Cabezas-García, Melania and Pamela Faber (2018). Phraseology in specialized resources: an approach to complex nominals. *Lexicography* 5(1), 55-83. Available at: http://link.springer.com/article/10.1007/s40607-018-0046-x

PHRASEOLOGY IN SPECIALIZED RESOURCES: AN APPROACH TO COMPLEX NOMINALS

Melania Cabezas-García and Pamela Faber

University of Granada

Abstract. In English, the international language of communication (Tono 2014), complex nominals (CNs) are frequently used to convey specialized concepts (Sager et al. 1980; Nakov 2013). These phraseological units have a nominal head that is modified by another element (e.g. *hydropower production*). Problems can arise in relation to their identification, their bracketing or internal structure disambiguation, their meaning access, and their translation or production in another language. Although they are not marginal phenomena in specialized language, they are rarely included in specialized resources. Even when they are included, their treatment is not systematic (Cabezas-García and Faber 2017a). This article describes the representation of CNs in EcoLexicon (www.ecolexicon.ugr.es), a terminological knowledge base, whose new phraseological module will include verb collocations (e.g. *a volcano spews lava*) as well as CNs. For that purpose, we used a wind power corpus in English and Spanish for term extraction, semantic analysis, establishment of interlinguistic correspondences, and definition crafting. We propose different access points to information (Kwary 2012), such as the CNs formed from a given term, a bilingual view in English and Spanish, or the syntactic-semantic combinations in CNs. The structure of the CN module is based on the semantics of these phraseological units, which facilitates the specification of mapping rules as well as knowledge acquisition (Faber 2012).

Keywords: complex nominal; phraseology; specialized language; terminological knowledge base.

1. Introduction

Phraseology, understood as the tendency for words to be co-selected by users to achieve meanings (Cheng et al. 2008: 236), is an area of great importance not only in general discourse (Bally [1909]1951; Sinclair 2000; Benson et al. 2009; *inter alia*) but also in specialized language (L'Homme 2009; Leroyer 2006; Buendía Castro 2013; *inter alia*). This is due to the frequency of phraseological units, given that approximately 80% of the words in discourse are selected in combination with other units (Sinclair 2000: 197).

Scientific texts in English, the international language of communication (Tono 2014), are characterized by the frequency of complex nominals (CNs) (e.g. *ozone depletion*) (Sager et al. 1980; Faber 2012; Sanz Vicente 2012; Nakov 2013). English CNs are expressions with a head noun preceded by a modifying element (i.e. nouns or adjectives) (Levi 1978). However, given the demand for the translation of scientific and technical texts (Krüger 2015: 40), CNs need to be adapted to the patterns of term formation of the target language. For example, in Spanish, the second most spoken language in the world by native speakers and learners (Instituto Cervantes 2016), CN structure is reversed. In other words, the nominal head is postmodified by adjectives or prepositional phrases.

These phraseological units often pose problems at different levels. Firstly, their identification can be difficult, because they often include general words and can be formed by many constituents. Once they are identified, accessing their meaning is not an easy task since they are composed of juxtaposed concepts whose semantic relation is not explicit. Additionally, in CNs formed by more than two terms, bracketing or disambiguating their internal structure is a necessary step (e.g. [*wind turbine*][*power curve*] and [*wind power output*] *fluctuation*). As previously mentioned, the translation or production of these phraseological units in another language can also be problematic given the different patterns of term formation, such as premodification in English and postmodification in Spanish. The complex nature of CNs thus highlights the need to include them in specialized resources, especially because they are a relevant part of conceptual systems (Sager et al. 1980; Sager 1990; Sanz Vicente 2012). However, phraseological units in general are not usually recorded in specialized resources and their treatment is not systematic (L'Homme and Pimentel 2012; Xu 2013; Buendía Castro 2013). Up until now, research on CNs has focused mostly

on two-term CNs formed by nouns, which are usually referred to as 'noun compounds', but has not considered other types of CNs, namely those including adjectives and other parts of speech.

In view of the complex nature of CNs, this article describes the representation of these phraseological units in EcoLexicon (www.ecolexicon.ugr.es), an environmental knowledge base which is the practical application of Frame-Based Terminology (Faber 2012). CNs will be part of the new phraseological module of EcoLexicon, and will provide different access points to information (Kwary 2012), such as the CNs formed from a given term, a bilingual view in English and Spanish, and the syntactic-semantic combinations in CNs. This research presents the preliminary design of two of these views, namely the formation of CNs from a specific term and the bilingual view, thus providing solutions for some of the difficulties of these multi-word terms. The representation of CNs in EcoLexicon is thus based on their semantics, which is the starting point in the specification of mapping rules and also facilitates knowledge acquisition (Faber 2012). For the purposes of the study, we used a wind power corpus in English and Spanish for term extraction, semantic analysis (by means of knowledge patterns, verb paraphrases and free paraphrases), establishment of interlinguistic correspondences, and definition crafting, since corpora have been proved to be essential in lexicographic and terminographic work (Huang et al. 2016).

Our preliminary results showed that this analysis of CNs provided valuable insights into the formation of these phraseological units, which should be considered if they are to be usefully included in linguistic resources. Furthermore, according to other surveys assessing the usability, functionality, and efficiency of terminographic resources (López-Rodríguez et al. 2012; León-Araúz and Reimerink 2018), users were found to appreciate the inclusion of conceptual relations, equivalents and synonyms, phraseology, contexts, and definitions, as proposed in this study. Nevertheless, an evaluation of the CN module of EcoLexicon will be carried out once all the views will be implemented, with a view to assessing the opinion of potential users.

The reminder of this article is organized as follows. Section 2 provides a description of the characteristics of CNs in specialized discourse. Section 3 focuses on Frame-Based Terminology and its practical application, EcoLexicon. In Section 4, we show the preliminary representation of CNs in the phraseological module of EcoLexicon. Section 5 explores CN formation in English and Spanish. Finally, Section 6 presents the conclusions that can be derived from this study as well as our plans for future research.

2. Specialized complex nominals

Different types of phraseological unit include idioms, collocations, and CNs. However, in specialized discourse, the separation of the different combinations has often been questioned since both collocations and CNs provide relevant information for the conceptual structuring of a specialized domain (Meyer and Mackintosh 1996). Accordingly, in Frame-Based Terminology (§3) 'terminological phrasemes' are specialized phraseological units which include collocations and CNs (Buendía Castro 2013).

In particular, CNs (e.g. *global warming*) are frequently used to designate specialized concepts in English (Sager et al. 1980; Faber 2012; Sanz Vicente 2012; Nakov 2013). This is not surprising because morphologically poor languages, such as English, usually create CNs[1] by adding nominal or adjectival pre-modifiers to a head noun (e.g. *waste management*). In Romance languages, such as Spanish, the modifiers are placed on the right of the nominal head and are often adjectives or prepositional phrases (Fernández-Domínguez 2016: 67) (e.g. *gestión de residuos*). These interlinguistic differences highlight the need for knowledge-based resources, namely in specialized translation, because effective knowledge acquisition is vital for rendering a term into another language (Faber 2012). This is particularly relevant in the case of CNs, whose semantic content must be adapted to the term formation rules in the target language.

[1] We refer to endocentric CNs, which are the focus of this study and the most frequent type of CNs in specialized texts (Nakov 2013).

CNs can be regarded as nodes of compressed knowledge. They combine concepts of the terminological system to form one new concept and, thus, can be used to extract information regarding conceptual hierarchies (Sager et al. 1980; Sager 1990). In other words, CNs represent hyponymic concepts that are the result of the specification of the head, which is the hypernym, by means of the addition of other terms in the form of modifiers. For instance, when modifiers are added to *pollution*, hyponyms such as *oil pollution* and *water pollution* are created. This conceptual complexity is increased when more than two concepts are juxtaposed, as in *electrically-excited synchronous generator*. Long CNs are often difficult to identify and bracket or disambiguate (Utsumi 2014).

In our view, concepts are not randomly paired in CNs, but rather are the result of underlying semantic constraints (Warren 1978; Pinker 1989; Wisniewski 1997; Štekauer 1998; Kageura 2002; Rosario et al. 2002; Maguire et al. 2010). This alludes to micro-contexts, which are essential to our semantic analysis. To understand the notion of 'micro-context', the head of a CN can be considered to open slots that are filled by specific conceptual categories (Wisniewski 1997; Rosario et al. 2002; Maguire et al. 2010). These categories have a semantic role[2], which is a key factor in the formation of CNs that are hyponyms of the head. For instance, *erosion* opens two slots, the first related to the entities that *cause* erosion, and the second related to those that are *affected* by it. The slot for the causes of erosion is filled by semantic categories such as WATERBODY or SUBSTANCE, which have the semantic role of AGENT (as in *sea erosion* and *chemical erosion*). Alternatively, the slot for the entities affected by erosion is filled by categories such as LANDFORM or LAND, which have the role of PATIENT (e.g. *dune erosion* and *soil erosion*). By opening these slots, the meaning of the head determines which concepts can specify it. Therefore, the micro-context of a CN includes such conceptual information by means of this slot-filling mechanism and is essential to the understanding of CNs.

In addition to this juxtaposition of concepts, the complex semantic nature of CNs is accentuated by the omission of the semantic relation between their constituents (Vanderwende 1994; Nakov 2013; Ó Séaghdha and Copestake 2013; *inter alia*). This becomes especially problematic in structurally similar CNs (e.g. in *oil pollution*, the pollution is *caused_by* oil, whereas in *water pollution*, the pollution *affects* the water) (Cabezas-García and León-Araúz 2018). Moreover, in long CNs different semantic relations can be established between internal groups. For example, in [[*power generation*] *system*], the semantic relations are 'system *has_function* power generation' and 'generation *has_result* power'. Thus, the meaning of CNs is not fully compositional (Utsumi 2014; Smith et al. 2014). The only information that can be derived from the structure is that it denotes something (conveyed by the head) that is somehow related to the modifiers (Jespersen 1942; Smith et al. 2014: 100). Consequently, in the formation of CNs, the principle of formal economy is prioritized over semantic transparency, especially in English (Fernández-Domínguez 2016). Therefore, for users to understand CNs, it is necessary for them to have access to knowledge underlying their structure.

In this respect, the Generative Lexicon theory (Pustejovsky 1995) explains how compositionality contributes to lexical semantics. This theory has been used in various studies on CNs and their interpretation. For example, Johnston and Busa (1999) used qualia structure to explain the compositional interpretation of CNs, Bouillon et al. (2012) provided an annotation scheme for CNs based on the Generative Lexicon, Rallapalli and Paul (2012) presented a hybrid approach for the interpretation of CNs using an ontology, Bassac and Bouillon (2013) used the Generative Lexicon to study the telic (purpose) relation in CNs, and Yadav et al. (2017) applied it to the study of the semantic relations in CNs.

