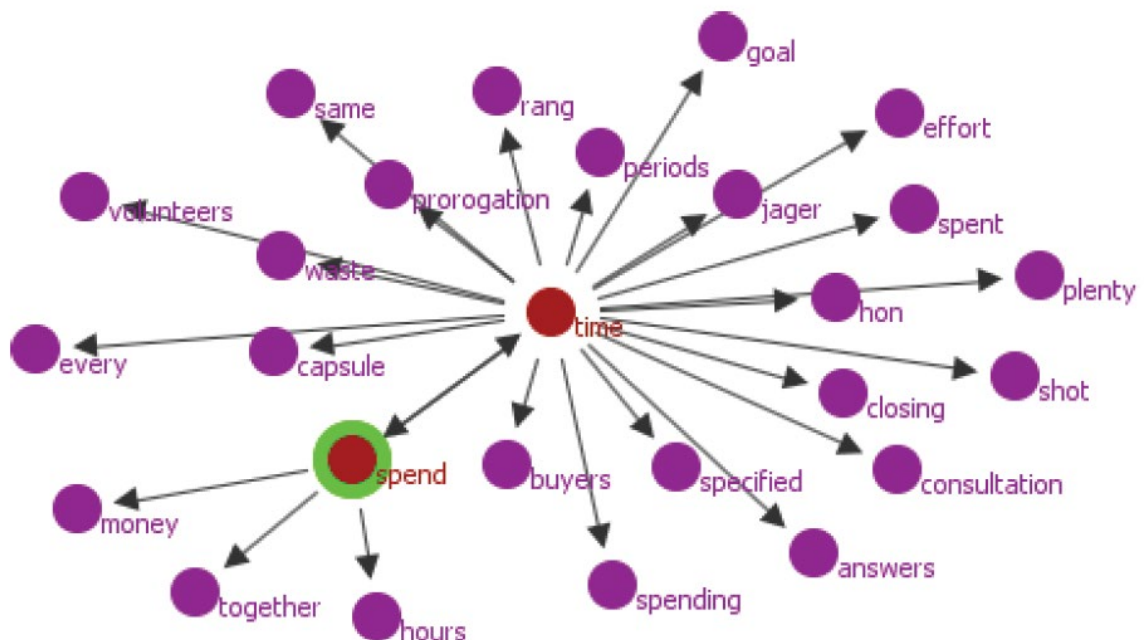


Figure 3. Collocational network for *time* in Brezina et al. (2015).

Lexical Computing, in turn, has opted to represent selected relations by means of colour clustering, as shown in Figure 4, in which objects, subjects and modifiers for the word *drop* have been visually classified in purple, light green and dark green, respectively:

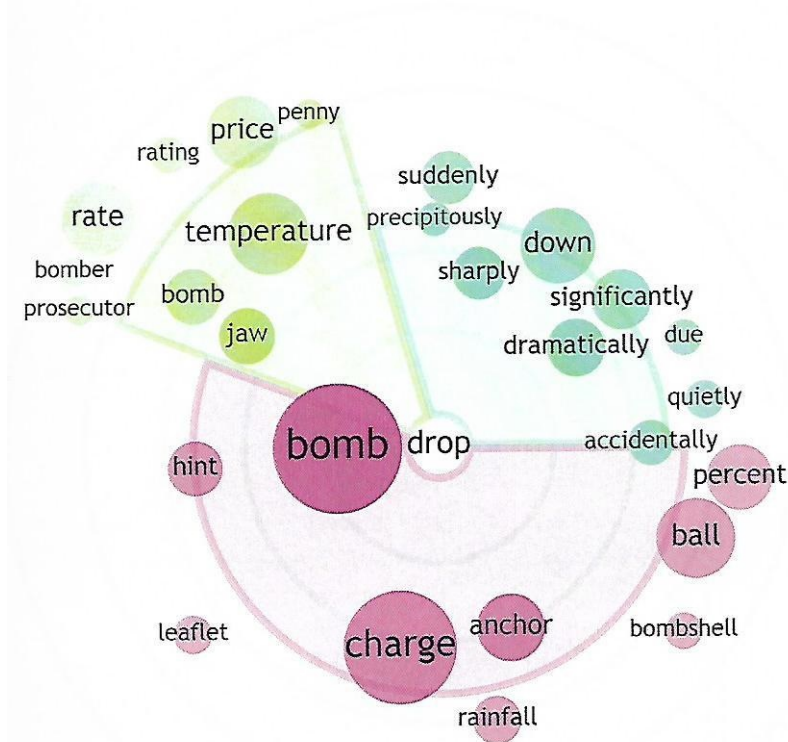


Figure 4. Word sketch visualization for the word drop in Kocincová (2015: 87).

This kind of representation serves as a tool for representing the lexical relations in a given language or sublanguage. The analysis of large corpora shows that the lexicon is organized into networks, so these graphic representations provide an adequate picture of lexical organization in its inherent complexity. Nevertheless, in spite of the designers' best efforts (cf. Kocincová 2015: 79; Kocincová et al. 2015: 18), since lexical relations involve a large number of lexical units, these networks are not easily readable by non-experts, which speaks against their inclusion in dictionaries. Since the primary function of the networks we have just mentioned is not readability, some aspects of the way they present information must be reconsidered and adapted for dictionary users.

We can assume that information about collocations is included in dictionaries to help users build their text production skills and learn vocabulary; as a consequence, the type of representation chosen must contribute to enhancing cognition. In the next section, we briefly review some basic principles of graphic representation that must be taken into account. Our proposal does not focus on the concept of collocation itself, but on how to represent collocation by means of visual networks. Therefore, we will not discuss which specific collocations should be included in the networks, which approach to the concept of collocations is more suitable or which specific statistical measure may offer optimal results. Rather, given a set of collocations, we propose a set of criteria to make visual networks a more efficient resource for dictionary users.

2.2. *Information visualization*

Visualization techniques appeal to the human ability to gather meaning from graphic information sources. Indeed, graphic information has traditionally been a useful means of accessing knowledge: maps are a classic example. Ware (2008: 16-17), for example, discusses how the London Underground map, designed by Harry Beck in 1931, makes use of a schematic

layout and a series of colour codes and symbols for stations and connections that allows rapid access to information. In fact, Beck's system was so successful that most public transport suppliers around the world use the same principles to design metro maps.

Further examples of information visualization include charts, diagrams and calendars. As Pham (2012) discusses, visualization has also been used in electronic dictionaries to show lexical relations. For example, the Thinkmap Visual Thesaurus website² uses visual tools for representing lexical relations, so that 'results are displayed as a tree-like structure with the input word as its root; its nodes are expanded with animations' (Pham 2012: 25). Likewise, the Visuwords graphical online dictionary³ 'allows users to look up English words to find their meanings and associations with other words and concepts' (Pham 2012: 26).

Spence (2001) drew attention to the fact that visualization is a cognitive activity. As noted by Ware (2008: 3), '[w]hen we interact with an information display, such as a map, diagram, chart, graph, or a poster on the wall, we are usually trying to solve some kind of cognitive problem'. The cognitive problem to be solved when using a map 'may be how to get from one location to another' (ibid.); when checking a graph, 'it may be to determine the trend' (ibid.). In the same way, the cognitive problem when navigating a collocational network is to find the word with which a given word collocates to express a specific semantic content.

In this cognitive activity, looking up information from a visual source entails a series of acts of attention and eye movements in order to find the information, called visual queries (Ware 2008). As a consequence, the visual representation has to be adequately designed in order to allow fast, effective access to information and prevent visual queries from becoming too demanding or time-consuming. Visualization displays, thus, must predict users' information-finding strategies and use visual tools to guide queries. These strategies are based on the way in which the human visual system processes information, specifically, the notion of visual patterns, as a visual system seeks out patterns in visual images. The main patterns are colour, shape, connecting contours, proximity, alignment, the enclosing contour, texture, and movement (Ware 2008: 58). Each of these patterns is processed in a different visual channel, so a good strategy when designing a visual display is to use a different pattern for each kind of information. Thus, different objects having the same shape, texture or colour are processed as similar or related objects, and objects grouped by an enclosing contour or a coloured background are processed as belonging to the same set. For instance, in a metro map, colours stand for metro lines, enclosing contour stands for payment zones, and grouping stands for connections.

