DE GRUYTER MOUTON DOI 10.1515/phras-2014-0004 — YoP 2014; 5: 57–94

Miriam Buendía Castro, Silvia Montero Martínez, Pamela Faber

# Verb collocations and phraseology in EcoLexicon[1]

**Abstract:** To satisfy the expectations of specialized audiences, translators are not only concerned with knowledge transfer, but also with the use of suitable terminological and phraseological units. Specialized resources with conceptual and linguistic information on a subject field are thus essential. This paper illustrates the treatment of verb collocations in Frame-based Terminology and the EcoLexicon knowledge base. Verbs are analyzed by studying their activation in texts and are classified in lexical domains and subdomains, according to the premises of the Lexical Grammar Model. Arguments are classified in sets of conceptual categories, along with their semantic roles and macro roles. Finally, verb-argument patterns associated with a specific category and domain are encoded in phraseological entries in EcoLexicon. English and Spanish verb collocations belonging to the EXTREME EVENT and the category NATURAL HAZARDS are used to illustrate the process.

**Keywords:** verb collocations; terminological knowledge bases; Frame based Terminology

**Correspondence address:** smontero@ugr.es

## 1 Introduction

Generally speaking, *phraseology* focuses on phrases, defined as "any multi-word expression up to sentence level" (Pawley 2001: 122). However, there is still no consensus regarding the limits (and denomination) of these units. More restricted views regard phraseological units as fixed, more or less idiomatic combinations of at least two words, which appear together and function as a noun, verb, adverb, or preposition (Corpas 2003: 134). From a broader perspective, other authors claim that phraseological units are all word combinations with a certain stability.

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

This approach was adopted by Hausmann (1984, 1985, 1989), Gläser (1994/95), and also by Roberts (1994/95), Heid (1994, 2001) and Montero (2003, 2008) for specialized discourse.

In a parallel way, approaches to collocations can be divided in semantically-based approaches and frequency-oriented approaches. The *semantically-based approach* (Mel'čuk et al. 1984–1999; Hausmann 1989; Benson et al. 1986, 2009) conceives collocations as binary units with a semantically-autonomous base or node and a semantically-dependent collocate. Since the components of the collocation are assumed to be directional, the selection of the collocate depends on the prior selection of the base (Hausmann 1989). In contrast, the *frequency-oriented approach* to collocation (Sinclair et al. 1970/2004) conceives collocations as statistically significant co-occurrences of two or more words. As such, collocations are based on frequency rather than on semantic characteristics.

In this paper, verb collocations in specialized discourse, a type of specialized phraseological unit (Meyer and Mackintosh 1994, 1996; Tercedor 1999), are regarded as highly frequent combinations of two or more words (Heid 1999; Siepmann 2005: 417), following a noun + verb or verb + noun pattern, where the noun is the *base* and the verb is the *collocate*. In this noun-centered type of collocation, the meaning of the collocate (verb) is modulated by the meaning of the base. However, at the same time, the collocate also constrains the number and type of arguments that combine with it. In other words, there is a process of co-selection (Martin 1992; Heid 2001). In addition, a certain degree of compositionality is assumed since each lexical unit in the collocation retains its meaning (L'Homme 2000: 106).

Although Cowie (1998: 225) argues that verb and noun complementation should be understood in terms of valence rather than collocation, constructional information for verbs are intimately related to collocation "since many noun-verb collocations require a specific distribution of semantic roles" (Siepmann 2005: 416). As such, in the collocations *hurricane damages* and *hurricane hits*, both elements maintain their respective meanings. Nonetheless, they are not free combinations since the predicates DAMAGE and HIT require noun phrases that designate entities, or processes with the role of AGENT and the potential to cause some kind of harm. In addition, *hurricane* takes predicates with the generic meaning of *to cause sth/sb to change for the worse* [DAMAGE, HIT]. Consequently, the semantic features encoded in the verb interact with or are in some degree modulated by those of the argument(s). This means that such combinations have a high frequency in texts that activate certain frames or situations typical of natural disasters and the environment.

This paper focuses on the theoretical approach and practical methodology used in Frame-based Terminology (FBT) (Faber 2009, 2012) for encoding and

describing verb collocations in terminographic resources (e.g. the EcoLexicon knowledge base), in which verbs and their arguments are classified in a set of conceptual categories, typical of a given specialized domain. This leads to the formulation of templates that generalize and predict this type of construction for entire semantic subdomains (Buendía 2013: 534). Such templates facilitate text comprehension and production. This process is illustrated within the environmental domain and the category NATURAL HAZARD.

# 2 Specialized translation and collocations

Translation is a process that begins when a client contacts a translator or translation bureau, and which ends when the client is satisfied with the product as the final result of the initial inquiry (Malmkjær 2000: 163). As such, professional translators must not only achieve accurate meaning transfer but also consider factors such as punctuality in delivery, adjustment to format specifications (Bonet 2002), and satisfaction of communicative expectations (Montero et al. 2001; Fuertes et al. 2005). To produce a good translation, specialized translators must be able to decode the source text, access its underlying knowledge structures, and encode a target text that is acceptable to the targeted group of receivers. This means understanding the concepts designated by different types of specialized lexical units, such as terms and terminological phrasemes (Montero et al. 2002). In FBT, following Meyer and Mackintosh (1994, 1996), terminological phrasemes are specialized phraseological units which include collocations and nominal compounds.

## 2.1 Conceptual and linguistic knowledge

Specialized texts codify communicative models arising from expert meaning configurations (Seibel and Jiménez 2002; Floros 2002). To a great extent, these models reflect the speakers' discursive positioning (Harré and Langenhove 1999) which involve a sender and receiver/s of different knowledge levels. Terminological and specialized phraseological units, such as terminological phrasemes, can also reflect different cognitive and cultural schemas (Cabré 1998: 51; Gläser 1998; Fuertes et al. 2005).

Consequently, specialized translators must select the linguistic choices that best reproduce cognitive schemas in order to fulfill communicative intentions and expectations. In other words, specialized translation requires at least three basic elements: (a) a good comprehension of the source language;

(b) knowledge of the specialized domain; (c) the skills needed to write like an expert in the field. Specialized translation is generally performed either by professional translators or by field experts. However, professional translators usually do not have the same level of domain-specific knowledge as the senders and receivers of the original text. As a result, they must rapidly learn to situate terms and specialized phraseology within their respective conceptual systems and, at the same time, increase their knowledge of the specialized domain (Montero and Faber 2009: 89).

Experts, on the other hand, have terminological and linguistic knowledge in their own language, but somewhat less knowledge of the foreign language (Bergenholtz and Nielsen 2002: 6). As such, terminographic tools that provide syntactic, semantic, and pragmatic information of terms and phraseology, as well as specialized conceptual knowledge are essential for translation purposes (Tercedor, López, and Faber 2012: 182).

## 2.2 Collocations in specialized resources

In specialized translation, collocations are extremely important for both text encoding and decoding. According to Rundell (2010: vii), collocations are as important as grammar since they make speakers/writers sound fluent. Unfortunately, few specialized resources actually contain word combinations (L'Homme and Leroyer 2009: 260), and even fewer include information regarding verb predicates and their argument structure (Buendía 2013: 23). This is paradoxical since verbs are the most important lexical and syntactic category in language (Lorente 2000; Lorente and Bevilacqua 2000).

An exception in this regard is Meaning Text Theory (MTT) (Mel'čuk et al., 1995; Mel'čuk 1984–1999). Its lexical component, the Explanatory Combinatorial Dictionary (ECD) provides a systematic description of lexemes or phrasemes which convey a specific meaning (L'Homme 2009: 271). The MTT also proposes an inventory of lexical functions (LFs). According to Mel'čuk (1996: 39), a lexical function *f* is a correspondence that associates a given lexical unit (L) (the argument or keyword, i.e. the base of the collocation) with a set of lexical items (L1) (the collocate) which express a specific meaning associated with *f*.

For instance, the lexical function *Magn* specifies for a noun *N* an adjective, or a word combination of adjective type, which expresses the meaning of great intensity or magnitude of the main quality of *N*. When applied to *hurricane*, the collocation designating maximum strength would vary, depending on the level of specialization. Accordingly, Magn (*hurricane*) = *powerful* is typical of general language and Magn (*hurricane*) = *category 5* of specialized language.

MTT also allows for the codification of predicate-argument structure in the form of *semantic actants*. Predicates are verbs and adjectives though nouns, especially deverbal and de-adjectival nouns, can also be predicates (L'Homme 2010: 142). Though actants can appear in different parts of the entry, they are most frequently stated in the definition and in the representation of the syntactic structure of lexical units.

Representation of collocations in specialized dictionaries follow different presentation modes, and, although there is an increasing number of resources that include these units, there is still no consensus as to which word combinations should be included and how they should be classified (Montero and Buendía 2012). Nevertheless, there is a certain agreement on the following (L'Homme 2009: 239):

- Collocations are listed under a headword that has been previously defined as a term in a specialized subject field;
- The keyword is usually a noun or a noun phrase;
- Lexical units that typically combine with terms can be verbs, nouns, or adjectives.

Buendía (2013: 201) also concludes that, in order to facilitate successful text encoding and decoding, collocational information in dictionary lemmas and termbase entries should follow these guidelines:

1. The configuration of phraseological information in entries should be theoretically based and conceived for both encoding and decoding purposes.
2. The format should be electronic without any difficult metalanguage.
3. The resource should include bilingual/multilingual correspondences for the phraseological units.
4. Users should be able to search and access collocations in various ways.
5. Entries should not contain an excessive number of collocations.
6. The classification of the information should be semantic.
7. The semantic and syntactic patterns associated with the meaning of each collocation should be accompanied by a description.
8. Each entry should contain various usage examples.

Accordingly, in FBT, the systematicity and coherence of verb-noun collocational information is based on theoretical insights from cognitive approaches to Terminology and linguistic models of predicate-argument structures. The phraseological entries and subentries are in consonance with translation needs and include conceptual and linguistic knowledge, a semantic and syntactic description of collocations, and usage examples. Finally, users have onomasiological and semasiological access to multilingual collocational information.

# 3 The EXTREME EVENT and the category of NATURAL HAZARD in ecolexicon

A basic premise in Frame-based Terminology (FBT) (Faber 2009, 2011, 2012) is that the configuration of specialized domains is based on events or frames. Each knowledge area can be said to have its own event template and can be represented accordingly (Grinev and Klepalchenko 1999).

In order to create these non-language-specific representations, FBT follows an integrated *top-down* and *bottom-up* approach. It combines principles of Corpus Linguistics, the Lexical Grammar Model (Faber and Mairal 1999), and Frame Semantics (Fillmore 1976, 1982, 1985; Fillmore and Atkins 1992). As is well known, frames are a cognitive structuring device, based on experience, which provide the background knowledge for the words in a language. As such, in order to understand word meaning, it is first necessary to have knowledge of the conceptual structures underlying their usage.