Concealed semantic relations between CN constituents are frequently made explicit by means of sets of semantic relations (e.g. *cause*, *result*). For instance, in *offshore wind farm*, the wind farm *is located* offshore (Vanderwende 1994; Barker and Szpakowicz 1998; Nastase and Szpakowicz 2003; Girju et al. 2005; Ó Séaghdha and Copestake 2013; *inter alia*). Most of these relations are based on general language, except for Rosario et al. (2002), who focused on the biomedical domain. Nevertheless, the use

[2] The set of semantic roles in EcoLexicon largely corresponds to those in the CREST implementation (Nirenburg 2000), the list of roles given by EAGLES (1996), and the inventory proposed by Gildea and Jurafsky (2002) (see an example in Figure 7). However, we are currently in the process of revising this roleset, as well as the inventory of semantic categories that so far had been designed ad-hoc.

of such inventories has been often questioned since there is no consensus as to the best set of relations or their partial semantic representation. (For example, the relation *affects* does not specify how something is affected.) In addition, different relations can be assigned to the same CN (e.g. a *museum book* can be a book *located at* a museum or a book published/*effected by* a museum). Sometimes, the meaning of a CN may not correspond to any relation at all (Nakov and Hearst 2013: 7; Hendrickx et al. 2013). For these reasons, authors such as Jespersen (1942) and Downing (1977) argued that inventories of semantic relations were not suitable for conveying the semantics of CNs.

In line with Finin (1980), who proposed the use of verbs to characterize CNs, Nakov and Hearst (2006) propose the use of paraphrases involving verbs and/or prepositions (e.g. a *bronze statue* is a statue made of/handcrafted from/dipped in bronze). Such verbs are better able to capture semantic features that relations cannot (Nakov and Hearst 2013: 3). Our semantic analysis of CNs combines semantic relations and paraphrases, because more abstract semantic relations can be further specified by means of the specific verbs elicited in paraphrases (Cabezas García and Faber 2017b). Paraphrases represent the sentential structure of CNs and evoke the CN formation processes in Levi (1978). These are predicate deletion (*a cyclone* originates over *the tropics* > *tropical cyclone*) and predicate nominalization[3] (*the weather* is predicted > *weather prediction*). They point to the existence of concealed propositions in the formation of CNs, as evidenced in the paraphrases. Thus, the study of CNs entails the analysis of their underlying predicates, which necessarily involves addressing their argument structure (Faber and Mairal 1999). Argument structure alludes to the specification of the number of arguments that a predicative unit (typically verbs, but also nominalizations) can take, their syntactic expression, and their semantic relation to the predicate. It plays a role in cross-linguistic correspondence since similar argument structures in different languages are regarded as a sign of linguistic equivalence (De Clerk et al. 2013; Buendía Castro and Faber 2016).

As previously mentioned, CNs are very frequent in specialized discourse, but they pose different problems related to their identification, their bracketing, and their underlying semantics (Lauer and Dras 1994). Additionally, their translation in another language can also be problematic, given the differences in term formation patterns. This is especially the case in the language pair English-Spanish. Such difficulties highlight the need to represent CNs in linguistic resources.

Nevertheless, phraseological units are rarely included in specialized resources (L'Homme and Pimentel 2012; Buendía Castro 2013). Even when CNs are addressed, the focus is on two-term CNs formed by nouns, whereas CNs including adjectives and other parts of speech tend to be disregarded. Furthermore, the treatment of CNs in lexicographic and terminographic resources is not systematic (Cabezas-García and Faber 2017a). Even when they are included, they are rarely defined (e.g. *Vocabulaire et cooccurrents de la comptabilité* [Caignon 2001]).

Other resources include argument structure but lack further details to clarify the meaning of the concept, as in *sedimentation*, which is represented as *sedimentation of particle* (*Dictionnaire fondamental de l'environnement. DiCoEnviro* [Observatoire de Linguistique Sens-Texte 2018]). Furthermore, CNs are often listed alphabetically (e.g. *Dictionary of Energy* [Cleveland and Morris 2015]). However, a representation of domain structure should reflect the relations of the CN with other terms (e.g. *Elsevier's Dictionary of Medicine Spanish-English English-Spanish* [Hidalgo 2014]). Some dictionaries represent equivalent CNs (and terms in general) in different entries with their own definitions, which do not reflect the relation between terms that represent the same concept in different languages (e.g. *Commercial Trucking Bilingual Dictionary*. *Diccionario Bilingüe de Transporte Comercial* [Moya 2004]).

There are also resources that include the CN only as sublemma of the head term, which is abbreviated, as in *p. septicémica* for *peste septicémica* (*septicemic plague*) (*Diccionario Etimológico de Medicina* [Segura 2004]). Other dictionaries show the modifiers and their possible heads in different lines, instead of including the entire CN (e.g. *Diccionario técnico inglés-español español-inglés* [Beigbeder 2006]). As for Romance languages, namely Spanish, there is a lack of specialized resources

[3] Predicate nominalization was also addressed by Halliday (1985) in his theory of the 'grammatical metaphor'.

that are regularly updated, which are essential in scientific and technical domains (López et al. 2010). For these reasons, there is a need for specialized resources that can successfully deal with the different problems that CNs can pose and facilitate the understanding and use of these phraseological units.

3. Frame-Based Terminology and EcoLexicon

This research follows Frame-Based Terminology, a cognitive approach to terminology that links specialized knowledge representation to cognitive linguistics in general and cognitive semantics in particular (Faber 2012). FBT is based on the Lexical Grammar Model (Martín Mingorance 1989; Faber and Mairal 1999), a lexical theory that focuses on the extraction and representation of conceptual and collocational relations in specialized discourse. FBT also combines premises of the Generative Lexicon (Pustejovsky 1995), which allows the organization and restriction of conceptual dimensions based on the semantics of concepts (León-Araúz 2009: 26). Finally, FBT adopts premises of Frame Semantics (Fillmore 1985, 2006), namely the notion of 'frame'. Knowledge is organized in frames (Minsky 1975; Fillmore 1985, 2006), which are cognitive structuring devices based on experience that provide the background knowledge and motivation for the existence of words in a language as well as the way those words are used in discourse (Faber 2009: 123). Moreover, frames make the semantic and syntactic behavior of terms explicit by means of the description of conceptual relations and terms' combinations (Faber 2009).

FBT focuses on: (1) conceptual organization; (2) the multidimensional nature of terminological units; and (3) the extraction of semantic and syntactic information through the use of multilingual corpora (Faber 2009: 123-124). This methodology is applied in the development of EcoLexicon (www.ecolexicon.ugr.es), a multilingual terminological knowledge base on environmental science that is the practical application of FBT. It was first implemented in 2003 and now includes 3,631 concepts and 20,342 terms in English, Spanish, German, French, Russian, Dutch, and Modern Greek. In line with digital lexicography and terminography (De Schryver 2003; Li 2005; Yamada 2013; Tono 2014), EcoLexicon has a visual interface with different modules that provide conceptual, linguistic, and graphical information, which can be individually selected by users (San Martín et al. 2017).

Furthermore, EcoLexicon has a phraseological module under construction, which is being redefined to also include CNs. As shown in Figure 1, verb collocations are provided in the form of the predicates that frequently combine with a given term (e.g. *hurricane* in Figure 1) (Buendía Castro 2013). The phraseological module is based on the semantic constraints that limit the combination of arguments (Pinker 1989; Buendía Castro 2013). Thus, it focuses on the conceptual structuring of predicates and its arguments. Predicative units are organized in the lexical domains of the Lexical Grammar Model (Faber and Mairal 1999), which encompass verbs sharing the same nuclear meaning and syntax (for example, the CHANGE domain includes verbs such as *affect* or *damage*). Alternatively, verb arguments are assigned semantic categories (e.g. NATURAL DISASTER or ATMOSPHERIC DISTURBANCE), which are generalizations of terms (Buendía Castro 2013: 376). The CN module, currently under construction, will be based on this conceptual organization with a view to providing a wide range of information regarding CNs, such as interlinguistic correspondences, CNs formed from a given term, syntactic and semantic combinations, etc. Frame-like representations, such as EcoLexicon, are an effective solution for including this phraseological information together with conceptual information (L'Homme and Robichaud 2014).

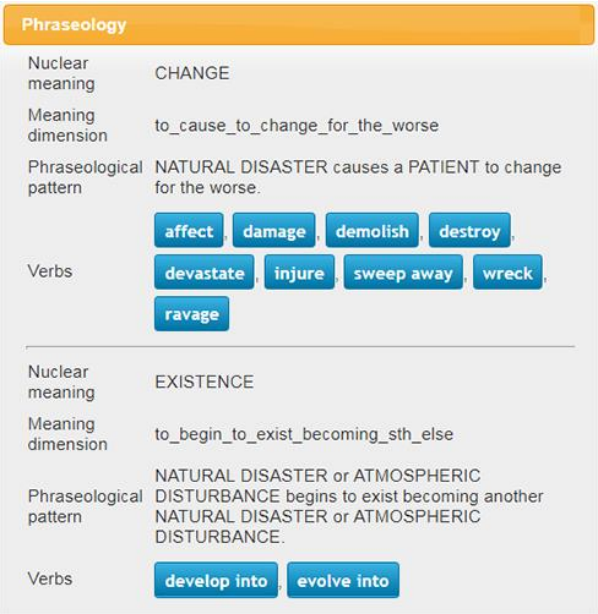


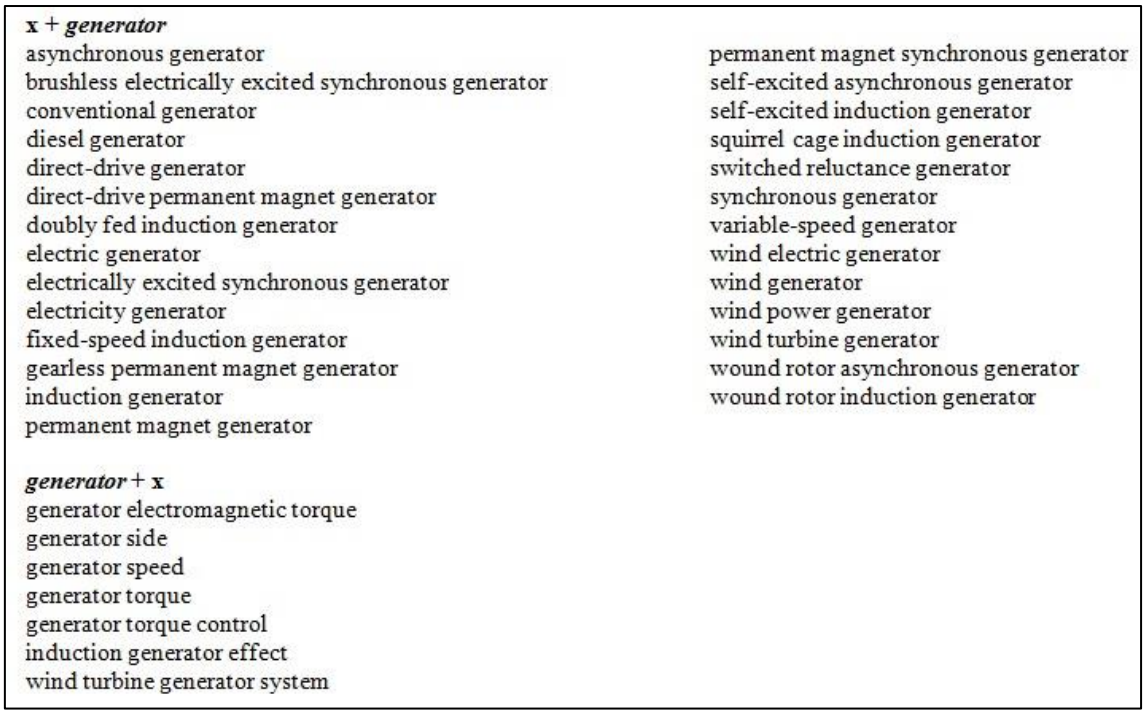
Fig. 1 Extract of the phraseological information of *hurricane*

4. Representation of complex nominals in the phraseological module of EcoLexicon

4.1. Extraction and semantic analysis of English complex nominals

For the purpose of the study, we manually compiled a corpus on wind power in English and Spanish of approximately 1.8 million words in each language. It was composed of highly specialized texts, namely journal articles and PhD dissertations. The English corpus was uploaded to the term extractor TermoStat (http://termostat.ling.umontreal.ca/) (Drouin 2003) in order to obtain a list of the most frequent single terms in the corpus from which CNs could be formed. We focused first on English, given its status of the *lingua franca* of specialized communication, from which texts are usually translated into other languages such as Spanish. It was found that *generator* was a very frequent term (15th of 1882 terms) that gave rise to the formation of many CNs.