The effectiveness of a visual display depends on the ease with which users understand the particular meaning of each pattern used, and the degree to which each pattern supports a specific query strategy. Visualization specialists call this *affordance* (Spence 2001: 17-20). Affordance is a concept used in design, borrowed from the perceptual psychologist J. J. Gibson to define certain features of objects that help people understand how to use them. For instance, in the context of information visualization, a button on a web page enables the user to perform an action; a hyperlink enables the user to navigate hypertextually; and a label enables the user to understand a function. According to the typology of affordances in Hartson (2003: 319) a *cognitive affordance* 'is a design feature that helps, aids, supports, facilitates, or enables thinking and/or knowing about something'. Thus, visual patterns in visual information displays may be understood as affordances to facilitate users' information retrieval.

Affordance represents a specific data structure, which acts as a guide to visual queries: structured data are more efficiently processed than unstructured (Ware 2008). From that point of view, the main shortcoming of the kinds of collocational networks presented in Section 2.1 when used as a tool for representing collocational patterns in dictionaries is that information is

not structured in a way that helps users find information easily. If collocational networks are to be used as a search aid, they may benefit from a more structured presentation in which different visual patterns are used to provide different types of information.

Our proposal aims to improve the representation of collocational information in dictionaries via graphical networks that use visual patterns to facilitate queries. The main objective is to provide ways to make finding information more efficient. This means that users must easily find the word or set of words with which a given lemma combines to express a certain idea or notion; additionally, it means that users have to be provided, when necessary, with some basic grammatical information in order to use the collocation adequately. To this end, we propose two improvements. First, an underlying data structure needs to be provided, using a variety of visual patterns to guide visual queries. As we shall discuss in detail in Section 3, we propose using grouping and enclosing contour patterns, in combination with labelling, to present semantically related collocates, as well as different colours and node shapes to provide information about the kind of collocate-base relation. Consequently, we recommend using hybrid graphics, which combine node-edge diagrams as a first order hierarchy level with other visualization techniques in various hierarchy levels of the structure. Second, we propose enriching the data provided by adding grammatical information, as will be explained in Section 4. Users of well-known collocational networks such as those mentioned in Section 2.1 would not expect to find grammatical information about the words they look up in these sources, but rather in grammar books or even in dictionary entries (e.g. a tag saying that a verb is ‘transitive’ or ‘intransitive’). The networks above, in fact, if used in isolation, fail to let the user know about basic syntactic data for text production such as the part of speech of the collocates: for example, Figure 3 presents *waste* as a collocate of *time*, but we are not told whether *waste* would be a noun or a verb (or either) in the frequent collocation. Additionally, if the network is trying to represent the noun *waste*, the user of the network is not told that they need the preposition *of* to link *waste* and *time* in sentence production. The networks described above also fail to present semantic differences: in Figure 3, the user is not told about the countable or uncountable nature of the noun *time* in English, and thus finds no differentiation in the representation of collocates such as *closing* (which collocates with both word forms, *time* and *times*) or *spending* (which collocates only with *time*). The need for this kind of syntactic and semantic information in resources for non-native learners of foreign languages is undisputed. Therefore, it seems natural to suggest the possibility of collocational networks incorporating such information as well.

3. Simplifying queries: Making information easier to find

Making information easier to find relies on structuring visual information with different visual patterns. In this way, visual patterns such as colour or grouping provide an implicit guide for looking up word combinations in collocational networks. Since lexical combinatorics depend on word meanings, which means that a word combines with different lexical units in accordance with the senses in which it is used (Hanks 2012), the main criterion for structuring collocational networks is semantic information. Thus, for example, collocates can be grouped according to semantic criteria. Secondly, grammatical information is used as an ancillary criterion to structure information for guiding queries; thus, for instance, shape or colour can be used to indicate part of speech in collocates. Currently available visual electronic dictionaries such as Visuwords or the Thinkmap Visual Thesaurus signal the part of speech of related words in a similar manner.

We propose using three different visual patterns to structure collocational networks and thus guide users’ queries: grouping, enclosing contour and colour. Each one of them carries a different kind of information. Additionally, labelling can be used to make information easier to find. Although it is not specifically visual information, Fekete and Plaisant (2003) argue for the

use of labelling as an additional source of information, which entails relating verbal and visual information.

3.1. *Collocate groupings and semantic labels*

As discussed in Section 2.1, collocational networks were initially conceived as a tool for representing lexical relations in texts or in natural languages, but not as a device for presenting collocations in dictionaries. In natural languages, lexical relations are complex and multiple; as a consequence, collocational networks tend to be informationally dense, presenting a large number of collocates in any network, each of which constitutes a distinct node in the network. Lexical relations are thus well represented. However, these networks would be less useful in dictionaries oriented toward non-specialist users, because they are difficult to read. Developing the suggestion in Torner and Bernal (2017), we propose a visual simplification of the information in the networks by grouping semantically related collocates together. Enclosing contours and background colour are used, then, as visual patterns to guide user queries.

In our proposal, grouping is first made by part of speech. Additionally, it is based on semantic grounds, according to the hypothesis that users look up word combinations to express a given semantic content. Grouping, thus, has its first *raison d'être* as a device to facilitate access to the information sought. Moreover, if used in learners' dictionaries, grouping has positive effects on vocabulary learning, assuming that vocabulary is more easily learnt by making semantic associations explicit to learners, as posited in previous studies on vocabulary learning (Lewis 1993 and 1997, Nation 2001 and 2008, among many others). Specifically in the domain of learning collocations, Higuera (2017: 256) states that 'relating collocations provides learners with a framework in which to connect new words with acquired knowledge'.

The primary semantic relation for grouping a set of collocates together is (near) synonymy. For example, the Spanish adjective *afectivo*, *afectiva* ('affective') collocates both with *lazo* ('tie'), *relación* ('relationship'), *vinculación* ('linking') and *vínculo* ('link'), which can be easily grouped into a single node in the network; likewise, the adjective collocates with *problema* ('problem') and *trastorno* ('disorder'), which may also be grouped together, as shown in the following partial collocational network. Note that, in what follows, we do not offer complete networks, but only those segments of networks which illustrate the question under discussion.

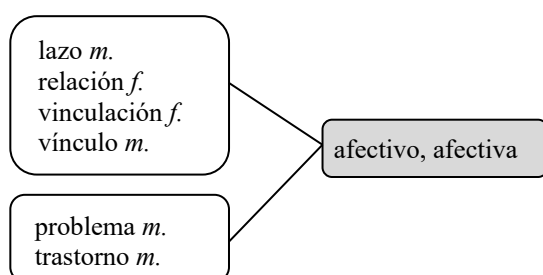


Figure 5. Partial collocational network for *afectivo*, *afectiva* (adjective).

Other semantic relations, however, must be taken into consideration. Frequently, a set of collocates form a lexical class (Bosque 2004, 2011), i.e., a set of words that express a similar semantic content and share collocational properties (such as 'nouns expressing positive feelings' or 'personal relations', etc.), or form a set through a hyponymy relation. Some formal linguistic models which deal with lexical combinatorics use the concept of semantic type to account for lexical selection. For instance, corpus pattern analysis (CPA) (Hanks 2004, 2013) assumes that verbs select lexical classes which share the semantic type in any given lexical

pattern as arguments; arguments are, then, conceived as type-labelled slots which must be fulfilled by members of the lexical class. CPA methodology has already been implemented in general language dictionaries (*Pattern Dictionary of English Verbs*, directed by Hanks), in learner’s dictionaries (*Diccionario de Aprendizaje del Español como Lengua Extranjera*, Renau and Battaner, 2012), as well as in specialised language tools focusing on English for Academic Purposes (Rees, 2018). In the CPA framework, lexical combinatorics is controlled by type matching mechanisms. As a consequence, grouping collocates into lexical classes not only serves as a visual device to make information easier to find, but also shows a linguistically motivated relation.