Consequently, in FBT, the ENVIRONMENTAL EVENT (EE) (Figure 1), derived from corpus and dictionary analysis, is the representation of the prototypical domain event (Barsalou 2003: 513; Faber et al. 2005; Faber 2011) and configures the most generic or base-level categories within the field of environmental science. In other

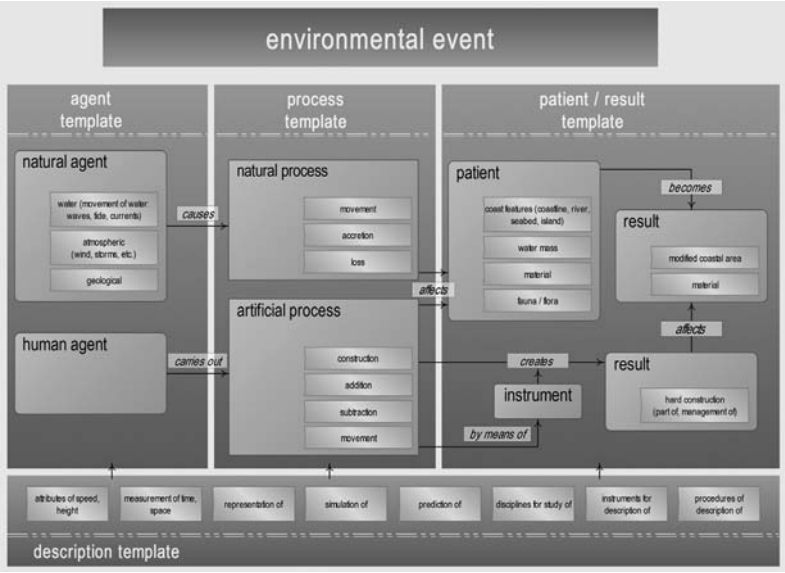



**Figure 1:** The ENVIRONMENTAL EVENT

words, it provides a frame or template for concepts that are linked by both hierarchical (e.g. *is-a*, *part-of*) and non-hierarchical relations (e.g. *affect*, *cause*, *create*).

As shown in Figure 1, the EE has two types of AGENT that can initiate processes: inanimate agents (natural forces) and animate agents (human beings). NATURAL AGENTS, such as earth movements and atmospheric phenomena, cause NATURAL PROCESSES in a geographic area. These processes affect other entities or PATIENTS, which as a RESULT, may suffer changes. HUMAN AGENTS can also use INSTRUMENTS to implement ARTIFICIAL PROCESSES (e.g. constructions), which can generate or prevent EFFECTS normally caused by natural processes. AGENT, PATIENT, RESULT, and INSTRUMENT are the most characteristic semantic roles of this specialized domain and the EE represents their relationships. However, there are also peripheral categories that include concepts used for the measurement, analysis, and description of the processes in the main event.

The practical application of FBT is EcoLexicon,[2] a visual thesaurus in which the environmental domain is configured in semantic networks and conceptual subdomains. Concepts are designated by terms in six languages: English, French, German, Greek, Russian, and Spanish. All of the entries in EcoLexicon are linked to the corresponding (sub)event and conceptual category. In other words, the conceptual, graphical, and linguistic information (including verb collocations) relative to entries are structured in terms of the underlying conceptual frame. The specification of the conceptual structure of subevents, such as the EXTREME EVENT, and the description of the lexical units in the frame are the result of a top-down and bottom-up approach. This includes the use of corpus, the factorization of definitional information, the application of LGM lexical domains, and the predicate-argument analysis of phraseological constructions such as verb collocations (Montero 2008; Montero and Buendía 2010, 2012).[3]

## 3.1 Corpus

A corpus of English and Spanish specialized texts belonging to the subdomain EXTREME EVENT was compiled for this research. In this process, quality and reliability of selected sources were guaranteed by an evaluation protocol (Buendía and Ureña 2009). Selection criteria were based on three parameters: (i) *authority*, which refers to the reputation and expertise of the authors; (ii) *content*, which includes coverage, accuracy, objectivity, currency, and audience; (iii) *design*,

**2** Available at:<http://ecolexicon.ugr.es> [2014/03/10].

**3** Because of space restrictions, this terminological description is only illustrated with English examples. However, reference to parallel work carried out in Spanish is given where necessary.

referring to the navigational aids, accessibility, presentation and management of web sites used as online resources. Thus, the *web for corpus* (WfC) approach (De Schryver 2002), that is, using the web as a source of texts for an offline corpus was used. Moreover, text scanning was also used for compiling when sources were not available in the web or in digital format.

The corpus selected can be described as a *bilingual* corpus, composed of two *written* subcorpora. More specifically, in the English subcorpus, texts are written in American and British English, and, in the Spanish one, in peninsular Spanish. It is also a *comparable* corpus, since the texts are on the same topic (i.e. the subdomain EXTREME EVENT within the general domain of the environment), but are not translations of each other. In this regard, it is a *special* corpus on natural disasters. Documents included are both full-text documents and samples. However, it can be referred to as a *full text* corpus since book chapters, for instance, have a considerable extension and a clear beginning and end. Regarding time period, the corpus is *synchronic* since texts date from 1996 to 2012. Finally, it is also a *plain* corpus because it does not have any type of linguistic annotation, though certain tags were added to enhance the retrieval and categorization of information (Buendía-Castro 2013: 325–236).

The English subcorpus includes 45 texts (five of which are books), which comes to a total of 779,995 tokens and 26,285 types, whereas the Spanish subcorpus contains 44 texts, which represents 449,416 tokens and 31,230 types. Initially, the objective was to compile texts with a similar number of words for each subcorpus, with a view to obtaining a more balanced bilingual corpus. However, in the end, the final size of the corpus was conditioned by the availability of texts, based on the selection criteria. This was particularly true of the Spanish subcorpus, since most of the textbooks in the domain were written in English. Consequently, it was fairly easy to retrieve textbooks on natural disasters in English in machine-readable format, but it was considerably more difficult to find textbooks in Spanish in this same format. Therefore, the textbooks in Spanish (Olcina 2006a, b; Keller and Blodget 2007) had to be scanned and converted to txt format with an OCR. For this reason, it was not possible to have two subcorpora of the same size, and this also explains why the Spanish corpus is smaller than the English one. However, the standardized type/token ratios (STTR) offered by *WordSmith Tools* present clear evidence that both corpus are comparable regarding lexical diversity. As such, in the English subcorpus, there are on average 40.61 different words for each text sequence of 1,000 tokens, whereas for the Spanish subcorpus the average is slightly higher (41.28).[4]

**4** The specific characteristics of the bilingual corpus, including source texts, can be found in Appendix 1.

Some materials in the Spanish corpus are translated texts such as the textbook by Keller and Blodget (2007) and the materials from *National Geographic*. The fact that it is hard to find specialized and semi-specialized materials originally written in Spanish means that experts and semi-experts within this field resort either to literature in English or to translated materials. Therefore, the role of translators cannot be underestimated, since they contribute to term formation by reconciling the linguistic constraints imposed by a language with the communicative expectations in a certain domain (Montero et al. 2001: 693). Descriptive terminological resources, such as Ecolexicon, must then contemplate the inclusion of reliable, high-quality translated materials, which contribute to the description of *in vivo* terminological and phraseological units. For example, Keller and Blodget (2007) was translated into peninsular Spanish by an expert translator (Pilar Gil Ruiz), and technically revised by several subject-field experts. It is a widely used textbook that is a crucial reference in undergraduate and postgraduate courses on Geology, Geophysics, Environmental Sciences, etc. throughout Spain.

In addition to the offline corpus, usage contexts and phrases were directly retrieved from the web following a *Web as Corpus* (WaC) approach (Baroni and Bernardini 2006; Fletcher 2007). The applications used included *SketchEngine*[5] and *WebCorp*.[6] This was done when the information in the main corpus did not satisfy communicative needs, when examples were not found, and when a wider range of contexts was required.

## 3.2 Conceptual and lexical scope

The conceptual structure of the EXTREME EVENT (in its sense of *natural hazard* or *natural disaster*) was first specified by formulating a preliminary description of the situation activated by the event. This was done by recurring to "armchair linguistics" (Fillmore et al. 2003: 299) and to specialized knowledge resources. In other words, the EXTREME EVENT frame was the result of intuition, dictionaries, thesauri, and specialized texts.[7]

**5** Available at: <https://www.sketchengine.co.uk/> [2014/03/10].

**6** Available at: http://www.webcorp.org.uk/live/> [2014/03/10].

**7** Among others, the concept NATURAL HAZARD was analyzed in *EuroVoc*, the multilingual thesaurus of the European Union; in the typology of hazards provided by the *European Environment Agency*, and in the organization of NATURAL DISASTERS provided by the website *Science Daily*.

Event-specific concepts include VOLCANO, LANDSLIDE, FLOOD, AVALANCHE, TROPICAL CYCLONE (including TYPHOON and HURRICANES), TORNADO, DROUGHT, EARTHQUAKE, FIRE, and TSUNAMI, all of which belong to the conceptual category NATURAL HAZARD. Other categories within the EXTREME EVENT are ATMOSPHERIC AGENT, WATER AGENT, ATMOSPHERIC CONDITION, MATERIAL ENTITY, AREA, CONSTRUCTION, ENERGY, HUMAN BEING, LANDFORM, WATER COURSE, DEATH, DAMAGE, LOSS OF LIFE/PROPERTY, PLANT, and EXPLOSIVE. Figure 2 shows a network of these concepts, along with the definition of EXTREME EVENT.[8]

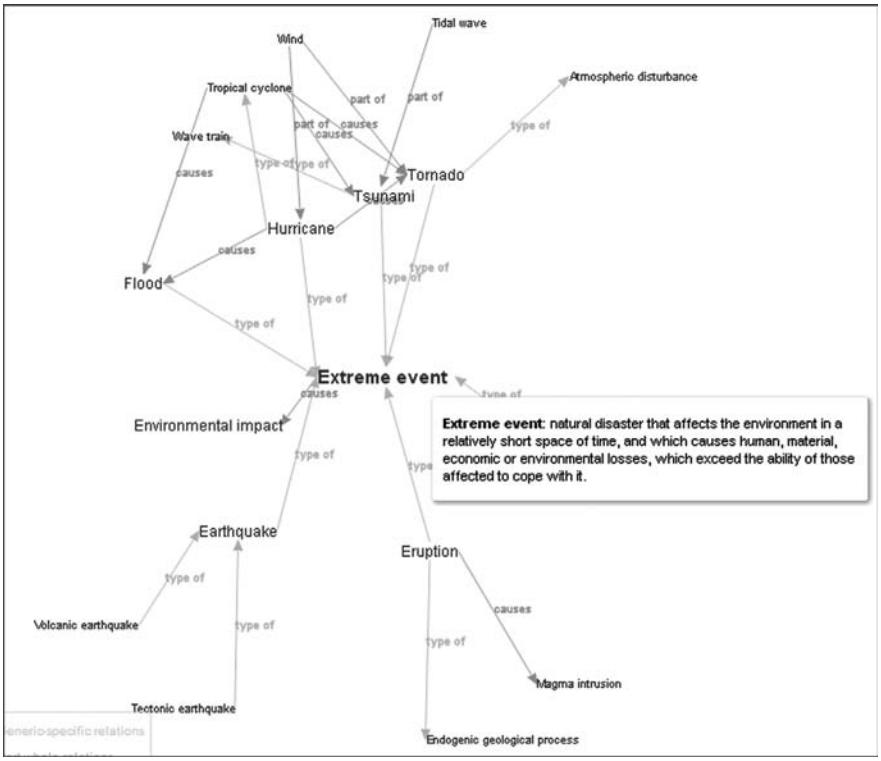



**Figure 2:** The EXTREME EVENT in EcoLexicon

**8** For the sake of clarity, the number of part-whole and non-hierarchical relations in the display has been kept to a minimum. For the full representation of the EXTREME EVENT, readers can access the environmental database EcoLexicon.