The corpus was then uploaded to Sketch Engine (www.sketchengine.co.uk/) (Kilgarriff et al. 2004, 2014), a corpus analysis tool allowing CN extraction and semantic analysis using different procedures, as shall be seen. The word sketches in Sketch Engine, which show a term's combinatorial potential, provided an overview of the terms that usually co-occurred with *generator* (e.g. *induction generator* and *generator torque*). We then performed CQL queries to extract CNs whose head was *generator* as well as those CNs in which *generator* was a modifier. The CQL formalism (Schulze and Christ 1996) allowed sophisticated queries based on regular expressions combined with POS-tags. For example, the following sequence was queried to extract CNs whose head was *generator*, and which could be pre-modified by nouns, adjectives and/or adverbs: [tag="N.*|JJ.*|RB.*"]{1,}[lemma="generator"]. Figure 2 shows a list of the 34 CNs extracted, with *generator* as head or modifier, which had a minimum frequency of ten occurrences in different texts of the corpus.



x + *generator*

asynchronous generator
brushless electrically excited synchronous generator
conventional generator
diesel generator
direct-drive generator
direct-drive permanent magnet generator
doubly fed induction generator
electric generator
electrically excited synchronous generator
electricity generator
fixed-speed induction generator
gearless permanent magnet generator
induction generator
permanent magnet generator
permanent magnet synchronous generator
self-excited asynchronous generator
self-excited induction generator
squirrel cage induction generator
switched reluctance generator
synchronous generator
variable-speed generator
wind electric generator
wind generator
wind power generator
wind turbine generator
wound rotor asynchronous generator
wound rotor induction generator

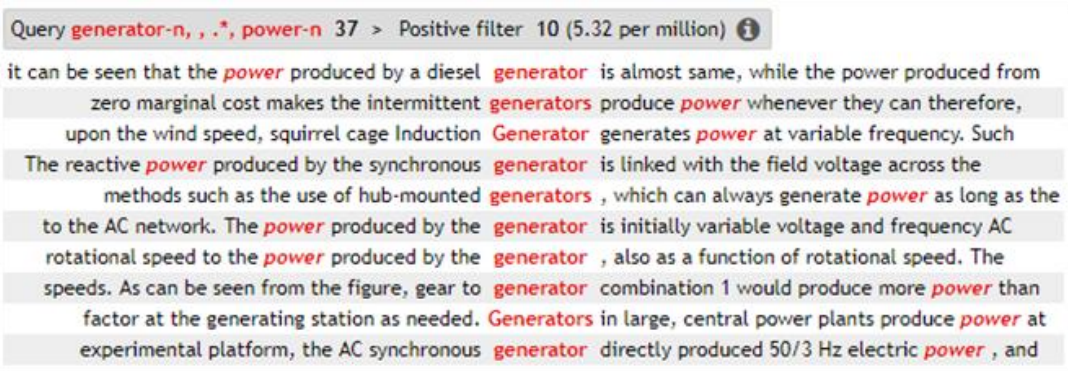
***generator* + x**

generator electromagnetic torque
generator side
generator speed
generator torque
generator torque control
induction generator effect
wind turbine generator system

Fig. 2 English CNs extracted

In addition, we included synonyms of these CNs, which were identified in the concordance analysis, during documentation in online texts and websites on wind power, and by means of synonymic knowledge patterns in the corpus query (e.g. *also called*, *referred to as*). Furthermore, long CNs, such as *permanent magnet synchronous generator*, were found to be usually hidden in the form of abbreviations (*PMSG*). Thus, CQL queries were also performed to find abbreviations.

Once the CNs were extracted, their semantics was accessed by means of a three-stage procedure involving knowledge patterns, verb paraphrases and free paraphrases. Knowledge patterns (KPs) are lexico-syntactic patterns that usually convey semantic relations in real texts (Meyer 2001; Marshman 2006). For instance, a well-known KP for indicating generic-specific relations is *X is a type of Y* (e.g. *wind power is a type of renewable energy*). KPs were used in the form of the 56 KP-based sketch grammars in León-Araúz et al. (2016), which allow the extraction of some of the most common semantic relations used in terminology: generic-specific, part-whole, location, cause and function. Most of the KPs used for retrieving these relations were not domain-specific, except for KPs such as *built for* or *built with*, which would only be found in construction-related domains (León-Araúz et al. 2016). Figure 3 shows an extract of the results of the query that targets the sentences annotated as word sketches between *generator* and *power* (e.g. to access the semantic relation in *wind power generator*), where *ws* means word sketch; "generator-n" and "power-n" are the terms annotated as part of a word sketch in the corpus; and "\"%w\".*" means any relation defined in the KP-based sketch grammars. As can be seen, these KPs reveal that the function of generators is power production.



Query generator-n, , .*, power-n 37 > Positive filter 10 (5.32 per million)

it can be seen that the *power* produced by a diesel	generator	is almost same, while the power produced from
zero marginal cost makes the intermittent	generators	produce *power* whenever they can therefore,
upon the wind speed, squirrel cage Induction	Generator	generates *power* at variable frequency. Such
The reactive *power* produced by the synchronous	generator	is linked with the field voltage across the
methods such as the use of hub-mounted	generators	, which can always generate *power* as long as the
to the AC network. The *power* produced by the	generator	is initially variable voltage and frequency AC
rotational speed to the *power* produced by the	generator	, also as a function of rotational speed. The
speeds. As can be seen from the figure, gear to	generator	combination 1 would produce more *power* than
factor at the generating station as needed.	Generators	in large, central power plants produce *power* at
experimental platform, the AC synchronous	generator	directly produced 50/3 Hz electric *power* , and

Fig. 3 Extract of the query results for KPs between *generator* and *power*:
[ws("generator-n","\"%w\" .*","power-n")]

Verb paraphrases (Nakov and Hearst 2006, 2013) were also used to further characterize the semantics of CNs by means of their concealed predicates. To this end, we performed CQL queries that elicited the verb linking the constituents of the CN. Figure 4 shows that the concealed predicates in *variable speed generator* are *operate* or *run*, which allude to the operation of the generator at variable speed.

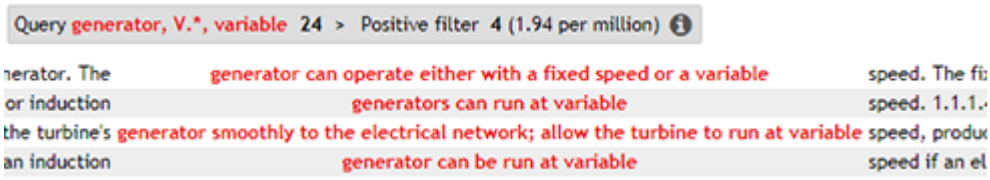


Fig. 4 Verb paraphrases for *variable speed generator*, obtained with the following CQL query:
[lemma="generator"][]{0,10}[tag="V.*"][]{0,10}[lemma="variable"] within <s/>

However, the frequent omission of constituents in CNs often complicates the extraction of verb paraphrases, because some of the constituents linked by the concealed verbs are often not specified in the CN. For this reason, free paraphrases (i.e. co-occurrences of the constituents of a CN in a sentence) were used as a final step in the semantic analysis of CNs. Figure 5 shows free paraphrases for *synchronous generator*, a CN that alludes to the synchronous speed of the rotor and stator of the generator.

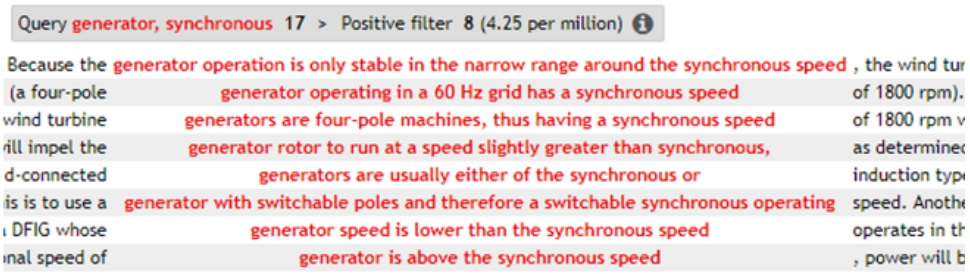


Fig. 5 Free paraphrases for *synchronous generator*, obtained with the following CQL query:
[lemma="generator"][]{1,10}[word="synchronous"][lemma!="generator"] within <s/>

In our analysis of CNs semantic relations are further specified by means of paraphrases, to provide a more complete representation of CN meaning (Cabezas-García and Faber 2017b).

4.2. Identification of Spanish equivalents

To identify the Spanish correspondences of English CNs, the head of the CN was first translated in order to ascertain the head of the Spanish term. Accordingly, *generator* was translated as *generador* in Spanish. This was confirmed by consulting specialized resources and verifying this correspondence in Spanish renewable energy texts. After identifying the most frequent modifiers of *generador* in the word sketches, a CQL query was performed to extract the structures that can co-occur with *generador* in Spanish. The query was the following:
[lemma="generador"][tag="A.*"]?[tag="S.*"]?[tag="N.*"]?[tag="S.*"]?[tag="N.*"]?[tag="A.*"]?. This query targets CNs whose head is *generador* ([lemma="generador"]), which can be postmodified by adjectives ([tag="A.*"]?) or prepositional phrases including nouns and also adjectives ([tag="S.*"]?[tag="N.*"]?[tag="S.*"]?[tag="N.*"]?[tag="A.*"]?). Figure 6 shows a list of the 34 Spanish CNs extracted, with *generador* as head or modifier, which had a minimum frequency of ten occurrences in different texts of the corpus.