On other occasions, collocates can be grouped together because they express similar or equivalent relationships with the central node. For example, for *abuelo* (‘grandfather’), the collocates *materno* (‘maternal’) and *paterno* (‘paternal’) can be grouped together, as they express equivalent subtypes; likewise, in both the network of *biológico* (‘biological’) and *adoptivo* (‘adoptive’ or ‘adopted’) the collocates *familia* (‘family’), *madre* (‘mother’), *padre* (‘father’) and *hijo* (‘son, child’) can be grouped into a single node in the network.

In fact, grouping collocates semantically has been highlighted as a feature that allows collocational dictionaries to be used as codifying tools rather than mnemonic tools (Tutin 2010). In this respect, Ferrando (2012) analyses 22 collocational dictionaries in different languages (English, Spanish, French and Italian) and finds that 12 of them structure entries by grouping collocates on semantic grounds. Works such as the OCD and BBI in English, the DCL in Italian, and the DCM in French follow this organisational tendency.

Two additional visual patterns may be used to make queries easier. First, when a node contains several collocates (five or more seems a reasonable number⁴), we advocate labelling the boxes with semantic headings that gloss the core meaning shared by all of the collocates. According to Ferrando (2012), labelling groups of collocates semantically is one of the ways to make the semantic structure of collocational dictionary entries explicit. This system improves the usability of the dictionary because it guides user queries. That is, semantic labels act as an index to make a first selection between sets of collocates in complex networks. In our proposal, semantic labels must be consistent among networks, for which purpose a semantic ontology may be used, such as CPA ontology (Hanks 2004, 2013) or Brandeis Semantic Ontology (Pustejovsky et al. 2006, Havasi 2007).

Figure 6 shows a partial collocational network with semantic labels:

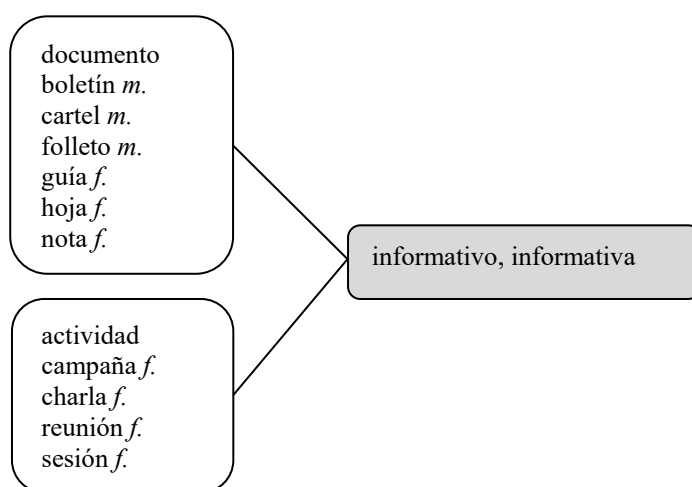


Figure 6. Partial collocational network for *informativo, informativa* (adjective).

Notice that the first box contains nouns for different kinds of informative documents (*boletín* ('bulletin'), *folleto* ('leaflet'), *hoja* ('sheet'), etc.), tagged with the semantic label *documento* ('document'). The second box is headed by the semantic label *actividad* ('activity'), standing for the type of nouns it includes (including *campana* ('campaign'), *charla* ('talk'), etc.).

Secondly, we propose using different background colours for collocate boxes to account for different parts of speech. In spite of the limitations acknowledged in the discussion of the experiment conducted by Dziemianko (2015), recent empirical tests have suggested that 'colour apparently enhance[s] stimulus-driven selection of information from the visual display' in online lexicographic tools (ibid.: 50). Thus, colour acts as a visual pattern that offers a second dimension of structure in networks. The Spanish verb *respetar* ('to respect'), for example, collocates with nouns (or sets of nouns) —*ley* ('law'), *acuerdo* ('agreement'), or *autoridad* ('authority')—, with adverbials —*estrictamente* ('strictly'), *escrupulosamente* ('scrupulously')— or with other verbs —*hacer* ('to make'). In order to guide user queries, the boxes containing collocates of each of these grammatical categories must have, according to our proposal, a different background colour.

3.2. Grouping of the bases

3.2.1. *Multiple networks.* Collocational networks as a tool for analysing text and retrieving collocations from corpora are built (semi-)automatically using statistical measures. They offer a global picture of the collocational patterns of the central node in the network, as they bring about the words with which it collocates. However, as posited by Alonso et al. (2011), lexicographic analysis of the results of automatic retrieval shows that, frequently, the collocations included in a network correspond to different meanings of the central node. This finding suggests that the collocations in a network could be grouped according to semantic criteria. For instance, the Spanish verb *adoptar* ('adopt') collocates with a) nouns such as *niño*, *niña* ('child'), *hijo*, *hija* ('son', 'daughter'), *bebé* ('baby') or *menor* ('minor'); b) nouns such as *decisión* ('decision'), *propuesta* ('proposal') or *medida* ('measure'); and c) nouns such as *costumbre* ('habit'), *comportamiento* ('behaviour'), *moda* ('fashion, trend'), *actitud* ('attitude'); each of these three sets of collocates corresponds to a different meaning of the verb, each of which presents different grammatical and collocational patterns.

Finding collocates for different meanings of a word in the same network may be perceived as unnecessary noise for a user looking up the collocates for a specific meaning. Note, for instance, that in Figure 3, the user is not told about the difference between *time* acting as an uncountable noun, functioning as a prehead modifier in *time capsule*, or a countable noun being modified by an adjective in *same* or *every time*. For a more efficient use of networks, we propose building different networks for each meaning of the central node. For instance, *adoptar* ('adopt') generates a total of three networks, which account for the following meanings: (1) to welcome or accept someone originally external [typically, a child] as part of your family; (2) to start using a new approach or idea; and (3) to start using a particular way of speaking, thinking, or behaving that differs from the one you frequently use. These three semantic patterns are reproduced below by following Hanks' (esp. 2004, 2013) CPA methodology:

- Pattern 1 [[Human 1 = Parent]] *adoptar* [[Human 2 = Child | Animal = Pet]]
- Pattern 2 [[Human]] *adoptar* [[Idea]]
- Pattern 3 [[Human]] *adoptar* [[Attitude | Behaviour]]

A similar account must be given whenever there are category shifts in the central node, that is, when a base may be used as belonging to different parts of speech. In languages such as Spanish, for example, many nouns may be used as adjectives, without any formal mark for the

category shift. We can use both as nouns or adjectives, for instance, units such as *vecino* (‘neighbour’ and ‘neighbouring’), *joven* (‘young person’ and ‘young’), or *suavizante* (‘softener’ and ‘softening’). Category shifts are also frequent in other languages, such as noun-verb shifts in English or adjective-adverb shifts in German. In those cases, the words with which the lemma collocates vary for different parts of speech. For instance, *vecino*, *vecina* collocate with *asociación* (‘association’), *asamblea* (‘assembly’), *junta* (‘meeting’) and with *barrio* (‘neighbourhood’), *calle* (‘street’) when used as nouns, and with *país* (‘country’), *nación* (‘nation’), *ciudad* (‘city’) and *casa* (‘house’), *finca* (‘land’) when used as adjectives. We propose building different collocational networks for each part of speech of the central node, as shown in the following partial collocational networks for *vecino*, *vecina*:

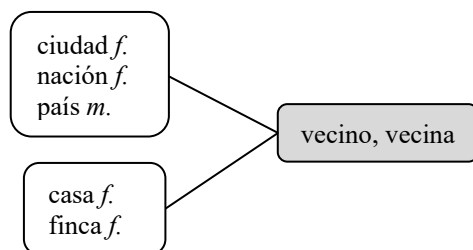


Figure 7. Partial collocational network for *vecino*, *vecina* (adjective).

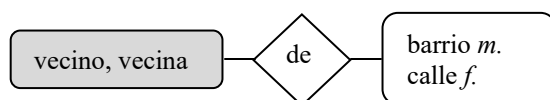


Figure 8. Partial collocational network for *vecino*, *vecina* (noun).