After the factorization of dictionary definitions for the most salient concepts in the EXTREME EVENT, candidate verbs and arguments in corpus texts (bottom-up approach) were extracted.

### 3.2.1 Lexical units

To identify the lexical units in the EXTREME EVENT frame, candidate terms were retrieved from corpus. The extraction was performed with *TermoStat*, a tool developed by Drouin (2003) at the University of Montréal. It statistically identifies terms in such a way as to compare their frequencies in a domain-specific and a general reference corpus. The search in the English subcorpus provided the verbs in Table 1.

**Table 1:** The most frequent verbs extracted from the English subcorpus *(TermoStat)*

Résultats

Liste des termes | Nuage | Statistiques | Structuration | Bigrammes

| Candidat de regroupement | Fréquence | Score (Spécificité) | Variantes orthographiques | Matrice |
|---|---|---|---|---|
| occur | 900 | 50.8 | *occur*<br>*occurs*<br>*occurred*<br>*occurring* | Verbe |
| cause | 850 | 40.71 | *cause*<br>*causes*<br>*causing* | Verbe |
| trigger | 256 | 38.07 | *trigger*<br>*triggers*<br>*triggered*<br>*triggering* | Verbe |
| flood | 225 | 35.34 | *flood*<br>*floods*<br>*flooded*<br>*flooding* | Verbe |
| affect | 544 | 33.19 | *affect*<br>*affects*<br>*affected*<br>*affecting* | Verbe |
| increase | 810 | 32.74 | *increase*<br>*increases*<br>*increasing* | Verbe |
| associate | 400 | 30.18 | *associate*<br>*associated*<br>*associating* | Verbe |
| evacuate | 155 | 29.22 | *evacuate*<br>*evacuates*<br>*evacuating* | Verbe |

The term list (first column of Table 1) displays the candidate verbs [*candidat de regroupement*] in their lemmatized form. They are followed by their absolute frequency in the text, the score specificity assigned to each unit, the orthographic variants (i.e. non-lemmatized forms) as well as the spelling variants. The most frequent verbs in the English corpus are *occur*, *cause*, *trigger*, *flood*, *affect*, *increase*, *associate*, and *evacuate*. The analysis of the English corpus produced a total of 352 verbs and that of the Spanish corpus, 323 verbs. However, some of these candidate verbs had to be discarded since they corresponded to other parts of speech (e.g. the English *crater*, *cluster*, *firefight*, *map,* and the Spanish *caliza*, *riesgo*, *ventisca, sureste*). The other candidate terms suggested by *TermoStat* were initially accepted. Even though some of them were not specifically linked to the ENVIRONMENTAL EVENT and the EXTREME (sub)EVENT (e.g. the English verbs, *feel*, *study*, *analyze*, *observe*, *detect;* and the Spanish verbs, *variar*, *oscilar*, *analizar*, *acrecentar*), they were considered useful for the configuration of the EXTREME EVENT (cf. Figure 4).

*TermoStat* also provided a general overview of other lexical units such as nouns, adjectives and adverbs. The *tag cloud* utility generated the 100 terms with the highest scores of the corpus, arranged alphabetically and with different font sizes, depending on their frequency. Figure 3 shows the tag cloud for the English corpus.

As can be observed in Figure 3, some of the most frequent lexical units are the terms designating event-specific concepts, such as *avalanche*, *cyclone*, *earthquake*, *drought*, *eruption*, *event*, *fire, flood*, *flooding*, *flow*, *hazard, hurricane*, *landslide*, *rainfall*, *storm*, *tsunami*, *tornado*, and *volcano*. Other frequent words are *change*, *damage*, *disaster*, *evacuation, famine*, *fatality*, *impact*, *loss*, and *risk*, which are some of the results/consequences of NATURAL HAZARDS. Also on the list are nouns, such as *height*, *intensity*, *instability*, *level*, *probability*, *speed*, and *velocity,* which point to the attributes of these phenomena. The occurrence of *scale* indicates how intensity is measured, and *frequency* points to the cyclic nature of these events. *Pressure*, *wind*, and *wave* reflect the fact that NATURAL HAZARDS are closely linked to these weather conditions.

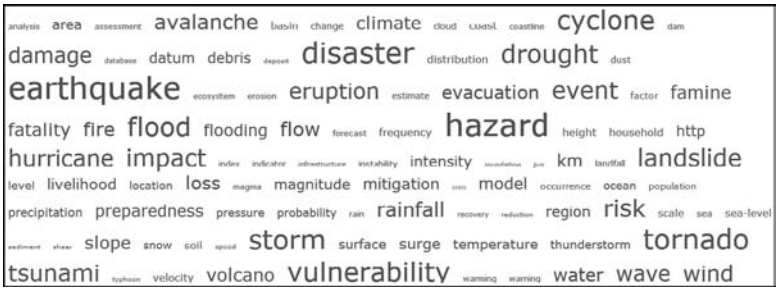



**Figure 3:** Tag cloud for the English subcorpus (*TermoStat)*

In addition, other words, such as *area*, *coastline*, *location*, *surface*, and *sea*, are evidence of the connection between NATURAL HAZARDS and the place where they generally happen. Other nouns, such as *household* or *population*, reflect that NATURAL HAZARDS negatively affect people and are threats to human life and property.

The bilingual corpus was also analyzed with the modules *Wordlist* and *Concord* of *Wordsmith Tools*.[9] *Wordlist* was used to list the words in the corpus and their frequency, which made it possible to identify more lexical units that designated concepts belonging to the EXTREME EVENT. *Concord* was used to generate concordance lines for each lexical unit. This information was used to extend the EXTREME EVENT and specify verb collocations.

### 3.2.2 Lexical (sub)domains

The most prominent lexical domains in the EXTREME EVENT were identified in order to classify candidate verbs. Consequently, verbs were grouped in sets that lexicalize all or part of a conceptual domain and which share the same nuclear meaning and syntax (Faber and Mairal 1999: 59).

The LGM divides the verb lexicon into twelve lexical domains. Each domain has one or two generic verbs or superordinates, in terms of which all the members of the domain are directly or indirectly defined. Table 2 shows the lexical domains (in square brackets) and their superordinate verbs (italics) (Faber and Mairal

**Table 2:** Lexical domains in LGM (Faber and Mairal: 1999: 88)

| | |
|---|---|
| (i) | *to be* [EXISTENCE] |
| (ii) | *to become different* [CHANGE] |
| (iii) | *to have/give* [POSSESSION] |
| (iv) | *to say* [SPEECH] |
| (v) | *to feel* [EMOTION] |
| (vi) | *to do/make* [ACTION] |
| (vii) | *to use* [MANIPULATION] |
| (viii) | *to know/think* [COGNITION/MENTAL PERCEPTION] |
| (ix) | *to move (go/come)* [MOVEMENT] |
| (x) | *to become aware (notice/perceive)* [GENERAL PERCEPTION] |
| (xi) | *to see/hear/taste/smell/touch* [SENSE PERCEPTION] |
| (xii) | *to be/stav/put* [POSITION] |

**9** Available at: <http://www.lexically.net/wordsmith> [2014/03/10].

**Table 3:** Lexical subdomains (CHANGE)

| CHANGE |
|---|
| to become different by going from one state to another |
| to become larger in size/amount number/degree |
| to become less in size/amount/degree'/importance/intensity |
| to cause sth to become less in size |
| to cause sth to increase in moisture |
| to cause sth to lessen in moisture |
| to cause sth to increase in temperature |
| to cause sth to decrease in temperature |
| to cause a change in sth so that it loses its power or control |
| to cause sth/sb to change for the worse |
| to begin to be different in the way that is stated |
| to separate into two or more pieces |
| to cause sth not to be changed by sth else |

1999: 88). For example, the superordinate for CHANGE,[10] *to become different*, marks the semantic territory covered by this domain and is the *genus* of the definition of each lexeme in the domain.

Lexical domains can be further subdivided into *subdomains*. Each subdomain focuses on a particular area of meaning and reflects a different specification of its content. For instance, in Table 3, CHANGE is subdivided into 13 meaning areas, which include subdomains such as *to cause sth/sb to change for the worse*.

To specify the most prominent lexical domains and subdomains activated within the EXTREME EVENT in English and Spanish, the verbs extracted with *TermoStat* were classified based on the *genus* and *differentiae* of their definitions. The *differentiae* are the semantic information that distinguishes each verb from others belonging to the same lexical (sub)domain.

Within CHANGE, the subdomain *to cause sth/sb to change for the worse* includes English verbs such as *affect*, *damage*, destroy, *deepen,* and *exacerbate*. As shown in Table 4, all of these verbs have a *genus* that directly or indirectly corresponds to the superordinate in the hierarchy. For instance, the verb *damage* is defined as "to cause physical harm to sth/sb without destroying them". However, the verb *destroy* is defined in terms of its hypernym "to damage sth so much that it does not exist anymore". In addition, the *differentiae* in both definitions allows

**10** In this paper, lexical domains appear in uppercase font in contrast to specialized domains that appear in small capital font.

**Table 4:** English verb definitions within the EXTREME EVENT

| To cause sth/sb to change for the worse (English) |
|---|
| **affect:** *to cause sth to change for the worse.* |
| **damage:** *to cause physical harm to sth/sb without destroying them.* |
| **destroy:** *to damage sth so much that it does not exist anymore.* |
| **devastate:** *to destroy sth (usually an area or place) completely.* |
| **ravage:** *to devastate.* |
| **demolish:** *to destroy sth (usually a building) completely.* |
| **wreck:** *to destroy sth completely (especially as the result of an accident).* |
| **sweep away:** *to destroy sth completely and without leaving a trace.* |
| **burn1:** *to cause sth to be damaged, injured or destroyed by fire or extreme heat.* |
| **injure:** *to damage some part of sb's body.* |
| **deepen:** *to cause sth to become worse.* |
| **exacerbate:** *to cause sth to become a lot worse (formal).* |

for the distinction between "to damage sth *without destroying them*" and "to damage sth *so much that it does not exist anymore*", respectively.[11]

The concepts activated by the EXTREME EVENT (e.g. HURRICANE, TSUNAMI, FIRE, etc.) are generally processes that begin to exist, exist over a period of time, and finally cease to exist. During their EXISTENCE, they ACT in a certain POSITION or LOCATION causing negative CHANGE (i.e. damage) in an affected entity. It is thus not surprising that the most frequent verbs in the EXTREME EVENT frame generally belong to the domains of EXISTENCE, ACTION, MOVEMENT, POSITION, and CHANGE. In fact, verbs of ACTION, MOVEMENT, and CHANGE are especially relevant.