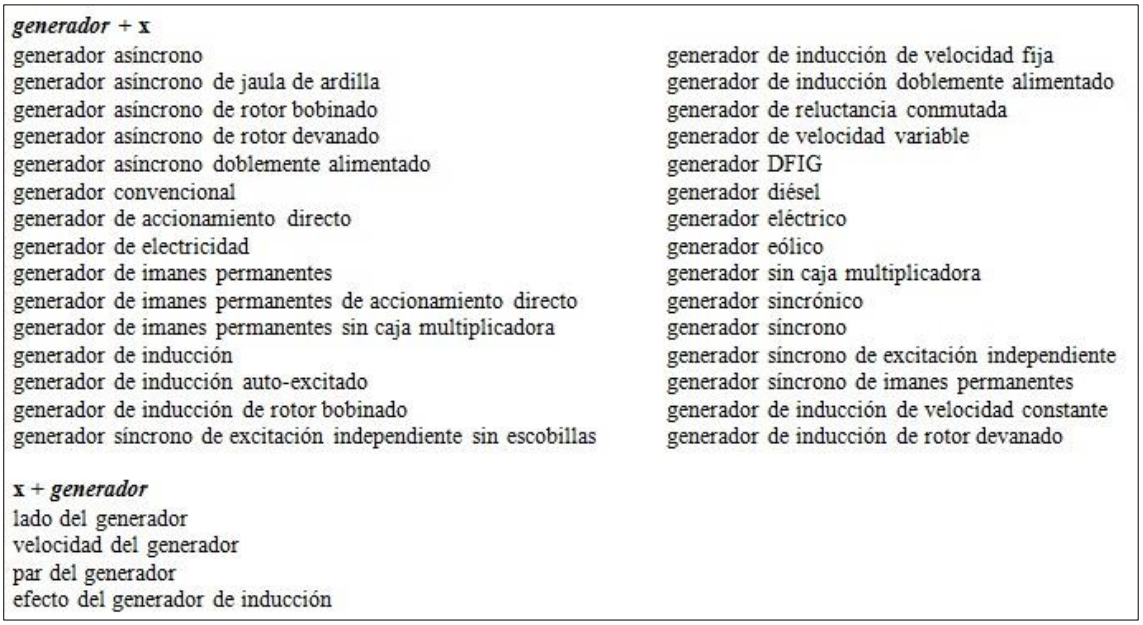


Fig. 6 Spanish CNs extracted

Since CNs are often abbreviated, CQL queries were also used to find abbreviations in Spanish. Finally, synonyms were identified in the concordance analysis, during documentation, and by means of synonymic KPs in the corpus query. This made it possible to extract equivalents that did not have the same length as the source term (e.g. *synchronous generator* can be translated as *generador síncrono* or *alternador* [*alternator*]). This problem is known as 'fertility' and is not often considered in bilingual terminology extraction (Daille et al. 2004). The synonyms showed that the target term was not always composed of the translations of the constituents of the source term (Daille et al. 2004). For example, *wind electric generator* is translated as *generador eólico*, which does not specify that the generator is *electric*. The procedure was replicated for those CNs in which *generator* was a modifier of the head. Spanish CNs were considered to be term candidates when they appeared at least ten times in different texts of the corpus.

After extracting the Spanish CNs, cross-linguistic correspondences were established between English and Spanish multi-word terms. Spanish CNs were semantically analyzed by means of KPs[4], verb paraphrases, and free paraphrases (§4.1) to verify that they designated the same concept as their English counterpart. Micro-contexts were essential to establish interlinguistic correspondences, because similar micro-contexts (i.e. slot-filling of the head of the CN by specific semantic categories and roles) were found in English and Spanish equivalents. Figure 7 shows the micro-contexts of *electrically excited synchronous generator* and its Spanish equivalent, *generador síncrono de excitación independiente*. The slots opened by *generator* and *generador* allude to the speed of the generator's rotor and stator and its excitation. Both are filled by the same semantic categories (ATTRIBUTE and EXCITATION, respectively), which have the same semantic roles (SPEED and STIMULUS). Nevertheless, CNs designating the same concept (in the same language or in different ones) can emphasize different characteristics of the concepts. This is the case of *electrically excited synchronous generator* and *generador síncrono de excitación independiente*. The English CN alludes to the electrical current that provides the rotor magnetization (*electrically*), whereas the Spanish CN highlights the fact that this current is provided by an independent machine (*independiente*). Thus, micro-contexts can be used to establish mapping rules, even when equivalence is not so evident.

[4] Since KP-based sketch grammars have not been developed for Spanish yet, our analysis of Spanish CNs by means of KPs was based on the translation of the KPs in the 56 English sketch grammars in León-Araúz et al. (2016) and their use in the form of CQL queries.

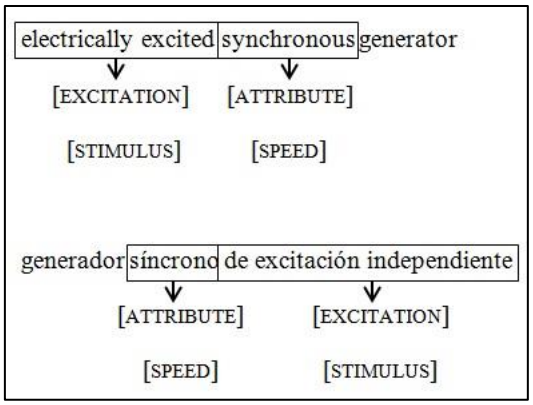


Fig. 7 Micro-contexts of *electrically excited synchronous generator* and *generador síncrono de excitación independiente*

4.3. Definition crafting

One of the problems of CNs in lexicographic and terminographic resources is that they are not often defined even though the juxtaposition of components makes their meaning far from transparent. In this regard, FBT proposes the use of definitional templates (Faber et al. 2001) (see Table 1), which specify the semantic relations that a certain category usually establishes. This is conducive to homogeneous definitions and the conceptual organization of terms. As shown in Table 1, hyponyms inherit the properties of the superordinate concept and add new specific values to them (e.g. the operating speed of the rotor and stator of an induction generator).

generator	
IS_A	machine
HAS_PART	rotor, stator
HAS_FUNCTION	convert rotational mechanical power into electrical power

induction generator	
IS_A	generator
HAS_PART	rotor magnetic field that operates at faster speed than that of the stator magnetic field

Table 1 Definitional templates for *generator* and *induction generator*

Thus, the semantic information previously elicited by means of KPs, verb paraphrases and free paraphrases was applied to these templates. This led to definitions based on the classical structure of *genus* and *differentiae*. Although many terminological models do not take argument structure into account (L'Homme and Robichaud 2014), the definitions in our study included the argument structure of predicative terms which should be considered when defining processes (Mel'čuk et al. 1995; Faber and Mairal 1999).

4.4. Complex nominals in the phraseological module of EcoLexicon

As previously mentioned, EcoLexicon will soon have a new phraseological module. It will include verb collocations as well as CNs. In line with current trends in lexicography and terminography, this CN module will offer different access points to the information (Kwary 2012), such as (i) the CNs formed from a given term; (ii) bilingual correspondences in English and Spanish (with the other languages in

EcoLexicon being subsequently implemented); (iii) syntactic combinations; and (iv) semantic combinations. This study focuses on the "Modifiers + Head" view and the "EN-ES" view.

Users will be able to access the CN module either through a specific tab in the main interface of EcoLexicon or through a term entry. This will give them the phraseological information associated with that term. Figure 8 shows the design of the information that is first presented to users in the Modifiers + Head view of the CN module, taking the entry of *generator* as an example. Even though the tabs are shown in English, users can change the language at any time during their query in the phraseological module. They can also switch to a different view (e.g. syntactic or semantic combinations).

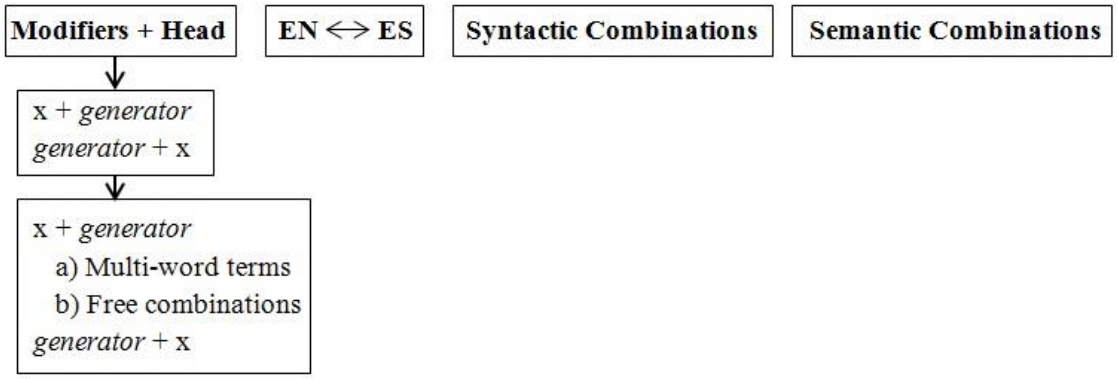


Fig. 8 Information initially presented to users in the Modifiers + Head view of the CN module

As shown in Figure 8, users must first decide which of the four views they prefer. In the Modifiers + Head view, they can obtain the CNs formed from a given term (e.g. *fluctuation* > *air pressure fluctuation*, *bathymetric fluctuation*, *beach fluctuation*, etc.). The EN-ES view focuses on translations (e.g. *tidal power* [EN]; *energía mareomotriz* [ES]). The Syntactic Combinations section offers the parts of speech that form CNs (e.g. N+N+N > *sea level rise*, *bed shear stress*, etc.). Finally, the Semantic Combinations view provides co-occurrences of semantic categories, roles and relations (e.g. RESOURCE+ENERGY > *wind power, wave power, tidal power*, etc.). The next step involves choosing whether the given term (*generator*) is a head or modifier in the CN. Then, users can focus on multi-word terms[5] or free combinations, or both. In the bilingual view, the specification of the position of the term in the CN is preceded by the choice of the directionality in translation (EN-ES or ES-EN). Then, the selected type of information is shown (see the preliminary interface of multi-word terms in the Modifiers + Head view in Figure 9).

[5] Multi-word terms are distinguished from free combinations because they designate a concept that is included in the conceptual system of the domain and establishes semantic relations with other concepts in the semantic network. They are often formed by means of the systematic slot filling of their micro-contexts by specific semantic categories and roles (e.g. *wind energy*, *wave energy*, *solar energy*, etc.); they can be replaced by abbreviations, and usually have a higher frequency than free combinations. Free combinations (e.g. *conventional generator*) are shown as a set of frequent combinations, without definitions or direct interlinguistic equivalences, because they do not represent concepts of the domain conceptual system and, thus, they can vary in different languages.