The design of different networks for several meanings or parts of speech of a same word may raise a problem from the point of view of information access, as the user has to be directed to the right network. Our proposal is that different sets of collocations must be provided for each sense definition, not for a whole dictionary entry. Thus, the lexicographical entry of *vecino*, *vecina* in the dictionary would include two different networks, each of which would be related to the corresponding part of speech with which the word may be used. At the same time, users navigating networks, and clicking on any collocate of a specific network, should ideally be redirected to the network corresponding to the part of speech or meaning that is relevant for their query.

3.2.2. Base plurals. It is still a disputed issue in the specialized literature whether collocational networks must include inflected forms or lemmas. For example, the collocational networks reproduced in section 2.1 are built following different criteria in this respect: whereas Alonso et al. (2011) include the canonical form of the collocates, Brezina et al. (2015) include inflected forms. The decision to use inflected forms or lemmas has been argued to be important in the automatic process of collocation retrieval, as the combinatorics of a given word may be different for some inflected forms, so that ‘the collocational network can vary depending on the form of the lexical unit’ (Alonso et al. 2011: 15). On the other hand, using lemmatized words allows a ‘complete panorama of the total environment of a word’ (*ibid.*).

In order to use collocational networks as a means of representing the collocational patterns of a lemma in a dictionary, our criteria for choosing between lemmatized words or inflected forms must be both clarity and informativity. The decision affects the representation of the singular and plural form of nouns. In this case, the most common situation is for units to

be combined either in the singular or the plural, with no semantic difference. For example, corpora show results for both *alianza matrimonial* (‘wedding ring’) and *alianzas matrimoniales* (‘wedding rings’). In these cases, we recommend representing noun + adjective collocations in the singular, following the lexicographic tradition of listing headwords by their canonical singular forms. A different situation holds when the base of the collocation may be either singular or plural but the collocate only occurs in the plural, as in *convivencia entre culturas* (‘coexistence of cultures’) or *separación de bienes* (‘separation of property’), in which both *culturas* and *bienes* are mandatorily plural. In those cases, the collocate is plural in the collocational network.

The most complex situation occurs where some collocations are only possible with a plural base; that is, when some collocational patterns require plural forms. Our proposal is to treat these units similarly to cases of polysemy: by building a separate network for collocations in which the plural of the base is mandatory. This is the case, for instance, in *condición, condiciones* (‘condition(s)'). Meaning ‘situation’ or ‘status’, in combination with a positive adjective like *favorable(s)* (‘favourable’), it is only used in the plural. In those cases, we propose displaying the base in the plural in a second central node of the network. Both collocational networks (those corresponding to singular and plural forms) need not be presented separately, but can appear as a single network with singular and plural combinatorics connected.

A similar treatment has been proposed in Torner and Bernal (2017). As argued, the Spanish noun *amigo, amiga* (‘friend’) combines with several collocates, both plural and singular, but is mandatorily plural when combined with nouns such as *pandilla* (‘gang’), *círculo* (‘circle’) or *grupo* (‘group’). Thus, collocates that can be used with both singular and plural bases, such as *viejo, vieja* or *antiguo, antigua* (‘old’), *verdadero, verdadera* (‘true’) and *íntimo, íntima* (‘close’), *del alma* (‘kindred spirit’) or *entrañable* (‘dear, close’), must be presented separately from collocates in which the base is always plural, as shown in the following partial collocational network:

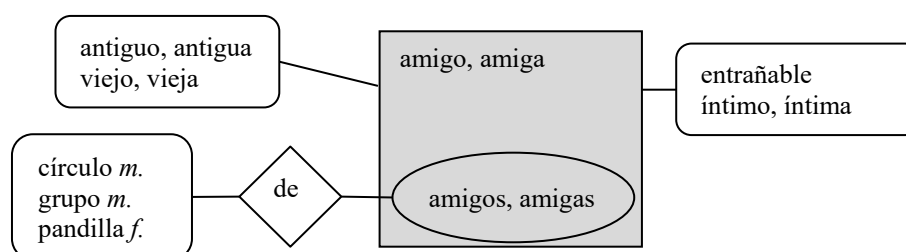


Figure 9. Partial collocational network for *amigo, amiga* (noun).

4. Enriching networks: Explaining grammar visually

As has been argued above, when one thinks of collocational networks, grammatical features do not seem to typically spring to mind. However, the need for this kind of information in dictionaries, especially in learner’s dictionaries, should be addressed from a lexicographic point of view (cf. Torner and Bernal, 2017: 168). This need has typically been acknowledged in the planning of monolingual or bilingual dictionaries. Consider, for instance, how Hunston and Francis (1999: 44) report on the importance of simplifying the representation of word patterns noted in the elaboration of the Collins COBUILD English Dictionary in 1995, under the assumption that ‘[a] learner’s dictionary requires a way of encoding grammar that is comprehensive and yet transparent’ (ibid.). However, the visual explanation of grammar is yet to be explored in the design of networks which, as has been shown above, do not even offer basic information for language production, such as the parts of speech involved in each

collocation. This section focuses on our proposal to *enrich* networks so that they provide users with relevant grammar information for the collocations they may be interested in.

As explained above, the purpose of collocational networks is to explain *visually* how words go together. In the case of networks presented as complementary to a dictionary entry, the visual display of information is not bound to include *all* the grammatical information that lexicographers intend to convey about each unit. Rather, our proposal focuses on basic morphosyntactic data which could be included to provide users with information needed for text production. Specifically, we argue for including information on the grammatical gender of lemmas (morphological information), canonical word orders and mandatory uses of a) specific verb complements and b) certain prepositions (syntactic information). The following subsections report on these lexicographic recommendations.

4.1. Gender

As in all Romance languages, Spanish nouns and adjectives have gender inflection. Gender inflection is not always easy to predict according to formal features, so Spanish dictionaries show inflection for noun and adjective lemmas without exception. In dictionaries not oriented towards native speakers, either bilingual or monolingual, inherent gender of nouns is usually shown for all nouns included both in the examples and definitions, because this feature of language is not intuitive for all learners, especially for those whose mother tongue does not have gender inflection. Therefore, incorporating this morphological information in the network would be justifiably useful.

The way gender is presented in the networks depends on each node's part of speech. If the network's node is a noun, adjective modifiers in collocate-node boxes are listed in the inflected form that agrees in gender with the noun (and not in their lemma form, which shows both masculine and feminine inflectional forms), as in *convivencia civilizada* or *niñez traumática*, in which both adjectival modifiers are feminine. We believe this presentation is clearer for the user than alternative ones, such as *convivencia civilizado, da* or *niñez traumático, ca*, in which adjectives show formal shifts when inflected in masculine or feminine. Figure 10 shows the latter example.

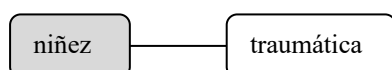


Figure 10. Representation of adjective agreement in noun central nodes.

This lexicographic decision is consistent with common solutions adopted in collocational dictionaries. The most well-known collocational dictionaries of Spanish currently available, *Práctico* and *DiCe*, present collocational adjectives in the same way: thus, for instance, *Práctico* offers the feminine forms of adjectives *abrasadora* ('burning'), *ciega* ('blind'), *ardiente* ('ardent'), or *incendiaria* ('incendiary'), among others, forms that agree grammatically with the lemma in the lexicographical entry for the feminine noun *pasión* ('passion');⁵ while *DiCe* suggests, for example, *intensa* ('intense') or *verdadera* ('true'), also in feminine inflection. An analogue solution has been adopted in dictionaries for other languages with gender inflection, such as DdC or DCM in French, DCL in Italian, and FWD in German.

On the other hand, when the network's node is a variable adjective, the inflection is expressed with an alternation, as in *único, única* or *conflictivo, conflictiva* (see also *informativo, informativa* in Figure 5 and *vecino, vecina* in Figure 7). As can be seen, the alternation shows the formal difference between the masculine and the feminine adjective. Our proposal goes beyond the common practice in monolingual Spanish dictionaries, such as DRAE23 or DUEAE, which only show the alternation from the last shared consonant on (e.g. *conflictivo*,

conflictiva), since online lexicography allows both full forms to be included. This helps the user to infer the form that must be used according to the gender of the noun collocate. By way of illustration, let us compare our proposal to current practice in Spanish lexicographic works on collocation: Figure 11 shows the treatment of the adjectival inflection in *pasivo*, *pasiva* in *Práctico*.⁶

<p>pasivo, va adj. CON SUSTS. sujeto · actitud <i>No entiendo su actitud tan pasiva ante el problema</i> · papel resistencia eutanasia fumador, -a voz <i>una oración en voz pasiva</i></p>
--

Figure 11. Treatment of the adjective *pasivo*, *pasiva* in *Práctico*.