However, verbs associated with the domains of SPEECH (e.g. *predict*), MENTAL PERCEPTION (e.g. *verify*) and GENERIC PERCEPTION (e.g. *observe*); MANIPULATION (e.g. *control*), and POSSESSION (e.g. *unload*) are also found in texts within this specialized domain.[12] Verbs of MANIPULATION (e.g. *measure*) clearly activate relations of *instrument_of*. In contrast, CHANGE (e.g. *destroy*, *devastate*), ACTION (e.g. *hit*, *strike*), MOVEMENT (e.g. *shake*, *sweep)*, and POSITION (e.g. *flood*, *engulf*) activate the relation of *affects*. Therefore, it is possible to establish a correspondence between domains and the relations most prototypically activated within the EXTREME EVENT (see Figure 4).

**11** The complete list of verb definitions, within the context of the EXTREME EVENT, along with their classification in lexical domains and subdomains can be found in Buendía (2013: 577–604).
**12** The only lexical domain that was not activated was that of SENSE PERCEPTION.

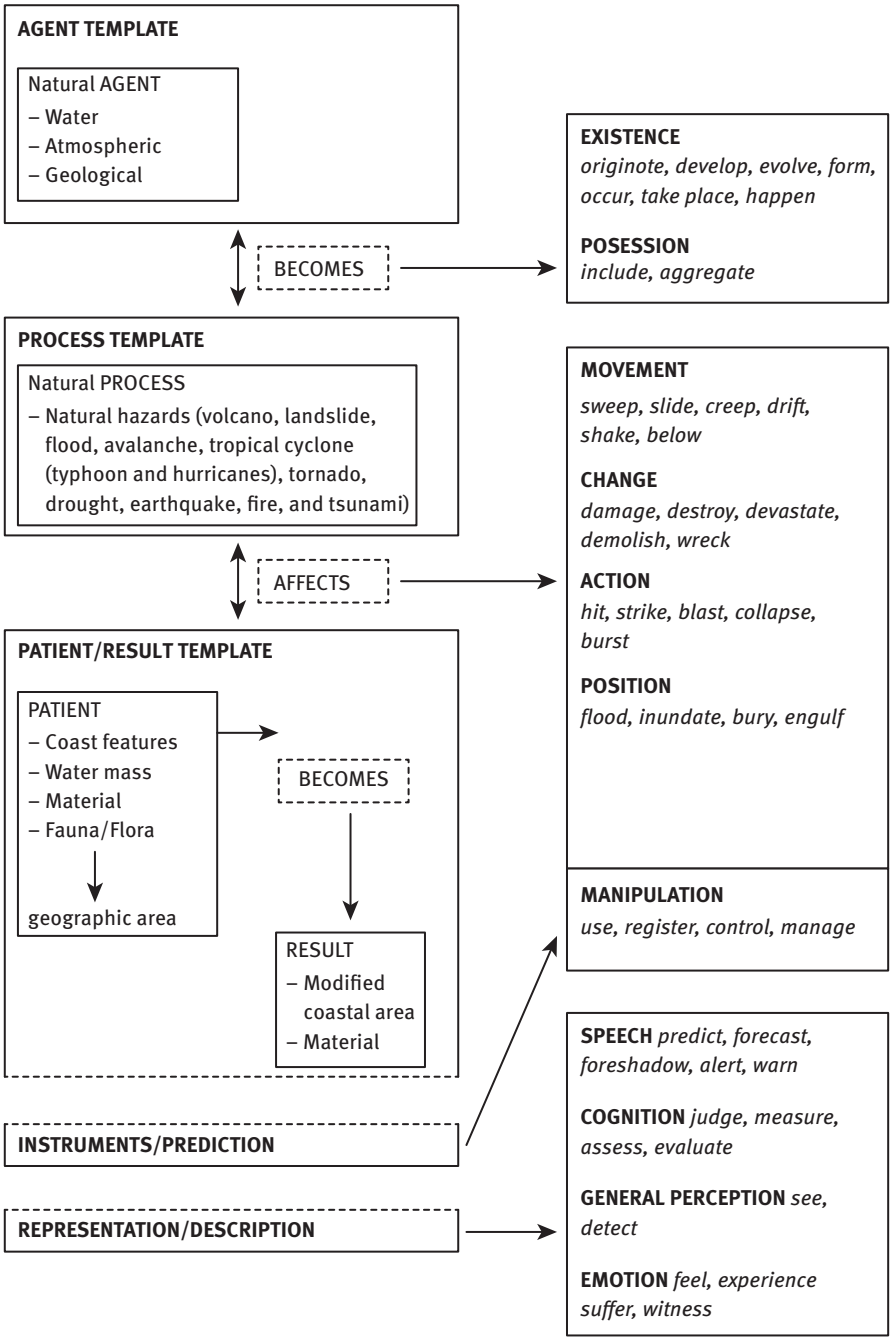



**Figure 4:** Activation of lexical domains within the EXTREME EVENT

# 4 Verb collocations in EcoLexicon

In FBT and EcoLexicon, verb collocations refer to highly frequent combinations of two or more words (Heid 1999; Siepmann 2005: 417), following a noun + verb or verb + noun pattern, where the noun is the *base* and the verb is the *collocate*. In specialized discourse, verb activation is often modulated by the event evoked by the meaning of the noun. At the same time, the verb constrains the semantic nature of its arguments. For example, in the collocation *the volcano spits*, 'volcano' takes a verb designating something being forced out of it (e.g. 'spit'). However, it is also true that 'spit' only combines with noun phrases designating something being forced out of a mouth or a mouth-like orifice (e.g. 'volcano', in which the crater is considered to be the mouth). Consequently, co-selection makes such collocations partially compositional, since base and collocate retain their meaning to a certain extent.

## 4.1 Analysis and description

The process of analysis, description, and storage of verb collocations within FBT is illustrated within the EXTREME EVENT. More specifically, typical arguments for NATURAL HAZARD and VERBS in the lexical subdomain *to cause to change for the worse* were analyzed. In order to identify verb collocations related to NATURAL HAZARD, a bottom-up analysis was performed. For this purpose, predicates and their arguments were identified and analyzed as follows: (i) semantic roles and macro-roles were assigned to arguments in the EXTREME EVENT; (ii) linguistic realizations of arguments were specified in the category NATURAL HAZARD; (iii) verb collocations were described in the templates.

### 4.1.1 Semantic roles and macroroles

Semantic roles generally express the set of properties that a verb requires for its argument/s. The semantic roles used in this study largely coincide with the most general thematic relations in Role and Reference Grammar (RRG) (Van Valin and LaPolla 1997; Van Valin 2005) and with the argument roles proposed by Goldberg (1995, 2006). Other roles were taken from VerbNet (Kipper 2005), ADESSE (*Alternancias de Diátesis y Esquemas Sintáctico-Semánticos del Español*) (Vaamonde et al. 2010), and Sensem (*Sentence Semantics*) (Fernández and Vázquez 2012). Along with semantic categories and semantic roles, arguments were assigned

**Table 5:** Semantic roles in the EXTREME EVENT

| Role | Definition and example |
|---|---|
| Agent | Human entity that carries out an action with intentionality<br>(e.g. 'The **man** started the fire intentionally') |
| Natural Force | Unvolitional agent that causes an action<br>(e.g. 'The **hurricane** destroyed the city') |
| Destination | Endpoint of path<br>(e.g. 'At the time of the final eruption, ash was ejected into the atmosphere **towards the north-east**') |
| Experiencer | Animate entity which continues to exist or ceases to exist in a natural disaster. It is usually a human being or a personified construction, plant, or area<br>(e.g. 'There were few **people** who survived the earthquake'). |
| Frequency | The rate at which something occurs over a particular period of time<br>(e.g. 'Many of Earth's volcanoes have erupted **dozens of times** in the past few thousand years') |
| Geographical Location | Underspecified place that usually designates where a situation takes place<br>(e.g. 'The tornado started at 3:30 pm **about 1.2 miles north of Poplar Grove in Boone county** and ended at 3:48 pm') |
| Manner | The way in which an action is carried out.<br>(e.g. 'This was the start zone where the avalanche **suddenly** originated'). |
| Path | Trajectory of the motion.<br>(e.g. 'Sufficiently large volumes of gases were lofted **into the stratosphere**') |
| Patient | Entity that is affected by a natural force.<br>(e.g. 'The hurricane devastated **the coast'**) |
| Situation/ experience | Situational context in which an event occurs.<br>(e.g. 'In addition to the ash, a large amount of sulfur dioxide gas and aerosols were ejected **in the eruption**') |
| Origin | The location origin, i.e. starting point, of a motion/trajectory or the entity from which another entity forms.<br>(e.g. 'A hurricane developed **from a tropical storm'**)<br>(e.g. 'The 1919 eruption of Mt Kelat on Java expelled water **from a crater lake**, covering 200 km2 of farmland') |
| Theme | Entity that undergoes a change of state (e.g. 'The **tropical storm** evolved into a hurricane'), which develops from sth else (e.g. 'The **hurricane** developed from an easterly wave'), which simply exists without entailing an action (e.g. 'The **hurricane** lasted 3 days'), or the thing or substance with which an area is filled (e.g. 'Fields were inundated **with water**). |
| Time | The time at which the action occurs.<br>(e.g. 'The volcano erupted **in November 1995**') |
| Result | An entity that is caused or produced by something else, a consequence or outcome.<br>(e.g. 'The hurricane developed **into a tropical storm** in the central Atlantic') |

**Table 6:** Semantic roles for *hurricane*

| | Semantic category | Semantic roles | Macroroles | Examples |
|---|---|---|---|---|
| ***Hurricane*** | Natural hazard | Theme | Undergoer | 'The **hurricane** evolved into a tropical storm' |
| | Natural hazard | Natural force | Actor | 'The **hurricane** struck the city of New York' |

macroroles, as proposed by RRG (i.e. ACTOR and UNDERGOER). The number of macroroles that a predicate can take is 0, 1, or 2.

Table 5 lists a non-exhaustive inventory of semantic roles for the EXTREME EVENT (Buendía 2013: 380). A definition is provided for each role as well as an English usage example from the corpus with the linguistic realization in bold font.

The linguistic realization associated with a certain semantic category does not always have the same semantics. For example, as shown in Table 6, a *hurricane* is always understood as a NATURAL HAZARD. However, depending on the verb collocate, it may activate either the semantic role THEME, when it merely exists but does not entail action or change, or the role NATURAL FORCE, when it entails action and produces damage (e.g. *The hurricane struck the city of New York*).

Regarding macroroles, in the first example in Table 6, *hurricane* is an UNDERGOER since it experiences a process and is transformed into something else (*tropical storm*). However, in the second example, *hurricane* is an actor because it negatively affected the city of New York.