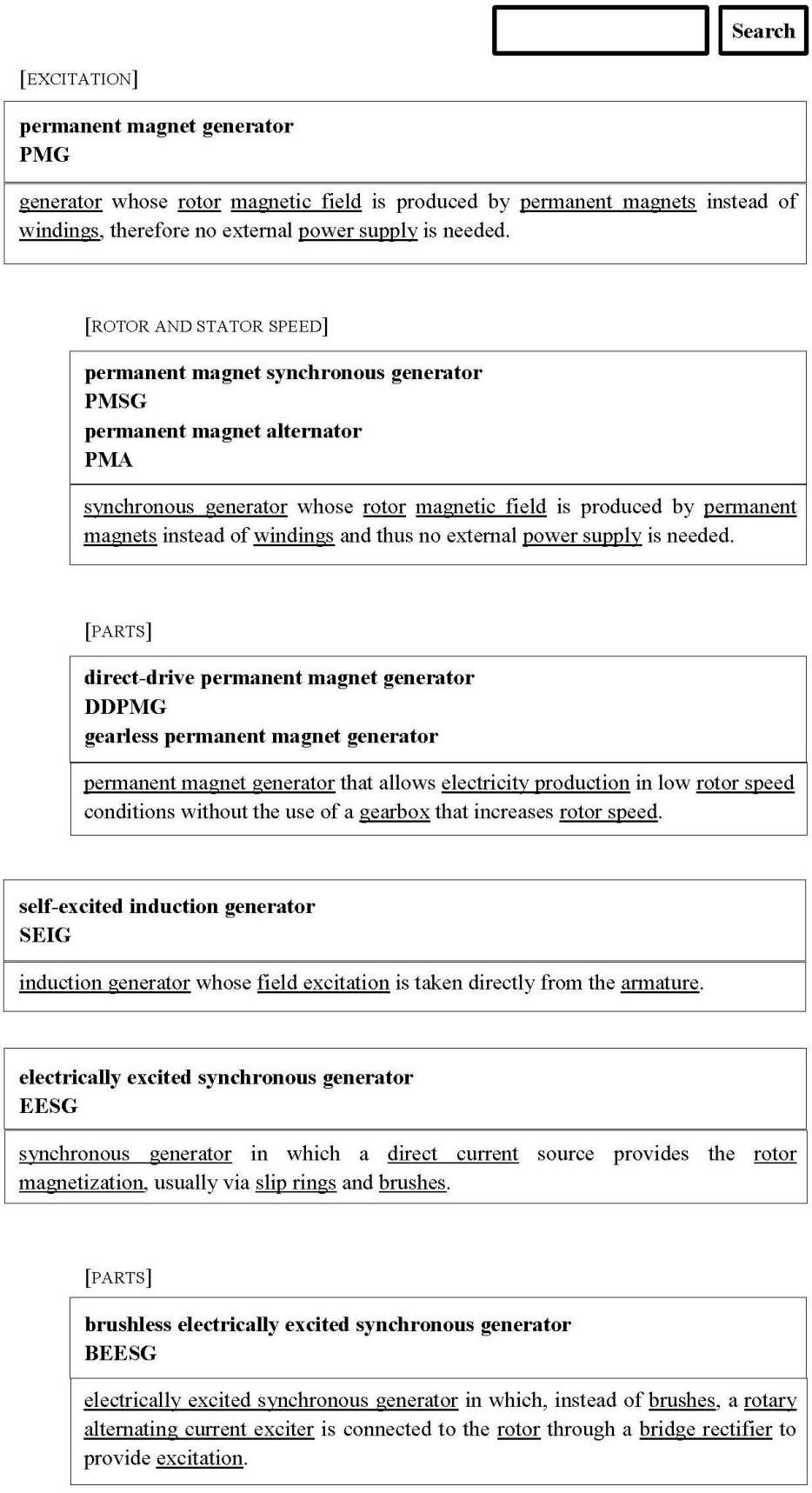


Fig. 9 Extract of the Modifiers + Head view of the multi-word terms of *generator*

As previously stated, the Modifiers + Head view focuses on the CNs formed from a given term, which are defined and semantically organized. This semantic organization is present in all the views of

the CN module and is based on the conceptual dimensions evoked by these phraseological units. For instance, depending on their excitation, generators can be permanent magnet generators, self-excited induction generators or electrically excited synchronous generators. These conceptual dimensions are listed based on the frequency of the CNs activating them. The CNs in each dimension are also organized based on their frequency (e.g. *permanent magnet generator* is more frequent than *electrically excited synchronous generator*).

Hyponyms are indented and preceded by the dimension activated (e.g. *permanent magnet synchronous generator* highlights the dimensions of EXCITATION and ROTOR AND STATOR SPEED). Because of their multidimensionality, terms can emphasize different dimensions of the same concept (Kageura 1997). This is frequently the case in CNs, which can either focus on one dimension, or combine several, as in the example of *permanent magnet synchronous generator*, which alludes to its excitation with permanent magnets and the speed of its rotor and stator, which is synchronous. As a result, CNs that combine several dimensions usually have more than one hypernym and are thus included under all of them (e.g. *permanent magnet synchronous generator* is indented under *permanent magnet generator* and *synchronous generator*), with a view to showing the different conceptual hierarchies that can be established.

As shown in Figure 9, each term appears with its synonyms. Even though CNs have a high level of variation (Cabezas-García and Faber 2017c), the CN module in EcoLexicon only includes abbreviations and those synonyms that have sufficient frequency and a linguistic form that is significantly different from the CN in question. Furthermore, the definitions show hyperlinks of the terms in the database, which allow users to access their term entry by clicking on the term or to visualize its definition in a floating window.

Figure 9 also shows the search box in the upper right corner of all the views of the CN module. This box facilitates the retrieval of a CN as well as a proximity search that shows the closest CNs. The proximity search can be very helpful given the identification problems often posed by CNs. For instance, when a user enters the search "excited generator", the following CNs containing those terms appear: *self-excited induction generator*, *electrically excited synchronous generator*, *brushless electrically excited synchronous generator*.

However, for the translations of CNs, users should choose the bilingual view. This view currently shows English and Spanish correspondences, but it will be extended to include the other languages in EcoLexicon. Figure 10 shows an extract of the bilingual view of the CN module of EcoLexicon.

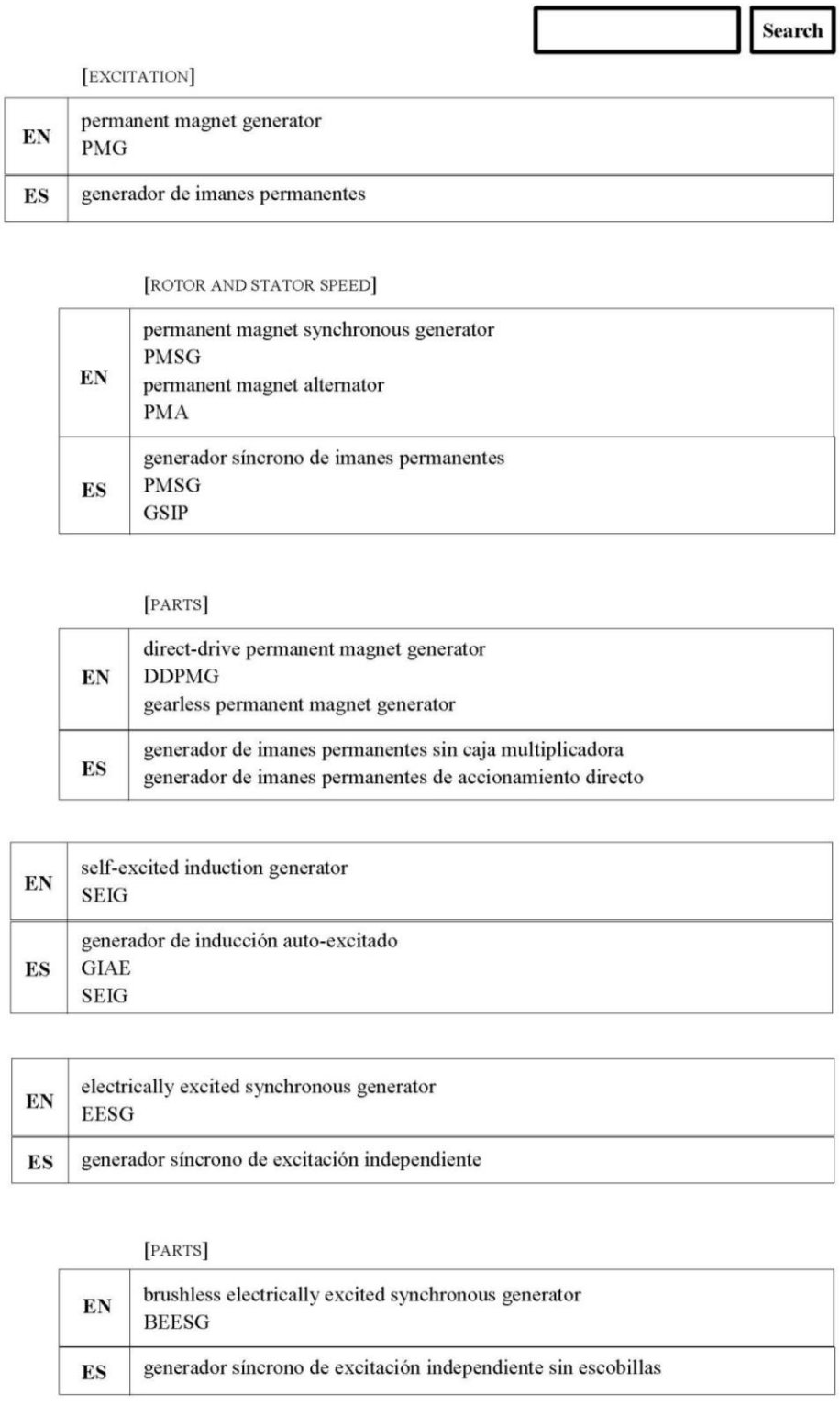


Fig. 10 Extract of the EN-ES view of the multi-word terms of *generator*

The CN module of EcoLexicon offers additional functionalities. The main options available in all the views are the following: (i) internal semantic relations between the constituents of the CN; (ii) usage examples; (iii) verb collocations that represent the same phraseological pattern; (iv) notes; and (v) the term entry in EcoLexicon. The non-specification of the semantic relations between the constituents of CNs is one of the main difficulties of these phraseological units. For example, in *direct-drive permanent magnet generator*, *direct drive* alludes to the gearbox that the generator does not have and thus refers to a *part of* the generator, which *is excited by* permanent magnets. Usage examples are also

provided, which allow the visualization of the CN in real contexts. As for verb collocations, the connection between the CN and the verb collocation sections of the phraseological module will be based on micro-contexts. In other words, there will be an association of CNs and collocations whose predicates are organized in the lexical domains of Faber and Mairal (1999) and complemented by arguments filled by specific semantic categories and roles. For example, the CN *erosion control structure* and the verb collocation *a structure controls erosion* follow the same phraseological pattern: a CONSTRUCTION [INSTRUMENT] *controls* [ACTION] an ENVIRONMENTAL PROCESS [PATIENT], which also gives rise to other more specific CNs and collocations, such as *groyne erosion* and *groynes control beach erosion*. The CN module also includes usage notes, such as the widespread use of an English abbreviation in Spanish (e.g. *PMSG* [*permanent magnet synchronous generator*], more frequent in Spanish than *GSIP* [*generador síncrono de imanes permanentes*]) or the polysemy of certain terms (e.g. *wind generator* can designate either a wind turbine or one of its parts, namely the generator that converts rotational mechanical power into electrical power). Finally, the term entry in EcoLexicon can also be accessed, where definitions, translations, conceptual networks, images, etc. are provided. As for the bilingual view, access to definitions is also offered in these secondary options since the main interface of this view focuses on English and Spanish equivalences. Figure 11 illustrates the additional functionalities of the CN module, taking the Modifiers + Head view as an example.