The solution shown in Figure 11 is clear in that it provides the adjective inflection by following the standard practice in monolingual lexicography (*pasivo*, *va*); however, the user is assumed to know the inherent gender of the noun collocates to produce a phrase with gender agreement, because gender information about collocates is not provided. Thus, for example, the different gender of *sujeto* (‘subject’) (masculine) and *actitud* (‘attitude’) (feminine) is not made explicit. Only for the specific cases of *actitud* and *voz* can feminine gender be inferred from the examples.

In our proposal, the gender of nominal collocates is specified in all cases, which leads to a representation like the one in Figure 12. As can be seen in the figure, the masculine nouns *consentimiento* (‘consent’) and *permiso* (‘permit’), which collocate with the adjective base *paterno*, *paterna* (‘paternal’), may be easily differentiated from the feminine collocate *autorización* (‘authorisation’), by the dictionary user, thus guiding their decisions in language production.

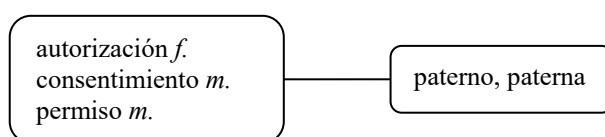


Figure 12. Proposal for the representation of gender in adjectival nodes.

4.2. Syntactic information

This section argues that networks may be used to convey linguistic information about the represented collocations, beyond morphological features. We propose including some basic syntactic information to inform users not only about the words with which a given lemma combines but also about specific syntactic features that need to be borne in mind in order to produce an accurate grammatical construction. We focus on three syntactic features which could be specified in networks: the prototypical order in which word combinations occur in corpora; the syntactic alternations involved in lexical combinatorics; and, finally, the specific prepositions needed to head prepositional complements, which is also a key aspect for collocational competence in many languages, including all Romance languages (Vincze et al. 2016).

4.2.1. Canonical word order. While constituent order is not very flexible in languages such as English or German, Romance languages are, in general, less restrictive in this sense. There are, however, a few restrictions which pose a challenge for the non-native user, since word order is not entirely arbitrary. Specifically, there are some complex syntactic rules governing word order in noun-adjective combinations, as well as constituent order in noun modifiers (Demonte 1999, Rigau 1999, RAE 2009, Castañeda and Chamorro 2014). Consequently, dictionary users could benefit from networks portraying the most prototypical (i.e. frequent in corpus) constituent order. On some occasions, a specific constituent order is mandatory according to syntactic rules;

on other occasions, two different orders are potentially possible, but occurrence in corpora shows that one of them is clearly more frequent than the alternative one. Indeed, frequency could be used as the general principle to deal with word order representation in networks in any language. Similarly, unmarked sentence constituent order in Spanish specifies the relative order between sentence subject, verb and complement, although it may be modified due to information structure.

We propose presenting constituents in their canonical word order in collocational networks. In terms of on-screen information distribution, in the absence of empirical data that suggest proceeding otherwise, we propose that networks be designed in accordance with the reading practices of each language. For example, in Spanish, users are presumed to read the networks from the left-hand side to the right-hand side of the screen. Consider the following examples in Figure 13 and Figure 14, which show instances in which the node is a noun (*permiso* ‘permit’) and an adjective (*rico, rica* ‘rich’) Notice that examples reproduce partial networks; boxes placed to the left and to the right of the central node represent independent collocations, and should not be read as a chain of collocations (Ramos and Wanner 2007):

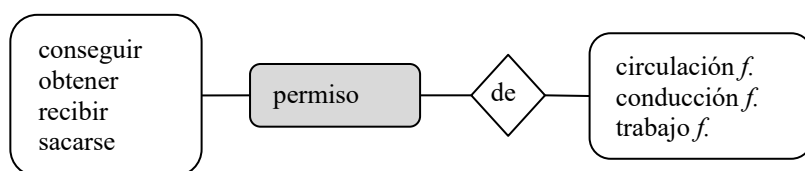


Figure 13. Proposal for the representation of canonical word order in noun nodes.

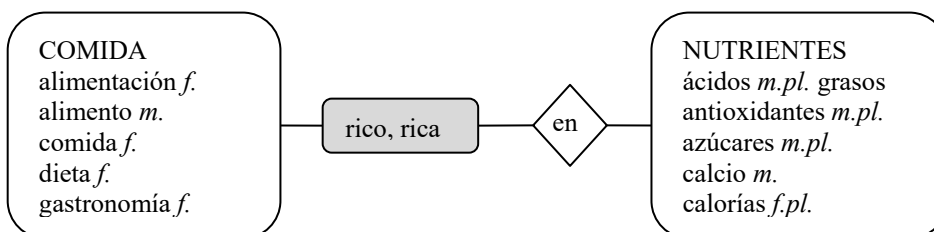


Figure 14. Proposal for the representation of canonical word order in adjective nodes.

In both cases, this graphic distribution allows the dictionary user to take advantage of the collocates listed at either side of the node in order to easily produce Spanish sentences by imitation: *Obtuvo el permiso de trabajo*; *Es una dieta rica en calcio*, etc. This proposed graphic distribution is clearly different from that of collocational networks presented in section 2.1, where collocates are distributed randomly through space. As an additional option for the user of the networks, we propose the inclusion of pop-up examples showing basic sentences like the aforementioned, so that they may also be used as a model for language production.

The same principle applies for networks in which the node is a verb: users may easily be told visually whether noun collocates are typically the subject or the direct object of a node verb by positioning collocates on either side of the node, in accordance with their canonical position in combination with it, i.e., when word order is not modified for purposes of information structure. Thus, the subject collocates would occupy a position on the left in the network, whereas direct objects would occupy a position on the right, following canonical word order. Consider the difference in dealing with the collocations *amor perdurar* (‘love last’) vs. *dar*

amor ('give love'), which share *amor* ('love') as a common node, as shown in the partial collocational network for *amor* in Figure 15.

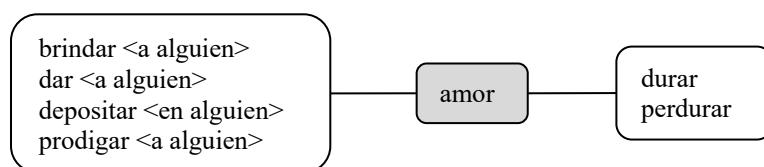


Figure 15. Proposal for the representation of canonical word order in verb nodes.

Likewise, modifiers that occupy a restricted position in combination with their heads should be indicated in the networks, as in *gran amor* ('big love'), with the adjective mandatorily preceding the noun, vs. *amor incondicional* ('unconditional love'), in which the adjective must follow the noun, and *mayor enemigo* ('greatest enemy') vs. *enemigo mortal* ('mortal enemy'), which show the same noun-adjective order behaviour. The same principle applies to adverb combinatorics: users gather information about their position with respect to verbs or adjectives in examples such as *oponerse abiertamente* ('openly oppose') vs. *abiertamente opuesto* ('openly opposed').

4.2.2. Prepositions and syntactic arguments. Prepositions are a frequent source of errors in learners' production of collocations (see e.g. Uriel Domínguez 2014 and Vincze et al. 2016), and even native speakers of a given language may struggle with prepositions in production. In Spanish, for example, it is frequent to have doubts about the restrictive *a* ('to') that introduces direct objects with the lexical type [[Human]], as in *padre cuidar <a hijo>* ('father look after son') (RAE 2009: 2630-2654). Thus, we propose including solutions for this type of doubt in the networks.

Prepositions take part in collocational constructions in two different ways. On the one hand, prepositions may be needed as a link between the base and the collocates; in these cases, informing users about the preposition to use prevents them from making an incorrect syntactic construction despite being right in the choice of the collocate. On the other hand, collocates frequently have their own complements, which are mandatory when combining with the base in a given sense, so users benefit from being informed both of the need to use the complement and the particular preposition it takes.