### 4.1.2 *Arguments and linguistic realizations*

Semantic categories are generalizations for a set of terms that are assumed to have a similar semantic and syntactic behavior. As such, arguments belonging to a semantic category not only share a common nuclear meaning but also similar syntactic projections. In FBT, categories are based on the following: (i) the semantic relations of the concepts expressed by the linguistic realizations in the corpus; (ii) its verification by means of tests based on Gross's (1994) *classes d'objets*.

Accordingly, in EcoLexicon, NATURAL HAZARD has the following characteristics: (i) natural hazards *cause* human/economic/material losses; (ii) natural hazards *affect* the environment; (iii) natural hazards *occur* in a short period of time

(cf. Figure 2). Thus, whenever a concept fulfills these three conditions, it belongs to the category of NATURAL HAZARD. For instance, the referents of the terms *earthquake*, *typhoon* or *fire* all cause human, economic, and material losses; they affect the environment; and they occur in a short period of time. Furthermore, when they are agents or actors, they appear with verbs, such as *hit, strike*, *devastate*, *destroy*, etc.

The terms (and arguments) in the corpus, which are lexical projections of the category NATURAL HAZARD in English, are the following: *avalanche*, *drought*, *earthquake*, *quake*, *cyclone*, *hurricane*, *tropical cyclone*, typhoon, *tsunami*, *landslide*, *tornado*, *eruption*, *volcano*, *flood*, *flooding*, *fire*, and *natural disaster.*

#### 4.1.3 *Verb template for the predicate* DAMAGE

The representation of phraseological information in EcoLexicon is based on verb descriptions encoded in templates (cf. Table 7). Each template includes:

- The lexical domain and subdomain activated by the predicate within the EXTREME EVENT (see Section 3.3);
- The specification of the subdomain with the semantic roles and categories activated by the arguments within the EXTREME EVENT;
- A user note, when necessary, to explain the meaning of the verb and/or its usage.

Subsequently, the potential arguments of the predicate are encoded and described in the template in terms of their semantic roles and macroroles, semantic category, linguistic realizations, and morphology or phrase type. Usage examples are also provided. Finally, a template with the requirements and constraints of the subdomain and of all member verbs is provided (cf. Table 8).

As previously mentioned, some of the most frequent verbs in the EXTREME EVENT belong to the lexical domain of CHANGE. More specifically, they activate the subdomain *to cause to change for the worse* (e.g. DAMAGE, in Table 7). These verbs generally have two arguments:

- NATURAL HAZARD with the role of NATURAL FORCE and the macrorole of ACTOR;
- PATIENT, with the macrorole UNDERGOER, which is the entity affected by the NATURAL FORCE, and which can be a CONSTRUCTION, HUMAN BEING, AREA, or PLANT.

Optionally, SITUATION/EXPERIENCE can also be specified, as well as LOCATION, TIME or MANNER. However, there are times when these verbs only occur with the patient since the NATURAL FORCE is implicit in the action of the verb or appears as

**Table 7:** Verb template for DAMAGE

**Predicate: DAMAGE**
**Lexical domain: CHANGE**
**Subdomain: to_cause_to_change_for_the_worse**
**Definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse.**
**Note: The PATIENT is normally a construction, human being, area, or PLANT. SITUATION/ EXPERIENCE can also be included. LOCATION, TIME and MANNER can also be specified.**

| **S. Role** | Natural force | Patient | Situation/ Exp. | Geo. location | Time | Manner |
|---|---|---|---|---|---|---|
| **Macrorole** | Actor | Under-goer | _____ | _____ | _____ | _____ |
| **Category** | natural disaster, part_of natural disaster (material), water agent, atmospheric agent, atmospheric condition | Construc-tion, human being, area, plant | _____ | _____ | _____ | _____ |
| **Linguistic realizations** | tornado, cyclone, earthquake, lava, hurricane, flooding, drought, rockfall, water, ice, wind | people, boat, homes, houses, dam, building, the coast, crops | earth-quake, floods | region, Fossa and Stiffe, Indonesia, Texas to Louisiana | 1985, the beginning of 2002 | seriously, severely, badly |
| **Morphology** | NP | NP | PP (in) | PP (in, at, from … to) | PP (in, at) | AVP |
| **Usage examples** | 1. Have there been instances of ships or boats being capsized or badly damaged by tornadoes?<br>2. In 1985 another cyclone killed 10,000 people, destroyed 17,000 homes and damaged a further 122,000.<br>3. The dam has already been damaged by frequent earthquakes in the region.<br>4. Lava spilling out of the mouth of the volcano damaged several houses.<br>5. This is done to protect the dam from being damaged by high water or overspill. | | | | | |

(*Continued*)

**Table 7:** (*Continued*)

| | |
|---|---|
| **Usage examples** | 6. Flooding in Indonesia at the beginning of 2002 killed at least 150 people and destroyed or damaged $177 million worth of infrastructure.<br>7. Over 7000 homes were damaged by wind.<br>8. About 12 landslides destroyed more than 100 homes and damaged railways and roads.<br>9. This hurricane damaged the coast from Texas to eastern Louisiana.<br>10. Crops were seriously damaged by drought.<br>11. 200 cars damaged in floods. |

a SITUATION or EXPERIENCE in which the action occurs. When this happens, the argument is a CONSTRUCTION, GEOGRAPHICAL AREA, HUMAN BEING, or PLANT. The argument has the role of PATIENT and activates the macrorole of UNDERGOER (e.g. *the house* was completely devastated).

Notwithstanding, different verbs in the same subdomain activate different types of PATIENT. For example, the PATIENT argument of *damage* is very general and, as such, can be a CONSTRUCTION (e.g. the *factory* was damaged in an earthquake in San Felice), a GEOGRAPHIC area (e.g. the hurricane damaged the *coast*), a PLANT (e.g. *crops* were seriously damaged by drought), or a HUMAN BEING (e.g. How many *people* were damaged by hurricanes?). In contrast, the PATIENT arguments of verbs such as *demolish* or *wreck* can only be CONSTRUCTION or GEOGRAPHIC AREA entities. Consequently, the usage contexts are thus more restricted. In this sense, *demolish* and *wreck* are more focalized than *damage* (cf. Table 4).

The number of sample sentences in the verb templates is variable, and depends on the verb sense and its specificity. As in FrameNet, sentence elements appear in different colors, depending on their semantic category. However, unlike FrameNet, only the heads of the phrases are annotated. In other words, instead of coloring the whole phrase (e.g. *the powerful hurricane*), only the head of the phrase (i.e. *hurricane*) is annotated. As can be observed in Table 7, for noun phrases and adverbial phrases with the roles of GEOGRAPHICAL LOCATION, TIME, and MANNER, the objective is not so much to identify linguistic realizations, but rather to identify the set of prepositions with which the heads generally occur (e.g. The hurricane developed *in the Atlantic ocean*). Accordingly, the preposition is specified in the phrase type section of GEOGRAPHICAL LOCATION in the template (i.e. PP (*in*)).

Finally, other frequent verbs in the EXTREME EVENT, such as *destroy*, *devastate* and *ravage*, also belong to the subdomain *to cause to change for the worse* (CHANGE). Nevertheless, since these verbs are not interchangeable in all contexts, a template showing restrictions is included (see Table 8 for an example).

**Table 8:** Restriction template for *to cause to change for the worse*

| **Subdomain: to_cause_to_change_for_the_worse** | | | | | | |
|---|---|---|---|---|---|---|
| **Semantic role** | Natural force | *affect damage destroy devastate ravage sweep away* | *burn1** | *demolish wreck* | *injure* | Patient |
| **Macrorole** | Actor | | | | | Undergoer |
| **Category** | natural disaster [fire]* | | | | | construction, area, human being, plant |
| **Morphology** | NP | | | | | NP |

In this regard, verbs such as *burn1* tend to appear with NATURAL HAZARD entities that are fire events (e.g. The *fire* burnt the house). The verbs in the first column (*affect*, *damage*, *destroy*, *devastate*, *ravage*, *sweep away*) generally have a PATIENT which can be a CONSTRUCTION, AREA, PLANT, or HUMAN BEING, i.e. the entire set of categories identified for PATIENTS. Verbs in the third column (*demolish* and *wreck*) usually have PATIENTS that are CONSTRUCTION or GEOGRAPHICAL AREA entities. Finally, the PATIENTS of *injure* are HUMAN BEINGS and PLANTS.

## 4.2 The entry for HURRICANE in EcoLexicon

EcoLexicon is designed to facilitate the monolingual and bilingual searches of different user groups: (i) science students wishing to access specialized knowledge; (ii) translators seeking linguistic correspondences, term usage, and conceptual knowledge; (iii) experts interested in text production in a second language. These groups correspond to three user profiles: laypeople, semi-experts, and experts with different knowledge levels (Bergenholtz and Tarp 2010: 34–35). In order to use EcoLexicon, users should have an adequate command either of English or Spanish (interface languages).

Searches in EcoLexicon can be carried out either by concept or by term. In addition, searches can be restricted to a specific (sub)domain within the environment (i.e. climatology, ecology, meteorology, oceanography, zoology, etc.).[13]

**13** For a detailed description of all the specialized (sub)domains included in EcoLexicon see León and San Martín (2011).

### 4.2.1 Conceptual, graphic, and linguistic information

Entries in EcoLexicon contain conceptual, graphic, and linguistic information. Figure 5 shows the entry for HURRICANE, which includes a definition based on a set of hierarchical and non-hierarchical conceptual relations:[14]

> "tropical cyclone [*type_of*] with sustained winds of 118 km per hour or greater [*attribute_of*], in the North Atlantic Ocean, Caribbean Sea, Gulf of Mexico, and in the eastern North Pacific Ocean [*has_location*], and which is called 'typhoon' [*type_of*], in the western Pacific and 'cyclone' [*type_of*], in the Indian Ocean".

In addition to the semantic relations in the definition, Figure 5 shows the conceptual network displayed in the center panel. The user may activate/deactivate the number

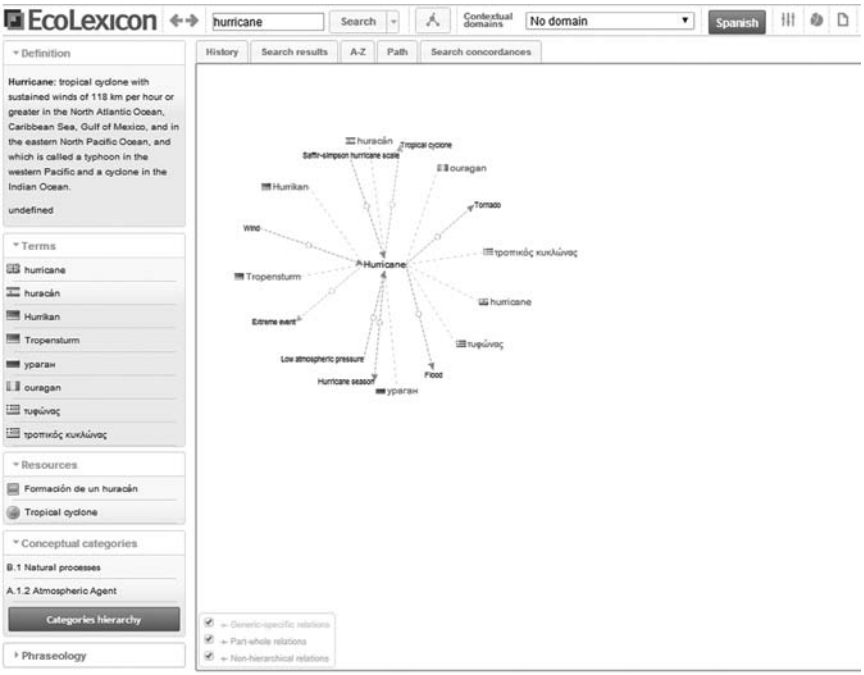



**Figure 5:** Entry for HURRICANE in EcoLexicon

**14** In the environmental domain, the hierarchical relations *is_a, type_of, part_of* and the non-hierarchical relations *made_of, phase_of, delimited_by, located_at,* and *attribute_of* are used for the construction of definitional templates. A complete description of this process in FBT and Ecolexicon is provided in León and Reimerink (2010).

and type of relations displayed in order to obtain a simpler/more complex conceptual representation. For example, the graphical representation shows that a HURRICANE *takes_place_in* the HURRICANE SEASON and *causes* TORNADO. Regarding the attributes of the concept, it is a LOW ATMOSPHERIC PRESSURE AREA, and its intensity is *measured_by* the SAFFIR-SIMPSON HURRICANE SCALE.