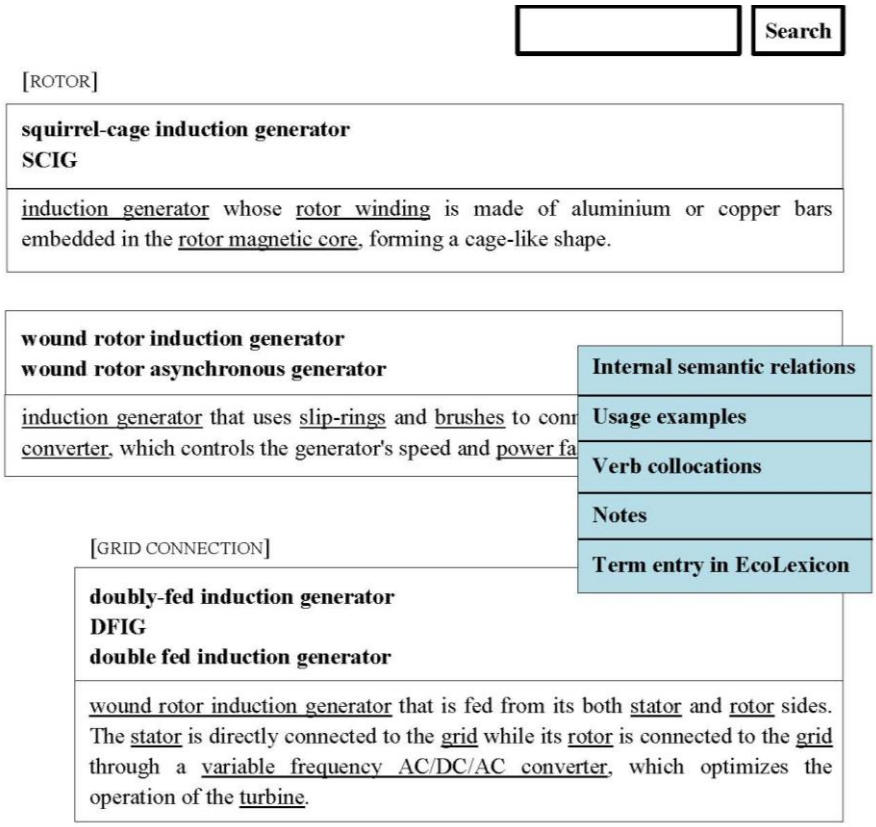


Fig. 11 Additional functionalities of the Modifiers + Head view

The design of the CN module of EcoLexicon presented in this section can help in cognitive situations, i.e. when users need encyclopedic knowledge related to language, specialized language, culture or any specific subject field (L'Homme and Leroyer 2009: 269), as well as in communicative situations, i.e. when they need dictionary assistance in some textual activity, such as reading or revising a text, translating a source text into a target text language or writing a text in the mother tongue or in a foreign language (L'Homme and Leroyer 2009: 270). Thus, users of different profiles can access terminographic resources such as EcoLexicon in these situations. They can be subject specialists, professional communication mediators (e.g. technical writers, translators, and interpreters), lexicographers and terminologists, information and documentation specialists, language planners, professional language users (e.g. publishers, language teachers), linguistic engineering and artificial intelligence professionals, and

laypeople (Sager 1990; Cabré 1999). Consequently, EcoLexicon users can be said to be English or Spanish native speakers, who need to perform environment-related tasks with different degrees of expertise for cognitive or communicative purposes (López-Rodríguez et al. 2012; León-Araúz and Reimerink 2018).

Thanks to the inclusion of different types of information, EcoLexicon meets the needs of these potential users that must understand or produce texts in English or Spanish, despite not being native speakers of either of these languages, a situation that is the order of the day in scientific communication (Faber 2012). In fact, the user evaluation of EcoLexicon (López-Rodríguez et al. 2012) showed that conceptual relations, equivalents and synonyms, phraseology, contexts, and definitions were among the most useful information types for translators. Conceptual organization, as shown in the CN module, was also preferred to alphabetically-ordered information since it contributed to knowledge acquisition (López-Rodríguez et al. 2012). Other surveys (Durán-Muñoz 2010; León-Araúz and Reimerink 2018) obtained similar results and underlined the usefulness of including CNs in knowledge resources, as well as their abbreviations and acronyms. In fact, León-Araúz and Reimerink (2018) argue that the search options should be improved in terminographic resources, an aspect that has been taken into account in the different types of query proposed. For these reasons, the design of the CN module of EcoLexicon and the information that it provides will be valuable for users of this resource since it takes into account the cognitive and communicative situations in which they may be involved.

5. Complex nominal formation in English and Spanish

In English, specialized concepts are frequently designated by CNs (Sager et al. 1980; Faber 2012; Sanz Vicente 2012; Nakov 2013). These phraseological units pose a wide range of problems affecting translators, language for specific purposes students (Kernerman 2007; Ding 2018), and experts, who wish to publish their scientific articles in English (Sanz Vicente 2012). However, there is a lack of studies and specialized resources addressing CNs, especially when it comes to dealing with languages other than English (Smith et al. 2014). Neither has the formation and internal structure of CNs been a focus of attention (Sanz Vicente 2012), especially in the case of CNs formed by more than two constituents and those including adjectives and other parts of speech. The CNs in this study allowed us to provide a preliminary design of the CN module in EcoLexicon and afforded insights into the formation of these phraseological units in English and Spanish.

The English CNs were formed by a nominal head, which was premodified not only by nouns but also by adjectives, and to a lesser extent, by adverbs and participles. Namely, 18 out of the 39 English CNs were premodified only by nouns (e.g. *induction generator*); almost the same number of CNs (16) were also modified by adjectives (e.g. *wind electric generator*); and 5 CNs included adverbs and/or past participles among the modifiers (e.g. *electrically excited synchronous generator*). This underlines the need to consider CNs formed not only by nouns, but also by adjectives, adverbs and participles. Alternatively, the heads of the Spanish CNs were postmodified by prepositional phrases or adjectives. In 21 out of the 42 Spanish CNs, the head was postmodified by a prepositional phrase mostly introduced by the preposition *de* [*of*] and usually followed by nouns, adjectives, and sometimes by adverbs and participles (e.g. *generador de inducción de rotor bobinado* [*wound rotor induction generator*]). On the other hand, in 19 CNs the head was postmodified by adjectives, which in some cases were followed by prepositional phrases (e.g. *generador asíncrono de jaula de ardilla* [*squirrel cage induction generator*]).

Regarding the number of constituents, the English CNs were found to be mainly formed by two (12 CNs), three (12 CNs), or four components (13 CNs), although there were also 2 CNs formed by five components. As for Spanish, the CNs were formed by two (14 CNs), three (5 CNs), four (9 CNs), five (6 CNs), or six components (6 CNs), with 2 CNs formed by seven components. These longer structures in Spanish are not surprising, because instead of the noun packing typical of Germanic languages, CNs in Romance languages have adjectival and prepositional postmodification.

Accordingly, CNs are formed from underlying predicates (Levi 1978) that can acquire different forms in each language or even in the same language. For instance, *electrically excited synchronous*

generator has two concealed predicates, namely *excite* (*excited*) and *generate* (*generator*). However, in its Spanish counterpart, *generador síncrono de excitación independiente*, the verb *excitar* [*excite*] does not appear as a past participle, but rather as a nominalization (*excitación* [*excitation*]). For this reason, equivalences must be based on meaning rather than form (Buendía Castro and Faber 2016). This means that micro-contexts are vital to establish interlinguistic correspondences and the formation of CNs.

In micro-contexts, the head of a CN can be regarded as having an argument structure (Rosario et al. 2002). More specifically, it opens slots that are filled by specific conceptual categories (Wisniewski 1997; Rosario et al. 2002; Maguire et al. 2010) that play a semantic role. These slots are perceivable in the definition of the head concept. When these slots are filled, this gives rise to the formation of CNs that make the meaning of the head more specific. For instance, an electrically excited synchronous generator is a "synchronous generator in which a direct current source provides the rotor magnetization, usually via slip rings and brushes". Namely one of its frequent parts, the brushes, give rise to the formation of its hyponym *brushless electrically excited synchronous generator*, which indicates the lack of this component.

Multidimensionality or the different perspectives from which the characteristics of a concept are usually specified (Kageura 1997) plays a relevant role in the formation of CNs (Cabezas-García and Faber 2017a). Therefore, the concepts designated by these phraseological units can emphasize different characteristics or dimensions. For instance, both *wound rotor induction generator* and *variable speed generator* can designate the same concept, although the former refers to its rotor and the latter alludes to its operation at variable wind speeds. Multidimensionality is also at the origin of long CNs, which in our sample were composed of up to five constituents in English and seven in Spanish. These long CNs often stem from the combination of dimensions, which is related to the specification of the micro-context. For example, *brushless electrically excited synchronous generator* refers to the rotor and stator speed in the generator (*synchronous*), its excitation (*electrically excited*), and the lack of a component (*brushless*). As can be seen, these longer strings are especially complex, because their internal structure must be identified in order to elicit the semantic relations among the different groups.

Long CNs are usually condensed in the form of abbreviations (e.g. *PMSG*, *DDPMG*, *SCIG*, etc.). Namely, 30% of the English term candidates[6] and 23% of the Spanish set of terms were abbreviations. However, the abbreviations used in Spanish did not always stem from the Spanish CNs. There were often English abbreviations in Spanish (such as *SCIG*, which stands for *squirrel cage induction generator*, and is used instead of *generador asíncrono de jaula de ardilla*), some Spanish terms were formed from a combination of English and Spanish constituents (e.g. *generador DFIG* [*doubly-fed induction generator*]), and in some cases the abbreviations in both languages coexist in Spanish (as in *generador de inducción auto-excitado* [*self-excited induction generator*], which is abbreviated either as *GIAE* or *SEIG*). These particularities, which are recorded in the CN module of EcoLexicon, are a sign of the instability of CNs. The frequent variation of these phraseological units is evidenced in their high number of synonyms, which were particularly based on morphosyntactic permutations. In English, 19 out of the 21 concepts analyzed had more than one variant, with up to nine denominations in some cases (see Figure 12). As for Spanish, 15 out of the 21 concepts were designated by more than one denomination, although some of them had up to thirteen designations (see Figure 12). These synonyms often entail minimal changes, such as hyphens, but they can also have significantly different forms, as evidenced in the synonyms and equivalents of *wind turbine generator system* (Figure 12):

[6] For the elaboration of these percentages, the sets of term candidates were augmented to 56 and 52 terms in English and Spanish, respectively, instead of the previous amounts of 39 and 42, which included CNs extracted by means of CQL queries and their synonyms, but did not take abbreviations into account with a view to exploring the number of constituents and parts of speech of the full forms of CNs.

EN	wind turbine generator system WTGS wind turbine wind generator windmill wind energy conversion system WECS wind machine aerogenerator
ES	aerogenerador turbina eólica generador eólico sistema eólico WECS molino de viento turbina de viento aeroturbina sistema de conversión de energía eólica sistema de energía eólica AG sistema de generación eólica máquina eólica

Fig. 12 Synonyms and Spanish equivalents of *wind turbine generator system*

This marked variation is often regarded as a sign of neology (Cabré 1999), which is also indicated by the frequent use of calques in Spanish (Cabezas-García and Faber 2017c). Some examples of calques are *generador de inducción doblemente alimentado* [*doubly-fed induction generator*] and *generador asíncrono de jaula de ardilla* [*squirrel cage induction generator*], which maintain the metaphor alluding to the form of the rotor. In conclusion, it has been shown that exploring the formation of CNs is an essential step towards their analysis and representation in linguistic resources, such as the CN module that we have presented in this paper.