With regard to the first case, prepositions as a link between the base and the collocate, we propose including prepositions in a separate box, as suggested in Alonso (2009). We advocate using a different box shape as a visual cue to inform the user that the preposition is not the lexical word that makes up the collocation but a linking word between the base and the collocate. It is worth noting that, typically, nodes have modifiers with parallel syntactic structure. An example of this can be found in *confirmación* ('confirmation'), which takes the preposition *de* ('of') to introduce object modifiers with the lexical type [[Event]]: *confirmación de reserva* ('reservation'), *de recepción* ('reception'), *de solicitud* ('request, application'). Therefore, prepositions are not repeated but listed only once in these cases, which makes the parallelism in the construction clearer. See the example in Figure 16.

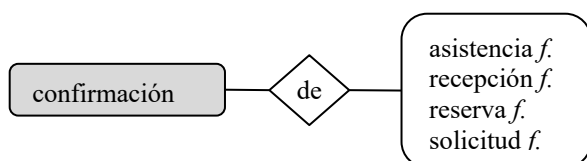


Figure 16. Proposal for the representation of prepositional complements.

The main limitation of this decision, however, is that semantically-connected modifiers become separate boxes in the network if two semantically related collocations have a different syntactic pattern. Given this, if we wish to help the dictionary user see the parallelism between *convivencia en paz* ('coexistence in peace') and *convivencia en armonía* ('coexistence in harmony'), the proposed solution will group together the collocates that are introduced by the same preposition. On the other hand, this solution displays the semantically related collocations *convivencia pacífica* ('peaceful coexistence') or *convivencia placida* ('placid') separately, because in our proposal the visual structuring of collocations in the network gives priority to grammatical aspects rather than to semantic ones. In order to show the close relation between both groups of collocations while preserving the grammatical information on prepositions, an enclosing box encompassing the two subgroups of collocations may be used, as shown in Figure 17.

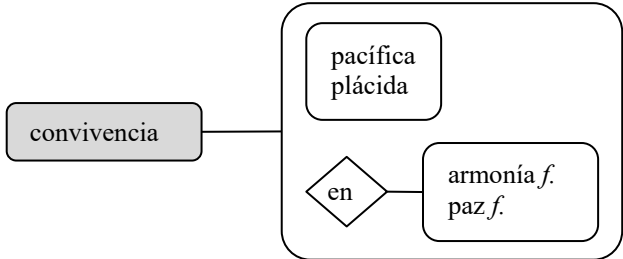


Figure 17. Proposal for the representation of enclosing boxes.

Regarding complements of collocates, our interest in facilitating production also leads us to favour listing prepositional complements of collocates in the network. For example, in building the network for *afecto* ('affection'), it may not suffice to let the dictionary user know that *mostrar* ('show') and *afecto* ('affection') are two words that go together: indeed, both words can only rarely be combined without a prepositional complement introduced by *a*: *mostrar <a alguien> afecto* ('to show affection to somebody'). The same applies to location complements such as *situar <en un lugar> casa* ('to place a house in a location') or predicative clauses like *hacer <a alguien> tío, tía* ('to make someone uncle, aunt'). We propose incorporating these complements in the collocational network, as can be seen in Figure 15, reproduced here for convenience:

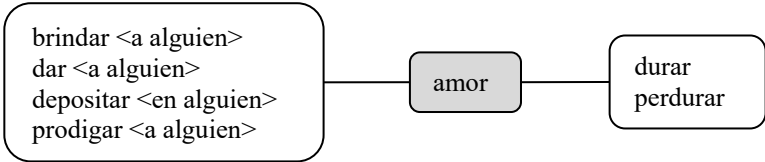


Figure 18. Representing collocate arguments.

Thus, the network provides information about the complements needed by the collocate in order to produce syntactic constructions. Nevertheless, the segmentation of information into

boxes prevents users from reading the collocates in the more natural word order, i.e. *brindar amor a alguien* ('to offer love to somebody'), which is indisputably more natural in Spanish than *brindar a alguien amor*, as offered in the network in Figure 18. To solve this potentially misleading representation of information, we propose providing examples of use in pop-up windows from which sentence word order may be inferred, following the suggestion of McCarthy et al. (2015) to account for grammatical information about collocations. Frankenberg-Garcia et al. (2018) propose a similar device in the form of a writing assistant to help writers in the use of collocations..

There are instances in which the preposition is mandatory only in one of two possible syntactic alternations of a semantic pattern (cf. Hanks 2013: esp. 186-207). Such is the case of *con* ('with') in *conciliar* ('reconcile'). This verb shows two well-delimited syntactic patterns: In the first one, it combines with plural nouns (e.g. *conciliar posturas*, 'reconcile positions'), while in the second one, it combines with a direct object and a prepositional phrase introduced by *con* ('with') (e.g. *conciliar vida laboral con vida familiar*, 'find a balance between work and family life'). When representing collocates connected to the latter pattern, we also recommend including the preposition explicitly.

5. Concluding remarks

Collocational networks were originally conceived as tools to (semi-)automatically identify collocations in corpora, as well as to represent collocational relations in texts (Philips 1983, Williams 1998, Alonso et al. 2011, Brezina et al. 2015, among many others). Their original function was not, thus, to be consulted by non-expert users, as they include much unstructured information which is often difficult to read, which is consistent with the fact that they were designed to show the complexity of lexical relations rather than be easy to read. Recent work, however, has highlighted the possibility of using them as powerful visualization tools in several contexts (Alonso et al. 2011, Jhang and Lee 2013, Masucci and Rogers 2006, McEnery 2006, Mouri et al. 2014). This paper has argued that their potential could be further exploited from a pedagogical, lexicographic point of view if some improvements are made by following the general principles of information visualization (Pham 2012, Spence 2001, Ware 2008): on the one hand, by simplifying access to the networks automatically retrieved from corpora, which include unnecessary noise for the non-expert dictionary user and, on the other, by enriching them with grammatical information, visual networks may become crucial tools in the current lexicographic landscape.

We have proposed a number of specific measures that could be adopted to use collocational networks as a complementary tool to represent lexical combinatorics in dictionaries. Specifically, this paper has argued that information needs to be made easier to find by structuring it with visual techniques, so that different visual resources like colour or grouping are used to guide user queries. In this respect, we advocate grouping collocates in accordance with both syntactic and semantic criteria: we have argued for collocates to be presented by part of speech and separated into different boxes to account for semantic relations, sometimes even adding semantic labels at the top of the boxes (McCarthy et al. 2015). Likewise, we have listed the advantages of offering different networks for different meanings (polysemous words) as well as dealing with inflected forms or lemmas in a systematic manner. Furthermore, we have proposed that networks, and especially those aimed at learners, should include grammatical information, including both morphological and syntactic information (namely gender inflection, canonical word order and use of prepositions, among others) which have been hitherto neglected in this type of lexicographic material. Thus, the sections above are an invitation for a new turn in the lexicographical approach to networks: rather than being sources

merely displaying automatically retrieved collocations, networks can become useful tools for language production whose design invites a pedagogical perspective.

Notes

1 The central node corresponds to the base of the collocation in those theoretical models which argue for a directional semantic selection in collocations (Mel'čuk 1997, Mel'čuk et al., 1995, among many others). In this paper, we will use the terms central node and base as equivalents to refer to the word which generates the collocational network, and collocates to refer to the words with which the base combines.

2 Available at <http://www.visualthesaurus.com/>.

3 Available at <http://www.visuwords.com/>.

4 Note that 5 is proposed here as a tentative number, since no experimental studies can confirm this proposal to date. The general consensus in perceptual number discrimination, since Jevons' studies in the 19th century, is that humans are able to discriminate sets made up of a maximum of 5 elements.

5 *Práctico* already improves the proposal made by the previously published dictionary *Redes*, which does not show gender inflection for collocates: for the same noun *pasión* ('passion'), the latter offers adjectives such as *abrasador* ('burning'), *amoroso* ('loving'), or *ardoroso* ('fervent'), among others, in the masculine form.