In the left column, this conceptual description is complemented with different types of resources, as well as associated conceptual categories. Resources also include encyclopedic explanations (such as the one referring to *tropical cyclone)*, and graphical information, including images of the concept, which enhance the content of the definition.[15] The conceptual description of HURRICANE also specifies its membership in one or more conceptual categories. In other words, the user can access its different conceptualizations in specialized discourse. As such, HURRICANE can be conceptualized either as an ATMOSPHERIC ENTITY that initiates a process or as the PROCESS itself.[16]

In addition to the conceptual and graphical information for HURRICANE, concept entries also include the terms that designate the phenomenon in different languages, namely *hurricane*, *huracán*, *Hurrikan*, *Tropensturm*, *ураган*, *ouragan*, and *τυφώνα* (cf. Figure 5, left column).

### 4.2.2 The term entry for hurricane

Table 9 shows the microstructure for the English term entry for *hurricane* with the following fields: language (*EN*); term type (*main term*); context (*hurric3a.txt*); part of speech (*common noun*); concordances; phraseology section; and a complete phraseological entry.

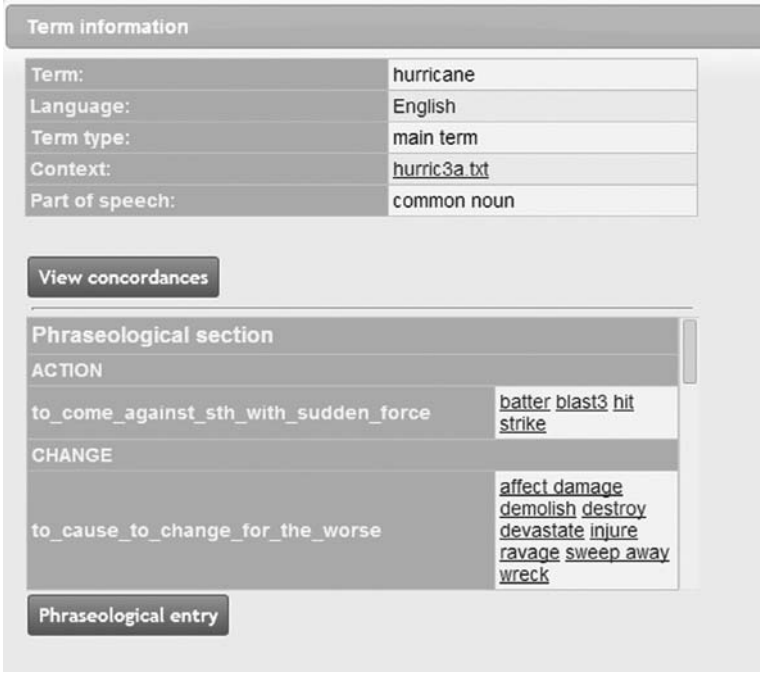
The context field includes a selection of meaningful and/or defining contexts (Reimerink et al. 2010). These textual fragments contain lexical markers (knowledge patterns) that are characteristic of knowledge elements that facilitate the understanding of a particular term or concept.[17] The concordance field contains a set of examples for *hurricane*. These are displayed in the form of KWIC (*key-word-in-context*) lines, which show all occurrences of the term in contexts of a fixed length within the corpus. These lines are selected based on knowledge patterns and are intended to exemplify term usage. For instance, the user can find the prepositions for the search word and the most frequent noun phrases (Montero et al. 2002; Montero 2008).

**15** To this end, depending on the type of knowledge activated, three kinds of images are used: iconic, abstract, and dynamic. Refer to Prieto (2009) and Prieto and López (2009) for a full description of graphic information in EcoLexicon.

**16** For more information about the ontology in EcoLexicon, see León and Magaña (2010).

**17** For more information regarding selection of knowledge contexts in EcoLexicon, see Reimerink et al. (2010) and Reimerink et al. (2012).

**Table 9:** Extract of the term entry for *hurricane*




### 4.2.3 Microstructure of phraseological (sub)entries

Verb collocations in EcoLexicon are first classified in terms of their *nuclear meaning* (i.e. the lexical domain) and then according to their *meaning dimension* (i.e. the lexical subdomain). Table 9 shows the partial phraseological subentries for *hurricane*. They include the dimension *to cause to change for the worse* (lower case) under the nuclear meaning CHANGE (upper case). The verbs (hyperlinks) in this dimension include *affect*, *damage*, *demolish*, *destroy*, *devastate*, *injure*, *sweep away*, *wreck* and *ravage*. By clicking on *damage*, the user accesses four usage examples as well as a note section with information about meaning restrictions. In this case, the note states that the PATIENT is usually a construction entity or area (see Table 10).

In addition to the information in phraseological subentries, the user has access to the complete phraseological information by means of the hyperlink *Phraseological entry* at the bottom of the term entry for *hurricane* (cf. Table 9). The phraseological entry includes a series of templates classified according to meaning. Table 11 shows the template for CHANGE (dimension *to cause to change for the worse*).

**Table 10:** Verb details for *damage* in EcoLexicon

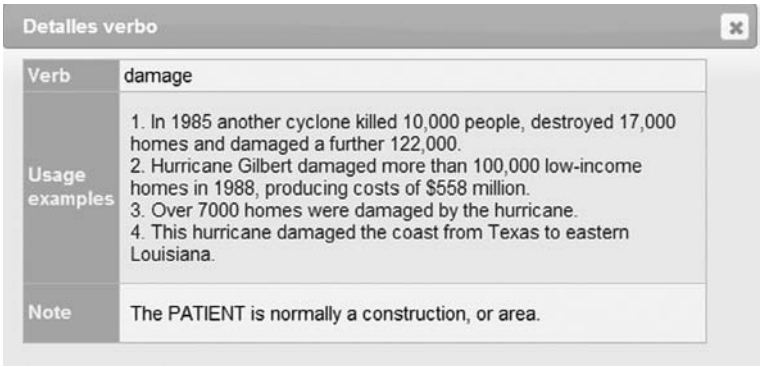
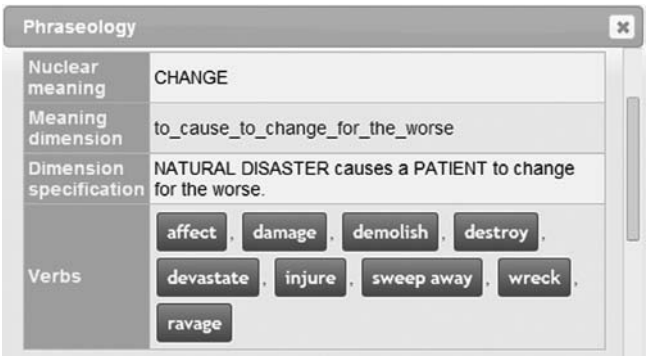



**Table 11:** English phraseological entry




The template also contains the *dimension specification*, which describes the dimension within the context of natural hazards or disasters (i.e. *NATURAL DISASTER causes a PATIENT to change for the worse*). Finally, it shows the verbs that can activate this meaning (*affect*, *damage*, *demolish*, *destroy*, *devastate*, *injure*, *sweep away*, *wreck* and *ravage*). Each verb can be clicked on to obtain a description like the one in Table 10.

### 4.2.4 Macrostructure of phraseological (sub)entries

The phraseological module in EcoLexicon enables users to access collocational information for *hurricane destroys* by means of the base (*hurricane*) as well as

the collocate (*destroy*). In addition, users seeking usage information can find interlinguistic correspondences for verb collocations stored in the knowledge base.

When phraseological information is accessed through the term, Figure 6 shows how the list of verb collocates for the search term *hurricane* is displayed in the column labeled *Phraseology*. A description of any of these units can be accessed just by clicking on it. In addition, EcoLexicon offers the possibility to access the complete phraseological entry for *hurricane* through the hyperlink *Phraseological Entry*.

The verb can also be entered in the search window in order to retrieve associated terms. EcoLexicon provides a list of terms for the verb, which, apart from *hurricane*, includes *tsunami*, *landslide*, *earthquake*, *flood* and *flooding* (see Figure 7). When users click on one of these terms, they retrieve the corresponding term entry, which shows the phraseological entries and subentries (cf. Table 9).

Finally, phraseological correspondences (e.g. English-Spanish verb collocations), can be accessed through the equivalent term entry. For instance, the terms in Figure 6, which designate HURRICANE in different languages (*hurricane*, *huracán*, *Hurrikan*, *Tropensturm*, etc.), have a similar semantic classification of verb collocations since they share conceptual features. This means that verbs related to *huracán* [hurricane] and *hurricane* are structured in similar lexical domains and subdomains. Consequently, they are stored in EcoLexicon in similar phraseological entries and subentries. If translators are looking for a Spanish collocation with *huracán* to convey the meaning *to cause to change sth for the worse*, the answer can be found under the same dimension in the entry for *huracán* (see Table 12).