6. Conclusions

Phraseological units play a major role in general language and specialized discourse because of their high frequency (Sinclair 2000) and the difficulties that they usually cause to non-native users (Benson et al. 2009; Inoue 2014). In English, the *lingua franca* of communication (Tono 2014), specialized concepts are usually conveyed by means of CNs (Sager et al. 1980; Faber 2012; Sanz Vicente 2012; Nakov 2013). However, these phraseological units, which are characterized by their syntactic-semantic complexity, are not usually recorded in specialized resources and their treatment is not systematic (Cabezas-García and Faber 2017a).

This article has described the representation of CNs in EcoLexicon (www.ecolexicon.ugr.es), a multilingual terminological knowledge base that focuses on environmental science. CNs will be incorporated into the phraseological module of EcoLexicon, which already includes verb collocations and will provide different access points to information (Kwary 2012). This research focuses on two of the views of the CN module, namely the formation of CNs from a specific term and the bilingual view. For that purpose, we used a wind power corpus in English and Spanish for term extraction, semantic analysis (by means of knowledge patterns, verb paraphrases and free paraphrases), establishment of interlinguistic correspondences, and definition crafting.

The Modifiers + Head view presented in this article focuses on the CNs formed from a given term (e.g. *generator* > *synchronous generator*, *induction generator*, *wind generator*, *diesel generator*, etc.), which are defined and semantically organized (Figure 9). This semantic organization is present in all the views of the CN module and is based on the conceptual dimensions evoked by these phraseological units (e.g. ROTOR, EXCITATION, GRID CONNECTION). On the other hand, the bilingual view prioritizes

translations (e.g. *switched reluctance generator* [EN]; *generador de reluctancia conmutada* [ES]) (Figure 10). Furthermore, secondary options are available in all the views, namely the internal semantic relations between the constituents of the CN, usage examples, verb collocations that represent the same phraseological pattern, notes, and the term entry in EcoLexicon, as well as definitions in the bilingual view.

Our main objective was to structure the representation of CNs in EcoLexicon based on their semantics, which is the basis for the establishment of mapping rules and facilitates knowledge acquisition (Faber 2012). This led to the study of CN formation in English and Spanish, which is vital to the semantic analysis and linguistic representation of CNs. Thus, premodification patterns were found to be prevalent in English, while postmodification was preferred in Spanish. CNs were composed of a varying number of constituents: from two to five in English, and from two to seven in Spanish, due to the fact that Romance languages do not admit noun packing. Furthermore, CNs were found to be formed from underlying predicates (Levi 1978), which are directly linked to micro-contexts. These are essential in the establishment of interlinguistic correspondences and the formation of CNs. Multidimensionality played an important role in CN formation, especially in the development of long CNs, which are often condensed in the form of abbreviations. Finally, CNs usually had different synonyms to name the same concept. This high instability has often been regarded as a sign of neology (Cabré 1999; Cabezas-García and Faber 2017c).

Different users can benefit from this proposal of CN representation, such as specialists and semi-experts in the environmental domain, and language professionals and students (Kernerman 2007; Ding 2018) since the design of the CN module and the information provided have been found to be especially valued by users of terminographic resources (López-Rodríguez et al. 2012; León-Araúz and Reimerink 2018). Moreover, it can be applied to both general and specialized dictionaries and databases in different languages, as well as to individual words. In future research the views regarding the syntactic and semantic combinations in CNs will be further developed. One of our research focuses will be recurrent semantic patterns in CN formation (e.g. SPEED + MACHINE: *variable speed wind turbine, fixed speed wind turbine, variable speed generator, fixed speed generator*) and their usefulness in the inference of semantic relations. The slot-filling of micro-contexts and their relevance for establishing correspondences will be assessed using inter-annotator agreement. Once all the views of the CN module are implemented in EcoLexicon, another evaluation of the resource by its potential users will be carried out following previous evaluations of EcoLexicon (López-Rodríguez et al. 2012) and of other related terminographic resources (León-Araúz and Reimerink 2018). For this purpose, different groups of subjects (e.g. translators, domain experts, etc.) will be assigned a cognitive or communicative task related to the environment and will be asked to use EcoLexicon and, in particular, the CN module to complete the assignment. Then, they will be asked to fill out a questionnaire on their user profiles and give their opinion of the tool's usability, functionality and efficiency (as highlighted in the ISO 9128 standard for software product evaluation), as well as problems encountered and possible improvements, with a view to enriching the CN module of EcoLexicon.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

Bally, Charles. 1951. *Traité De Stylistique Française*, 3rd ed. Paris: Librairie C. Klincksieck.

Barker, Ken, and Stan Szpakowicz. 1998. Semi-automatic recognition of noun modifier relationships. In Proceedings of the 17th International Conference on Computational Linguistics, ICCL '98, 96–102.

Bassac, Christian, and Pierrette Bouillon. 2013. The Telic Relationship in Compounds. In *Advances in Generative Lexicon Theory*, eds. James Pustejovsky, Pierrette Bouillon, Hitoshi Isahara, Kyoko Kanzaki, and Chungmin Lee, 109-126. Dordrecht: Springer.

Beigbeder, Federico. 2006. *Diccionario técnico inglés-español español-inglés*, 2nd ed. Madrid: Díaz de Santos.

Benson, Morton, Evelyn Benson, and Robert Ilson. 2009. *The BBI Combinatory Dictionary of English*, 3rd ed. Amsterdam/Philadelphia: John Benjamins.

Bouillon, Pierrette, Elisabetta Jezek, Chiara Melloni, and Aurélie Picton. 2012. Annotating Qualia Relations in Italian and French Complex Nominals. In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012), 1527-1532.

Buendía Castro, Miriam, and Pamela Faber. 2016. Phraseological correspondence in English and Spanish Specialized Texts. In *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives = Fraseología computacional y basada en corpus: perspectivas monolingües y multilingües*, ed. Gloria Corpas Pastor, 391-398. Geneva: Tradulex.

Buendía Castro, Miriam. 2013. Phraseology in Specialized Language and its Representation in Environmental Knowledge Resources. PhD Thesis. Granada: University of Granada.

Cabezas-García, Melania, and Pamela Faber. 2017a. A Semantic Approach to the Inclusion of Complex Nominals in English Terminographic Resources. In *Computational and Corpus-Based Phraseology*. ed. Ruslan Mitkov, 145-159. Cham: Springer.

Cabezas-García, Melania, and Pamela Faber. 2017b. Exploring the Semantics of Multi-word Terms by Means of Paraphrases. In *Temas actuales de Terminología y estudios sobre el léxico*, eds. Miguel Ángel Candel-Mora, and Chelo Vargas-Sierra, 193–217. Granada: Comares.

Cabezas-García, Melania, and Pamela Faber. 2017c. The role of micro-contexts in noun compound formation. *Neologica* 11: 101-118.

Cabezas-García, Melania, and Pilar León-Araúz. 2018. Towards the Inference of Semantic Relations in Complex Nominals: a Pilot Study. In Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018), 2511-2518.

Cabré, María Teresa. 1999. *Terminology: Theory, methods and applications*. Amsterdam/Philadelphia: John Benjamins.

Caignon, Philippe. 2001. *Vocabulaire et cooccurrents de la comptabilité*. Montréal: Linguatech.

Cheng, Winnie, Chris Greaves, John McH. Sinclair, Martin Warren. 2008. Uncovering the Extent of the Phraseological Tendency: Towards a Systematic Analysis of Concgrams. *Applied Linguistics* 30(2): 236–252.

Cleveland, Cutler, and Christopher Morris. 2015. *Dictionary of Energy*, 2nd ed. Amsterdam: Elsevier.

Daille, Béatrice, Samuel Dufour-Kowalski, and Emmanuel Morin. 2004. French-English multi-word term alignment based on lexical context analysis. In Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 04), 919-922.

De Clerck, Bernard, Timothy Colleman, and Dominique Willems. 2013. Introduction: A multifaceted approach to verb classes. *Linguistics* 51(4): 663-680.

De Schryver, Gilles-Maurice. 2003. Lexicographers' dreams in the electronic-dictionary age. *International Journal of Lexicography* 16(2): 143-199.

Ding, Jun. 2018. 'A study of Chinese medical students as dictionary users and potential users for an online medical termfinder.' *Lexicography* 3(2): 115-136. https://doi.org/10.1007/s40607-017-0031-9

Downing, Pamela. 1977. On the creation and use of English compound nouns. *Language* 53: 810–842.

Drouin, Patrick. 2003. Term extraction using non-technical corpora as a point of leverage. *Terminology* 9: 99–115.

Durán Muñoz, Isabel. 2010. Specialized lexicographical resources: a survey of translators' needs. In *eLexicography in the 21st century: New Challenges, new applications. Proceedings of ELEX2009. Cahiers du Cental. Vol. 7*, eds. Sylviane Grander and Magali Paquot, 55-66. Louvain-La-Neuve: Presses Universitaires de Louvain.

EAGLES (Expert Advisory Group On Language Engineering). 1996. *Text Corpora Working Group Reading Guide*. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale.

Faber, Pamela, and Ricardo Mairal Usón. 1999. *Constructing a Lexicon of English Verbs*. Berlin: Mouton de Gruyter.

Faber, Pamela, Clara Inés López Rodríguez, and Maribel Tercedor Sánchez. 2001. Utilización de técnicas de corpus en la representación del conocimiento médico. *Terminology* 7(2): 167–198.

Faber, Pamela. 2009. The Cognitive Shift in Terminology and Specialized Translation. *MonTI. Monografías de Traducción e Interpretación* 1: 107-134.

Faber, Pamela. 2012. *A cognitive linguistics view of terminology and specialized language*. Berlin/Boston: De Gruyter Mouton.

Fernández-Domínguez, Jesús. 2016. A morphosemantic investigation of term formation processes in English and Spanish. *Languages in Contrast* 16(1): 54–83.

Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di Semantica* 6(2): 222-254.

Fillmore, Charles J. 2006. Frame Semantics. In *Cognitive Linguistics. Basic readings*, ed. Dirk Geeraerts, 373-400. Berlin/Boston: De Gruyter.

Finin, Timothy. 1980. The Semantic Interpretation of Compound Nominals. PhD Thesis. Urbana: University of Illinois.

Gildea, Daniel, and Daniel Jurafsky. 2002. Automatic Labeling of Semantic Roles. *Computational Linguistics* 28: 245-288.

Girju, Roxana, Dan Moldovan, Marta Tatu, and Daniel Antohe 2005. On the semantics of noun compounds. *Computer Speech and Language* 19(4): 479–496.

Halliday, Michael A. K. 1985. *An introduction to functional grammar*. London: Arnold.

Hendrickx, Iris, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Stan Szpakowicz, and Tony Veale. 2013. SemEval-2013 Task 4: Free Paraphrases of Noun Compounds. In Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), 138–143.

Hidalgo, Ana. 2014. *Elsevier's Dictionary of Medicine Spanish-English English-Spanish*. Amsterdam: Elsevier.

Huang, Chu-Ren, Li Lan, and Su Xinchun. 2016. Lexicography in the Contemporary Period. In *The Routledge Encyclopedia of the Chinese Language*, ed. Chan Sin-Wai, 545-562. Abingdon/New York: Routledge.