6 On the other hand, this proposal represents an improvement compared to the treatment of adjectives in both *Redes* and *DiCe*, which do not provide the user with this information, and list all lemmas in the masculine form.

References

A. Dictionaries

Alonso Ramos, M. (ed.). In progress. *Diccionario de colocaciones del español*. Accessed on 10 June 2018. <http://www.dicesp.com.> (DiCe)

Battaner, P. (dir.). 2002. *Diccionario de uso del español de América y España*. Barcelona: SPES Editorial. (DUEAE)

Beauchesne, J. 2001. *Dictionnaire des cooccurrences*. Montréal: Guérin. Accessed on 10 June 2018. <http://www.btb.termiumplus.gc.ca/tpv2guides/guides/cooc/index-fra.html?lang=fra.> (DdC)

Benson, M., E. Benson and R. Ilson. 1986. *The BBI Combinatory Dictionary of English: A Guide to Word Combinations*. Amsterdam/Philadelphia: John Benjamins. (BBI)

Bosque, I. (dir.). 2004. *REDES. Diccionario combinatorio del español contemporáneo*. Madrid: Ediciones SM. (*Redes*)

Bosque, I. (dir.). 2006. *Diccionario combinatorio práctico del español contemporáneo: las palabras en su contexto*. Madrid: SM. (*Práctico*)

Francis, B. and I. McIntosh (eds.). 2002 and 2009. *Oxford Collocations Dictionary for Students of English*. Oxford: Oxford University Press. (OCD)

Häcki Buhofer, A., M. Dräger, S. Meier and T. Roth (eds.). 2014. *Feste Wortverbindungen des Deutschen: Kollokationen-Wörterbuch für den Alltag*. Tübingen: Francke. (FWD)

Le Robert. 2007. *Dictionnaire des combinaisons de mots*. Paris: Le Robert. (DCM)

Real Academia Española. 2001. *Diccionario de la lengua española*. 23rd ed. Accessed on 10 June 2018. <http://www.rae.es/recursos/diccionarios/drae.> (DRAE23)

Urzi, F. 2009. *Dizionario delle Combinazioni Lessicali*. Luxemburg: Convivium. (DCL)

B. Other literature

- Alonso Campo, A. 2009.** *Características del léxico del medio ambiente en español y pautas de representación en el diccionario general*. Ph.D. Thesis, Universitat Pompeu Fabra.
- Alonso Campo, A., Ch. Millon and G. Williams. 2011.** ‘Collocational Networks and their Application to an E-Advanced Learner’s Dictionary of Verbs in Science (DicSci).’ In Kosem, I. and K. Kosem (eds), *Electronic Lexicography in the 21st Century. New Applications for New Users. Proceedings of eLex 2011, Bled, 10-12 November 2011*. Ljubljana: Trojina, Institute for Applied Slovene Studies, 12–22.
- Bahns, J. and M. Eldaw. 1993.** ‘Should We Teach EFL Students Collocations?’ *System* 21.1: 101-104.
- Bosque, I. 2004.** ‘Combinatoria y significación. Algunas reflexiones.’ In Bosque, I. (dir.), *Redes. Diccionario combinatorio del español contemporáneo*. Madrid: SM, LXXVII-CLXXIV.
- Bosque, I. 2011.** ‘Deducing Collocations.’ In Boguslavsky, L. and L. Wanner (eds), *Proceedings of the 5th International Conference on Meaning-Text Theory*. Barcelona: Universitat Pompeu Fabra, vi-xxiii.
- Brezina, V., T. McEnery and S. Wattam. 2015.** ‘Collocations in Context: A New Perspective on Collocation Networks.’ *International Journal of Corpus Linguistics* 20.2: 139-173.
- Castañeda Castro, A. and M. D. Chamorro Guerrero. 2014.** ‘Adjetivos antepuestos y pospuestos al sustantivo. Problemas descriptivos y propuestas didácticas.’ In Castañeda Castro, A. (coord.), *Enseñanza de la gramática avanzada de ELE: criterios y recursos*. Madrid: SGEL, 223-266.
- Demonte, V. 1999.** ‘El adjetivo. Clases y usos. La posición del adjetivo en el sintagma nominal.’ In Bosque, I. and V. Demonte (eds), *Gramática descriptiva de la lengua española* 1.3: 129-216.
- Dziemianko, A. 2015.** ‘Colours in Online Dictionaries: A Case of Functional Labels.’ *International Journal of Lexicography* 28.1: 27-61.
- Evert S. and B. Krenn. 2005.** ‘Exploratory Collocation Extraction.’ In Cosme, C., C. Gouverneur, F. Meunier, M. Paquot (eds), *Phraseology 2005. The many faces of Phraseology*. Louvain: Université catholique de Louvain, 113–115.
- Fekete, J. D. and C. Plaisant. 2003.** ‘Interactive Information Visualization of a Million Items.’ In Bederson, B. and B. Shneiderman (eds), *The Craft of Information Visualization*. San Francisco: Morgan Kaufmann, 279-286.
- Ferrando, V. 2012.** *Aspectos teóricos y metodológicos para la compilación de un diccionario combinatorio destinado a estudiantes de E/LE*. Ph.D. Thesis, Universitat Rovira i Virgili.
- Frankenberg-Garcia, A., R. Lew, J. Roberts, G. P. Rees, and N. Sharma. 2018.** ‘Developing a writing assistant to help EAP writers with collocations in real time.’ *ReCALL*, 1-17.
- Granger, S. 1998.** ‘Prefabricated Patterns in Advanced EFL Writing. Collocations and Formulae.’ In Cowie, A. P. (ed), *Phraseology: Theory, analysis and applications*. Oxford: Clarendon Press, 145-160.
- Gries, S. Th. 2013.** ‘50-something Years of Work on Collocations: What Is or Should Be Next.’ *International Journal of Corpus Linguistics* 18.1: 137-166.
- Hartson, H. R. 2003.** ‘Cognitive, physical, sensory, and functional affordances in interaction design.’ *Behaviour and Information Technology*: 22, 315–338.
- Hanks, P. 2004.** ‘The Syntagmatics of Metaphor and Idiom.’ *International Journal of Lexicography* 17.3: 245-274.