**Figure 6:** Search for collocations via the base *hurricane*

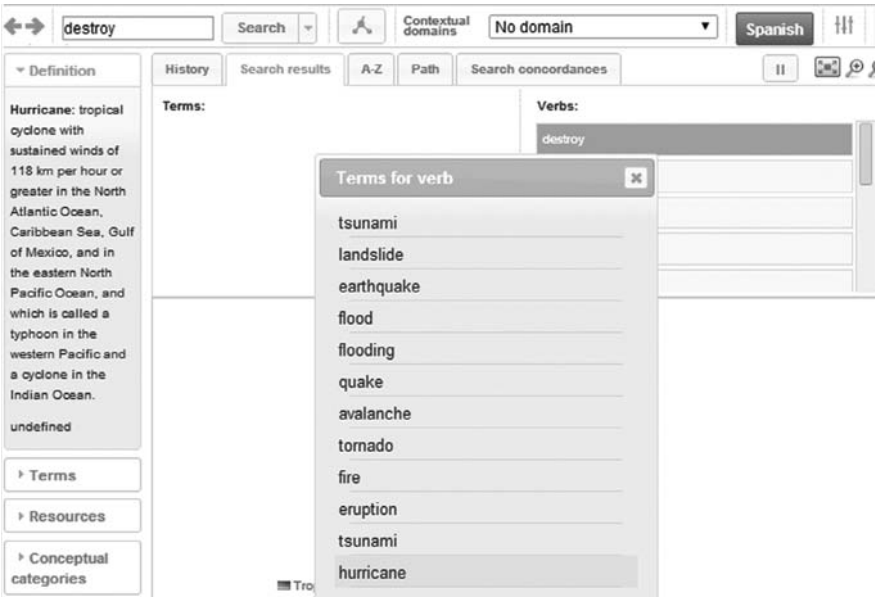
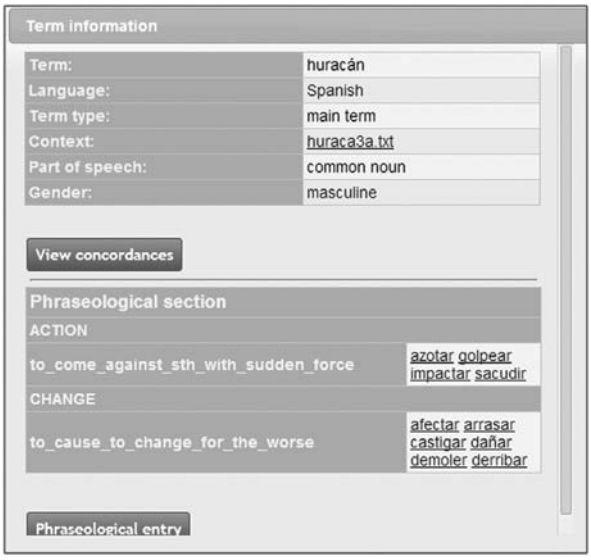



**Figure 7:** Search for collocations via the collocate *destroy*

**Table 12:** Extract of the Spanish term entry for *huracán*



However, Table 13 shows a more detailed description of the phraseological entry for *huracán*, where the template includes the verbs *afectar*, *arrasar*, *castigar*, *dañar*, *demoler*, and *derribar*.

In order to select a suitable collocate in the target language, the user often needs more information about meaning restrictions and/or linguistic usage contexts. As shown in Table 14, the note section for the Spanish *castigar* [damage], as a collocate for the term *huracán*, restricts the potential patients to constructions and geographical areas. This is in evident contrast to the normal use of *castigar* in general language, in which the second argument is generally a human patient.

However, to offer more dynamic bilingual (and multilingual) searches in EcoLexicon, meaning dimensions in one language are currently being linked to meaning dimensions in the others. There would thus be a direct correspondence between verb collocations. From a computational perspective, this is an extremely complex task since it means restructuring the links in the whole knowledge base.

**Table 13:** Spanish phraseological template

| | |
|---|---|
| Nuclear meaning | CHANGE |
| Meaning dimension | to_cause_to_change_for_the_worse |
| Dimension specification | NATURAL DISASTER causes a PATIENT to change for the worse. |
| Verbs | afectar, arrasar, castigar, dañar, demoler, derribar |

**Table 14:** Verb details for *castigar* in EcoLexicon

Detalles verbo

| | |
|---|---|
| Verb | castigar |
| Usage examples | 1. El huracán castigó la parte sur del país el pasado jueves.<br>2. Huracán castiga las Bahamas y amenaza el noreste de Estados Unidos..<br>3. 'Sandy' castiga la costa este de EE.UU.<br>4. Huracán Irene castiga a las Bahamas en su ruta hacia los Estados Unidos.<br>5. Antes de llegar a las Islas Caimán, el ciclón castigó a Jamaica, dejando al menos dieciséis muertos, miles de damnificados y cuantiosos daños. |
| Note | The PATIENT is normally a construction, or area. |

# 5 Conclusions

Any analysis of terminological and phraseological units in specialized translation must account for semantic, pragmatic, contextual, and cultural factors. It goes without saying that the textual adequacy of verb collocations, which is central to discourse, is a determining factor in translation quality.

In specialized language, verb meaning is restricted because of the constraints of specialized subject fields. Consequently, if arguments are classified and structured in a set of conceptual-semantic categories typical of a given domain, along with the semantic roles activated, the range of verbs generally associated with a certain category could be predicted within the framework of a specialized event. For this purpose, in order to identify verb collocations within the EXTREME EVENT and the category NATURAL HAZARD, verb candidates were analyzed by studying their activation in texts, as reflected in corpus concordances. They were further classified based on their definitions in order to provide the most prominent lexical domains activated within the EXTREME EVENT. Subsequently, an analysis of terms belonging to the category NATURAL HAZARDS was also performed in order to specify verb-argument patterns. As a result, linguistic realizations of arguments were identified and described in terms of the following: (i) semantic and thematic roles; (ii) semantic category; (iii) morphological and syntactic descriptions. All of this information was encoded in verb templates.

However, in order to provide access to conceptual and linguistic information that is so necessary for translation, phraseological entries and subentries in EcoLexicon are conceptually bound to a term entry. This enables users to situate verb collocations within specific conceptual categories and larger specialized (sub)domains. With a view to providing both conceptual and linguistic knowledge, collocations are classified and described in terms of their nuclear meaning, meaning dimension, dimension specification, meaning restrictions, and linguistic usage contexts. To facilitate user understanding, the microstructure of the entries does not contain difficult metalanguage regarding semantic roles and macroroles. The macrostructure of the entries permits the semasiological and onomasiological access to collocations, which is helpful both for the translation of a collocation from L1 to L2 and for the production of a collocation in L2 with a specific meaning.

In conclusion, this kind of relational representation of phraseological information, based on conceptual events and categories, facilitates knowledge acquisition for textual processing and production since, in written communication, the perceiver's knowledge of conceptual events plays a central role in sentence processing (Faber 2011: 16). The noun phrases in verb collocations, whether objects

or processes, are powerful cues for the wider event knowledge. In this regard, the choice of a term is sufficient to generate expectations and predictions of the range of events in which the term is likely to appear. This in turn also constrains the potential verb collocates that combine with it.

*University of Granada, Spain*

## References

Baroni, Marco & Silvia Bernardini. 2006. *Wacky! Working Papers on the Web as Corpus*. Bologna: GEDIT.

Barsalou, Lawrence W. 2003. Situated Simulation in the Human Conceptual System. *Language and Cognitive Processes* 18(5–6). 513–562.

Benito, Gerardo & Andrés *Díez* Herrero. 2004. *Riesgos naturales y antrópicos en geomorfología: libro de Actas de los simposios desarrollados durante la VIII Reunión Nacional de Geomorfología,* Toledo 22–25 de septiembre de 2004. Madrid: Sociedad Española de Geomorfología.

Benson, Morton, Evelyne Benson & Robert F. Ilson. 1986. *Lexicographic Description of English. Studies in Language Companion Series*. Volume 14. Amsterdam/Philadelphia: John Benjamins.

Benson, Morton, Evelyne Benson & Robert F. Ilson. 2009. *The BBI Combinatory Dictionary of English. 3rd edition*. Amsterdam/Philadelphia: John Benjamins.

Bergenholtz, Henning & Sandro Nielsen. 2002. Terms in the Language of Culture-Dependent LSP Dictionaries. *Lexicographica. International Journal of Lexicography* 18. 5–18.

Bergenholtz, Henning & Sven Tarp. 2010. LSP Lexicography or Terminography? The Lexicographer's Point of View. In Pedro Antonio Fuertes-Olivera (ed.), *Specialized Dictionaries for Learners*, 27–37. Berlin & New York: Mouton de Gruyter.

Bonet, Josep. 2002. La calidad de la traducción según sus objetivos. In Joaquín García Palacios & María Teresa Fuentes Morán (eds.), *Texto, terminología y traducción*, 169–188. Salamanca: Ediciones Almar.

Bryant, Edward. 2005. *Natural Hazards*. 2nd edition. Cambridge: Cambridge University Press.

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

Buendía Castro, Miriam & José Manuel Ureña Gómez-Moreno. 2009. Parameters of Evaluation for Corpus Design. *International Journal of Translation* 21(1–2). 73–88.

Cabré Castellví, María Teresa. 1998. Las Fuentes Terminológicas para la Traducción. En Purificación Fernández Nistal & José María Bravo Gonzalo (eds.), *La Traducción: Orientaciones Lingüísticas y Culturales*, 27–59. Universidad de Valladolid: SAE.

Chacón Montero, José & Clemente Irigaray Fernández (eds.). 1996. *VI Congreso Nacional y Conferencia Internacional de Geología Ambiental y Ordenación del Territorio. Riesgos Naturales, Ordenación del Territorio y Medio Ambiente = VI Spanish Congrress and International Conference on Environmental Gegology and Land-use Planning: Natural Hazards, Land-use Planning and Environment.* Madrid: Sociedad Española de Geologia Ambiental y Ordenación del Territorio.

Corpas Pastor, Gloria. 2003. *Diez años de investigación en fraseología: análisis sintáctico-semánticos, contrastivos y traductológicos*. Frankfurt am Main, Madrid: Vervuert, Ibero-America.

Cowie, Anthony Paul. 1998. *Phraseology. Theory, Analysis, and Applications*. Oxford: Clarendon Press.

De Schryver, Gilles-Maurice. 2002. Web For/as Corpus: a Perspective for the African Languages. *Nordic Journal of African Studies* 11. 266–282. http://tshwanedje.com/publications/webtocorpus.pdf (accessed 9 January 2014 ).

Drouin, Patrick. 2003. Term extraction using non-technical corpora as a point of leverage. *Terminology* 9(1). 99–117.

European Environment Agency. 2010. *Mapping the Impacts of Natural Hazards and Technological Accidents in Europe*. Luxembourg: Publications Office of the European Union.

Faber, Pamela. 2009. The Cognitive Shift in Terminology and Specialized Translation. *MonTI. Monografías De Traducción e Interpretación* 1(1). 107–134.

Faber, Pamela. 2011. The Dynamics of Specialized Knowledge Representation: Simulational Reconstruction or the Perception-action Interface. *Terminology* 17(1). 9–29.

Faber, Pamela (ed.). 2012. *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin & Boston: Mouton de Gruyter.

Faber, Pamela & Ricardo Mairal. 1999. *Constructing a Lexicon of English Verbs*. Berlin & New York: Mouton de Gruyter.

Faber, Pamela, Carlos Márquez & Miguel Vega. 2005. Framing Terminology: a Process-oriented Approach. *META* 50(4). CD-ROM. http://www.erudit.org/livre/meta/2005/000255co.pdf (accessed 10 April 2013).