Inoue, Ai. 2014. Newly observed phraseological units with noun forms of modal verbs. *Lexicography* 1(2): 137–157.

Instituto Cervantes. 2016. El español: una lengua viva. Informe 2016. http://www.cervantes.es/imagenes/File/prensa/EspanolLenguaViva16.pdf. Accessed 19 January 2018.

Jespersen, Otto. 1942. *A Modern English Grammar on Historical Principles*, Vol. VI. Copenhagen: Munksgaard.

Johnston, Michael, and Federica Busa. 1999. Qualia structure and the compositional interpretation of compounds. In *Breadth and Depth of Semantic Lexicons*, ed. Evelyne Viegas, 167-187. Dordrecht: Springer.

Kageura, Kyo. 1997. A preliminary investigation of the nature of frequency distributions of constituent elements of terms in terminology. *Terminology* 4(2): 199–223.

Kageura, Kyo. 2002. *The Dynamics of Terminology: A Descriptive Theory of Term Formation and Terminological Growth*. Amsterdam: John Benjamins.

Kernerman, Ilan. 2007. What's so good or bad about advanced EFL dictionaries. In *Dictionary Visions, Research and Practice: Selected Papers from the 12th International Symposium on Lexicography, Copenhagen, 2004*, eds. Henrik Gottlieb, and Jens Erik Mogensen, 139-145. Amsterdam/Philadelphia: John Benjamins.

Kilgarriff, Adam, Pavel Rychlý, Pavel Smrž, and David Tugwell. 2004. The Sketch Engine. In Proceedings of the 11th EURALEX International Congress, 105–116.

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, and Vít Suchomel. 2014. The Sketch Engine: ten years on. *Lexicography* 1(1): 7–36.

Krüger, Ralph. 2015. *The Interface between Scientific and Technical Translation Studies and Cognitive Linguistics*. Berlin: Frank & Timme.

Kwary, Deny Arnos. 2012. Adaptive Hypermedia and User-Oriented Data for Online Dictionaries: A Case Study on an English Dictionary of Finance for Indonesian Students. *International Journal of Lexicography* 25(1): 30-49.

Li, Lan. 2005. The growing prosperity of on-line dictionaries. *English Today* 83(21): 16-21.

Lauer, Mark, and Mark Dras. 1994. A probabilistic model of compound nouns. In Proceedings of the 7th Australian Joint Conference on Artificial Intelligence, 474-481.

León-Araúz, Pilar, Antonio San Martín, and Pamela Faber. 2016. Pattern-based Word Sketches for the Extraction of Semantic Relations. In Proceedings of the 5th International Workshop on Computational Terminology (Computerm2016), 73–82.

León-Araúz, Pilar, and Arianne Reimerink. 2018. Evaluating EcoLexiCAT: a Terminology-Enhanced CAT Tool. In Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018), 2374-2381.

León-Araúz, Pilar. 2009. Representación multidimensional del conocimiento especializado: el uso de marcos desde la macroestructura hasta la microestructura. PhD Thesis. Granada: University of Granada.

Leroyer, Patrick. 2006. Dealing with Phraseology in Business Dictionaries: Focus on Functions — Not Phrases. *Linguistik Online* 27(2): 183–194.

Levi, Judith. 1978. *The Syntax and Semantics of Complex Nominals*. New York: Academic Press.

L'Homme, Marie-Claude, and Patrick Leroyer. 2009. Combining the semantics of collocations with situation-driven search paths in specialized dictionaries. *Terminology* 15(2): 258-283.

L'Homme, Marie-Claude, and Janine Pimentel. 2012. Capturing Syntactico-semantic Regularities among Terms: An Application of the FrameNet Methodology to Terminology. In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012), 262-268.

L'Homme, Marie-Claude, and Benoît Robichaud. 2014. Frames and Terminology: Representing Predicative Terms in the Field of the Environment. In Proceedings of the 4th Workshop on Cognitive Aspects of the Lexicon, 186–197.

L'Homme, Marie-Claude. 2009. A Methodology for Describing Collocations in a Specialized Dictionary. In *Lexicography in the 21st Century*, eds. Sandro Nielsen, and Sven Tarp, 237-256. Amsterdam/Philadelphia: John Benjamins.

López Rodríguez, Clara Inés, Pamela Faber, Pilar León Araúz, Juan Antonio Prieto Velasco, and Maribel Tercedor Sánchez. 2010. La Terminología basada en marcos y su aplicación a las Ciencias Ambientales: los proyectos MARCOCOSTA y ECOSISTEMA. *Arena Romanistica* 7(10): 52-74.

López Rodríguez, Clara Inés, Miriam Buendía Castro, and Alejandro García Aragón. 2012. User needs to the test: Evaluating a terminological knowledge base on the environment by trainee translators. *Jostrans. The Journal of Specialized Translation* 18: 57-76.

Maguire, Phil, Edward J. Wisniewski, and Gert Storms. 2010. A corpus study of semantic patterns in compounding. *Corpus Linguistics and Linguistic Theory* 6(1): 49-73.

Marshman, Elizabeth. 2006. Lexical Knowledge Patterns for Semi-automatic Extraction of Cause-effect and Association Relations from Medical Texts: A Comparative Study of English and French. PhD Thesis. Montréal: Université de Montréal.

Martín Mingorance, Leocadio. 1989. Functional Grammar and Lexematics. In *Meaning and Lexicography*, eds. Jerzy Tomaszczyk, and Barbara Lewandowska, 227-253. Amsterdam: John Benjamins.

Mel'čuk, Igor, André Clas, and Alain Polguère. 1995. *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot.

Meyer, Ingrid, and Kristen Mackintosh. 1996. Refining the Terminographer's Concept-analysis Methods: How Can Phraseology Help? *Terminology* 3(1): 1–26.

Meyer, Ingrid. 2001. Extracting knowledge-rich contexts for terminography: a conceptual and methodological framework. In *Recent Advances in Computational Terminology*, eds. Didier Bourigault, Christian Jacquemin, and Marie-Claude L'Homme, 279–302. Amsterdam/Philadelphia: John Benjamins.

Minsky, Marvin. 1975. A framework for representing knowledge. In *The Psychology of Computer Vision*, ed. Patrick Henry Winston, 211-277. New York: McGraw-Hill.

Moya, Maria Ivon. 2004. *Commercial Trucking Bilingual Dictionary. Diccionario Bilingüe de Transporte Comercial*. New York: Delmar Learning.

Nakov, Preslav, and Marti Hearst. 2006. Using Verbs to Characterize Noun-Noun Relations. *Artificial Intelligence Methodology Systems and Applications* 4183: 233–244.

Nakov, Preslav, and Marti Hearst. 2013. Semantic Interpretation of Noun Compounds Using Verbal and Other Paraphrases. *ACM Transactions on Speech and Language Processing* 10(3): 1-51.

Nakov, Preslav. 2013. On the interpretation of noun compounds: Syntax, semantics, and entailment. *Natural Language Engineering* 19(03): 291–330.

Nastase, Vivi, and Stan Szpakowicz. 2003. Exploring noun-modifier semantic relations. In Fifth International Workshop on Computational Semantics (IWCS-5), 285–301.

Nirenburg, Sergei. 2000. CREST: Progress Report. Working Paper, NMSU CRL. Presented at the DARPA TIDES PI Meeting.

Observatoire de Linguistique Sens-Texte. 2018. *Dictionnaire fondamental de l'environnement. DiCoEnviro*. http://olst.ling.umontreal.ca/cgi-bin/dicoenviro/search.cgi? Accessed 22 January 2018.

Ó Séaghdha, Diarmuid, and Ann Copestake. 2013. Interpreting compound nouns with kernel methods. *Natural Language Engineering* 19: 331-356.

Pinker, Steven. 1989. *Learnability and Cognition*. Cambridge: MIT Press.

Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge: MIT Press.

Rallapalli, Sruti, and Soma Paul. 2012. A hybrid approach for the interpretation of nominal compounds using ontology. In 26th Pacific Asia Conference on Language, Information and Computation, 554-563.

Rosario, Barbara, Marti Hearst, and Charles Fillmore. 2002. The Descent of Hierarchy, and Selection in Relational Semantics. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, ACL '02, 247–254.

Sager, Juan C., David Dungworth, and Peter F. McDonald. 1980. *English Special Languages. Principles and Practice in Science and Technology*. Wiesbaden: Brandstetter Verlag.

Sager, Juan C. 1990. *A practical course in terminology processing*. Amsterdam/Philadelphia: John Benjamins.

San Martín, Antonio, Melania Cabezas-García, Miriam Buendía, Beatriz Sánchez-Cárdenas, Pilar León-Araúz, and Pamela Faber. 2017. Recent Advances in EcoLexicon. *Dictionaries: Journal of the Dictionary Society of North America* 38(1): 96-115.

Sanz Vicente, Lara. 2012. Approaching secondary term formation through the analysis of multiword units: An English–Spanish contrastive study. *Terminology* 18 (1): 105–127.

Schulze, Bruno Maximilian, and Oliver Christ. 1996. *The CQP User's Manual*. Stuttgart: Universität Stuttgart.

Segura Munguía, Santiago. 2004. *Diccionario Etimológico de Medicina*. Bilbao: Universidad de Deusto.

Sinclair, John McH. 2000. Lexical Grammar. *Darbai Ir Dienos* 24: 191-205.

Smith, Viktor, Daniel Barratt, and Jordan Zlatev. 2014. Unpacking noun-noun compounds: interpreting novel and conventional food names in isolation and on food labels. *Cognitive Linguistics* 25(1): 99–147.

Štekauer, Pavol. 1998. *An Onomasiological Theory of English Word-Formation*. Amsterdam/Philadelphia: John Benjamins.

Tono, Yukio. 2014. Lexicography in Asia: its future and challenges. *Lexicography* 1(1): 1-5.

Utsumi, Akira. 2014. A semantic space approach to the computational semantics of noun compounds. *Natural Language Engineering* 20: 185-234.

Vanderwende, Lucy. 1994. Algorithm for automatic interpretation of noun sequences. In Proceedings of the 15th Conference on Computational Linguistics, COLING 1994, vol. 2, 782–788.

Warren, Beatrice. 1978. *Semantic Patterns of Noun-Noun Compounds*. Göteborg: Acta Universitatis Gothoburgensis.

Wisniewski, Edward J. 1997. When concepts combine. *Psychonomic Bulletin and Review* 4: 167–183.

Xu, Hai. 2013. Phraseology and English Phrase Inclusion and Arrangement in Dictionaries. *Lexicographical Studies* 2013(5): 50-56.

Yadav, Prabha, Elisabetta Jezek, Pierrette Bouillon, Tiffany J. Callahan, Michael Bada, Lawrence E. Hunter, and K. Bretonnel Cohen. 2017. Semantic relations in compound nouns: Perspectives from inter-annotator agreement. *Studies in Health Technology and Informatics* 245: 644-648.

Yamada, Shigeru. 2013. Overview of Hand-held Electronic Dictionaries in Japan: Functions, Usage, and Impact on Print Dictionary Industry. In *Multi-disciplinary Lexicography: Traditions and Challenges of the XXIst Century*, eds. Olga M. Karpova, and Faina I. Kartashkova, 158-165. Newcastle-upon-Tyne: Cambridge Scholars.