- Hanks, P. 2012.** 'How People Use Words to Make Meanings: Semantic Types Meet Valencies.' In Boulton, A. and J. Thomas (eds), *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno: Masaryk University Press, 52-67.
- Hanks, P. 2013.** *Lexical analysis: Norms and exploitations*. London: The MIT Press.
- Havasi, C., A. Rumshisky and J. Pustejovsky. 2007.** 'An Evaluation of the Brandeis Semantic Ontology.' In Bouillon, P., L. Danlos and K. Kanzaki (eds), *Fourth International Workshop on Generative Approaches to the Lexicon*. Paris, France.
- Higueras, M. 2017.** 'Pedagogical Principles for the Teaching of Collocations in the Foreign Language Classroom.' In Torner, S. and E. Bernal (eds), *Collocations and other lexical combinations in Spanish*. London/New York: Routledge, 250-266.
- Hoey, M. 2005.** *Lexical Priming: A New Theory of Words and Language*. London/New York: Routledge.
- Howarth, P. 1998.** 'The Phraseology of Learners' Academic Writing.' In Cowie, A. P. (ed), *Phraseology: Theory, analysis and applications*. Oxford: Clarendon Press, 161-186.
- Hunston, S. and G. Francis. 1999.** *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam/Philadelphia: John Benjamins.
- Jhang, S.-E. and S.-M. Lee. 2013.** 'Visualization of Collocational Networks: Maritime English Keywords.' *Language Research* 49.3: 781-802.
- Kocincová, L. 2015.** *Interactive Visualization Methods for Sketch Engine*. M.A. Thesis, Brighton: Lexical Computing.
- Kocincová, L., V. Baisa, M. Jakubíček and V. Kovar. 2015.** 'Interactive Visualizations of Corpus Data in Sketch Engine.' In Grigonyté, G., S. Clematide, A. Utká and M. Volk (eds), *Proceedings of the Workshop on Innovative Corpus Query and Visualization Tools at NODALIDA 2015*. Vilnius: Linköping University Electronic Press, 17-22.
- Laufer, B. 2003.** 'Vocabulary Acquisition in a Second Language: Do Learners Really Acquire Most Vocabulary by Reading? Some Empirical Evidence.' *Canadian Modern Language Review* 59.4: 567-587.
- Laufer, B. 2011.** 'The Contribution of Dictionary Use to the Production and Retention of Collocations in a L2.' *International Journal of Lexicography* 24.1: 29-49.
- Lewis, M. 1993.** *The Lexical Approach*. Hove: Language Teaching Publications.
- Lewis, M. 1997.** *Implementing the Lexical Approach*. London: Language Teaching Publications.
- Masucci, A. and G. Rodgers. 2006.** 'Network Properties of Written Human Language.' *Physical Review* 74: 1-8.
- McCarthy, D., Kilgarriff, A., Jakubíček, M. and Reddy, S. 2015.** 'Semantic Word Sketches'. In *8th International Corpus Linguistics Conference (CL 2015)*. Lancaster University.
- McEnery, T. 2006.** *Swearing in English: Bad Language, Purity and Power from 1586 to the Present*. Abington: Routledge.
- Mel'čuk, I. 1997.** *Vers une linguistique sens-texte. Leçon inaugurale*. Paris: Collège de France.
- Mel'čuk, I., A. Clas and A. Polguère. 1995.** *Introduction à la lexicologie explicative et combinatoire*. Brussels: Duculot.
- Meunier, F. and S. Granger. 2008.** *Phraseology in Foreign Language Learning and Teaching*. Amsterdam/Philadelphia: John Benjamins.
- Millon, C. 2011.** *Acquisition automatique de relations lexicales désambiguïsées à partir du Web*. Ph.D. Thesis, Université de Bretagne-Sud.
- Mouri, K., H. Ogata, N. Uosaki and S. Liu. 2014.** 'Visualizing Ubiquitous Learning Logs Using Collocational Networks.' In Liu, C.-C., H. Ogata, S. C. Kong and A. Kashihara

- (eds), *Proceedings of the 22nd International Conference on Computers in Education*. Japan: Asia-Pacific Society for Computers in Education, 685-693.
- Nation, I. S. P. 2001.** *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. 2008.** *Teaching Vocabulary: Strategies and Techniques*. Boston: Heinle.
- Nesselhauf, N. 2003.** 'The Use of Collocations in Advanced Learners of English and some Implications for Teaching.' *Applied Linguistics* 24.2: 223-242.
- Nesselhauf, N. 2005.** *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.
- Oakes, M. P. 2012.** 'Describing a Translational Corpus.' In Oakes, M. P. and M. Ji (eds), *Quantitative Methods in Corpus-Based Translation Studies*. Amsterdam/Philadelphia: John Benjamins (Studies in Corpus Linguistics), 115-148.
- Pham, B. 2012.** *Visualize Online Collocation Dictionary with Force-directed Graph*. M. A. Thesis, University of Waikato.
- Phillips, M. 1985.** *Aspects of Text Structure: An Investigation of the Lexical Organisation of Text*. New York: Elsevier.
- Phillips, M. 1989.** *Lexical Structure of Text*. Discourse Analysis Monograph 12, University of Birmingham.
- Phillips, M. K. 1983.** *Lexical Macrostructure in Science Text*. Unpublished Ph.D. Thesis, University of Birmingham.
- Pustejovsky, J., C. Havasi, J. Littman, A. Rumshisky and M. Verhagen. 2006.** 'Towards a Generative Lexical Resource: The Brandeis Semantic Ontology' (demo). 5th *International Conference on Language Resources and Evaluation (2006)*, Genoa, Italy.
- RAE and Asociación de Academias de la Lengua. 2009.** *Nueva gramática de la lengua española*. Madrid: Espasa-Calpe.
- Ramos, M. and L. Wanner, 2007.** 'Collocation chains: how to deal with them?'. In Gerdes, K., T. Reuther and L. Wanner (eds), *Meaning – Text Theory 2007. Proceedings of the 3rd International Conference on Meaning-Text Theory*, Klagenfurt, May 20-24. Munich: Wiener Slavistischer Almanach, Sonderband 69, 11-20.
- Rees, G. P. 2018.** *A Phraseological Multi-discipline Approach to Vocabulary Selection for English for Academic Purposes*. Ph.D. Thesis, Universitat Pompeu Fabra.
- Renau, I. and P. Battaner, 2012.** 'Using CPA to represent Spanish pronominal verbs in a learner's dictionary' In Fjeld, R. V. and J. M. Torjusen (eds), *Proceedings of the 15th EURALEX International Congress*. Oslo, Norway: Department of Linguistics and Scandinavian Studies, University of Oslo, 350–361.
- Rigau, G. 1999.** 'La estructura del sintagma nominal: los modificadores del nombre.' In Bosque, I. and V. Demonte (eds), *Gramática descriptiva de la lengua española*. Madrid: Espasa Calpe, 311-362.
- Spence, R. 2001.** *Information Visualization*. New York: Addison-Wesley.
- Spina, S. 2010.** 'The Dictionary of Italian Collocations: Design and Integration in an Online Learning Environment.' In Calzolari, N., K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis, M. Rosner, D. Tapias (eds), *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta*, 3202-3208.
- Torner, S. and E. Bernal. 2017.** 'Collocations in Learner's Dictionaries.' In Torner, S. and E. Bernal (eds), *Collocations and other Lexical Combinations in Spanish*. London/New York: Routledge, 157-172.
- Tutin, A. 2010.** 'Le traitement des collocations dans les dictionnaires monolingues de collocations du français et de l'anglais.' In Neveu, F., V. Muni Toke, J. Durand, T.

- Klingler, L. Mondada, S. Prévost (eds), *2eme Congres Mondial de Linguistique Francaise*, 1075-1090.
- Uriel Domínguez, M. 2014.** *Las colocaciones en un corpus de aprendices valón y flamenco.* M.A. Thesis, Universitat de Barcelona/Universitat Pompeu Fabra.
- Vincze, O., M. García Salido, A. Orol and M. Alonso Ramos. 2016.** ‘A Corpus Study of Spanish as a Foreign Language Learners’ Collocation Production.’ In Alonso Ramos, M. (ed), *Spanish Learner Corpus Research: Current Trends and Future Perspectives.* Amsterdam/Philadelphia: John Benjamins, 299–332.
- Walker, C. 2009.** ‘The Treatment of Collocations by Learners’ Dictionaries, Collocational Dictionaries and Dictionaries of Business English.’ *International Journal of Lexicography* 22.3: 281-299.
- Ware, C. 2008.** *Visual Thinking: for Design.* Massachusetts: Morgan Kaufmann.
- Williams, G. 1998.** ‘Collocational Networks: Interlocking Patterns of Lexis in a Corpus of Plant Biology Research Articles.’ *International Journal of Computational Linguistics* 3.1: 151–171.
- Williams, G. 2001.** ‘Mediating between Lexis and Texts: Collocational Networks in Specialised Corpora.’ *ASp* 31-33, 63-76.
- Williams, G. 2008.** ‘The Good Lord and his works: A corpus-driven study of collocational resonance.’ In Granger, S. and F. Meunier (eds), *Phraseology: An interdisciplinary perspective.* Amsterdam: John Benjamins Pub. Company, 159–173.
- Williams, G. and C. Millon. 2010.** ‘Going Organic: Building an Experimental Bottom-up Dictionary of Verbs in Science.’ *Proceedings of the 14th Euralex Congress, Leeuwarden*, 6-10 July 2010. Leeuwarden: Fryske Akademy, 1251–1257.
- Williams, G., C. Millon and A. Alonso Campo. 2012.** ‘Growing Naturally: The DicSci Organic E-Advanced Learner’s Dictionary of Verbs in Science.’ In Vadvedt, R. and M. Torjusen (eds), *Proceedings of the 15th Euralex International Congress, Oslo*, 7-11 August 2012. Oslo: University of Oslo, 1008–1013.
- Wray, A. 2002.** *Formulaic Language and the Lexicon.* Cambridge: Cambridge University Press.