Fernández, Ana & Gloria Vázquez. 2012. Análisis Cuantitativo del Corpus SenSem. In Izaskun Elorza, Ovidi Carbonell i Cortés, Reyes Albarrán, Blanca García Riaza & Miriam Pérez-Veneros (eds.), *Empiricism and Analytical Tools for 21st Century Applied Linguistics*, 157–170. Salamanca: Ediciones de la Universidad de Salamanca. http://grial.uab.es/archivos/analisis cuantitativo del corpus sensem - aesla 2011 - def.pdf (accessed 22 October 2013).

Fillmore, Charles J. 1976. Frame Semantics and the Nature of Language. In Stevan R. Harnad, Horst D. Steklis & Jane Lancaster (eds.), *Origins and Evolution of Language and Speech*, 20–32. Annals of the NY Academy of Sciences, Vol. 280.

Fillmore, Charles J. 1982. Frame Semantics. In The Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm*, 111–137. Seoul: Hanshin.

Fillmore, Charles J. 1985. Frames and the Semantics of Understanding. *Quaderni Di Semantica* 6. 222–254.

Fillmore, Charles J. & Beryl T. Atkins. 1992. Toward a Frame-based Lexicon: The Semantics of RISK and its Neighbors. In Adrienne Lehrer & Eva Feder Kittay (eds.), *Frames, Fields and Contrasts: New Essays in Semantic and Lexical Organization*, 75–102. Hillsdale: Erlbaum.

Fillmore, Charles J., Miriam R.L. Petruck, Josef Ruppenhofer & Abby Wright. 2003. FrameNet in Action: The Case of Attaching. *International Journal of Lexicography* 16(3). 297–332.

Fletcher, William H. 2007. Implementing a BNC-Compare-able Web Corpus. In Cédrick Fairon, Hubert Naets, Adam Kilgarriff & Gilles Maurice De Schryver (eds.), *Building and Exploring Web Corpora*, 43–56. Lovain-la-Neueve: Cahiers du Central.

Floros, Georgios. 2002. *Kulturelle Konstellationen und ihre Ubersetzung*. Tubingen: Narr.

Fuertes Olivera, Pedro Antonio, Silvia Montero Martínez & Mercedes García de Quesada. 2005. Modelos Culturales y Discursivos en la Traducción de Textos de Comercio Internacional: Problemas y Soluciones. *BABEL Revue internationale de la traduction* 51(4). 357–379.

Gläser, Rosemarie. 1986. *Phraseologie der englischen Sprache*. Leipzig: VEB Verlag Enzyklopädie.

Gläser, Rosemarie. 1994/5. Relations between Phraseology and Terminology with Special Reference to English. *ALFA* 7(8). 41–60.

Gläser, Rosemarie. 1998. The Stylistic Potential of Phraseological Units. In Anthony Paul Cowie (ed.), *Phraseology: Theory, Analysis, and Applications*, 125–143. Oxford: Clarendon Press.

Goldberg, Adele. 1995. *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, Adele. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Grinev, Sergei & Igor Klepalchenko. 1999. Terminological Approach to Knowledge Representation. In *TKE '99: Proceedings of the 5th International Congress on Terminology and Knowledge Engineering*, 147–151. Vienna: TermNet.

Gross, Gaston. 1994. Classes d'objets et description des verbes. *Langages* 115. 15–30.

Harre, Rom & Luk Van Langenhove (eds.). 1999. *Positioning Theory: Moral Contexts of Intentional Action*. Malden: Blackwell.

Hausmann, Franz Josef. 1984. Wortschatzlernen ist Kollokationslernen: Zum Lehren und Lernen französischer Wortverbindungen. *Praxis des neusprachlichen Unterrichts* 31. 395–406.

Hausmann, Franz Josef. 1985. Kollokationen im deutschen Wörterbuch: ein Beitrag zur Theorie des lexikographischen Beispiels. In Henning Bergenholtz & Joachim Mugdan (eds.), *Lexikographie und Grammatik*, 118–129. Tübingen: Niemeyer.

Hausmann, Franz Josef. 1989. Le dictionnaire de collocations. In Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand & Ladislav Zgusta (eds.), *Wörterbücher / Dictionaries / Dictionnaires — Ein internationals Handbuch zur Lexikographie / An International Enyclopedia of Lexicography / Enyclopédie internationale de lexicographie*, 1010–1019. Berlin & New York: Walter de Gruyter.

Heid, Ulrich. 1994. On Ways Words Work Together –Topics in Lexical Combinatorics. In Willy Martin, Willem Meys, Margreet Moerland, Elsemiek ten Pas, Piet van Sterkenburg and Piek Vossen, (eds.), *Euralex 1994: Proceedings*. Amsterdam: International Congress on Lexicography. 226–257.

Heid, Ulrich. 1999. Extracting Terminologically Relevant Collocations from German Technical Texts. In K-D. Schmitz, (ed.), *TKE'99: Terminology and Knowledge Engineering*. Frankfurt: INDEKS-Verlag. 241–255.

Heid, Ulrich. 2001. Collocations in sublanguage texts: Extraction from corpora. In Wright, S.E. (ed.). *Handbook of Terminology Management*: *Applications-Oriented Terminology Management*. Vol. II. Amsterdam/Philadelphia: John Benjamins. 788–808.

Keller, Edward A. & Robert H. Blodgett. 2007. *Riesgos naturales: procesos de la Tierra como riesgos, desastres y catástrofes*. Madrid: Pearson Educación. [translation by Pilar Gil Ruiz].

Kipper, Karin. 2005. *VerbNet: a broad-coverage, comprehensive verb lexicon*. Pennsylvania: University of Pennsylvania PhD Thesis. http://verbs.colorado.edu/~kipper/Papers/dissertation.pdf (accessed 24 May 2013).

León, Pilar & Antonio San Martín. 2011. Distinguishing Polysemy from Contextual Variation in Terminological Definitions. In María Luisa Carrió, Josefa Contreras, Françoise Olmo, Hanna Skorczynska, Inmaculada Tamarit & Debra Westall (eds.), *Actas Del X Congreso de*

*la asociación europea de lenguas para fines específicos: la investigación y la enseñanza aplicadas a las lenguas de especialidad y a la tecnología*, 173–186. Valencia: Universitat Politècnica de València. http://lexicon.ugr.es/pub/leo-dis (accessed 15 January 2013).

León, Pilar & Pedro Javier Magaña. 2010. EcoLexicon: Contextualizing an Environmental Ontology. In *Proceedings of the Terminology and Knowledge Engineering (TKE) Conference 2010*. Dublin. http://lexicon.ugr.es/pdf/leonmagana2010.pdf (accessed 12 February 2013).

León, Pilar & Arianne Reimerink. 2010. Knowledge Extraction and Multidimensionality in the Environmental Domain. In *Proceedings of the Terminology and Knowledge Engineering (TKE) Conference 2010*. Dublin: Dublin City University. http://lexicon.ugr.es/pdf/leonreimerink2010.pdf (accessed 12 February 2013).

L'Homme, Marie Claude. 2000. Understanding Specialized Lexical Combinations. *Terminology* 6(1). 89–110.

L'Homme, Marie Claude. 2009. A methodology for describing collocations in a specialized dictionary. In Sandro Nielsen & Sven Tarp (eds.), *Lexicography in the 21st century*, 237–256. Amsterdam/Philadelphia: John Benjamins.

L'Homme, Marie Claude. 2010. Designing Terminological Dictionaries for Learners Based on Lexical Semantics: The Representation of Actants. In Pedro Antonio Fuertes-Olivera (ed.), *Specialised Dictionaries for Learners*, 141–153. Berlin & New York: Mouton de Gruyter.

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

Lorente Casafont, Mercè. 2000. Tipología Verbal y Textos Especializados. In Miguel González Pereira & Montserrat Souto Gómez (eds.), Cuestiones conceptuales y metodológicas de la lingüística, 143–153. Santiago de Compostela: Universidad de Santiago de Compostela. http://www.upf.edu/pdi/iula/merce.lorente/docums/ml01.pdf (accessed 22 February 2013).

Lorente Casafont, Mercè & Cleci Regina Bevilacqua. 2000. Los verbos en las aplicaciones terminográficas. In *Actas del VII Simposio Iberoaméricano de Terminología RITERM*, Lisbon 14–17 November. París: Unión Latina.

Malmkjær, Kirsten. 2000. Multidisciplinarity in process research. In Sonja Tirkkonen-Condit & Riitta Jääskeläinen (eds.), *Tapping and mapping the processes of translation and interpreting. Outlooks on empirical research,* 163–170. Amsterdam/Philadelphia: John Benjamins.

Martin, Willy. 1992. Remarks on Collocations in Sublanguages. *Terminologie et Traduction* 2(3). 157–164.

McGuire, Bill, Ian M. Mason, Ian Mason & Christopher R.J. Kilburn. 2002. *Natural Hazards and Environmental Change*. New York: Oxford University Press.

Mel'čuk, Igor. 1996. A Tool for the Description of Lexical Relations in the Lexicon. In Leo Wanner (ed.), *Lexical Functions in Lexicography and Natural Language Processing*, 37–71. Amsterdam/ Philadelphia: John Benjamins.

Mel'čuk, Igor et al. 1984–1999. *Dictionnaire explicatif et combinatoire du français contemporain. Recherche lexico-sémantiques I, II, III, IV.* Montréal: Les Presses de l'Université de Montréal.

Mel'čuk, Igor, André Clas & Alain Polguère. 1995. *Introduction à la Lexicologie Explicative et Combinatoire*. Lovain-la-Neueve, Belgium: Duculot/Aupelf-UREF.

Meyer, Ingrid & Kristen Mackintosh. 1994. Phraseme Analysis and Concept Analysis: Exploring a Symbiotic Relationship in the Specialized Lexicon. In Willy Martin, Willem Meijs, Margreet Moerland, Elsemiek Ten Pas, Piet Van Sterkenburg & Piek Vossen (eds.), *EURALEX 1994 proceedings. Papers read at the Sixth EURALEX International Congress on Lexicography*, 339–348. Vrije Universiteit, Amsterdam: EuraLex.

Meyer, Ingrid & Kristen Mackintosh. 1996. Refining the terminographer's concept-analysis methods: How can phraseology help? *Terminology* 3(1). 1–26.

Montero Martínez, Silvia. 2003. *Estructuración Conceptual y Formalización Terminográfica de Frasemas en el Subdominio de la Oncología*. Vol. 19. CSIC/Elies: Madrid. http://elies.rediris.es/elies19 (accessed 10 January 2014).

Montero Martínez, Silvia. 2008. A Constructional Approach to Terminological Phrasemes. In Elisenda Bernal & Janet DeCesaris (eds.), *Proceedings of the XIII International EURALEX Congress*, 1015–1022. Barcelona: IULA.

Montero Martínez, Silvia. & Miriam Buendía Castro. 2010. Las construcciones fraseológicas en la Terminología basada en Marcos. Plenary session presented at the *Seminario de lingüística española: los retos informáticos del español en Lexicología, Terminología y Fraseología*, University of Antwerpen, 26–27 Februrary.

Montero Martínez, Silvia & Miriam Buendía Castro. 2012. La sistematización en el tratamiento de las construcciones fraseológicas: el caso del medio ambiente. In Izaskun Elorza, Ovidi Carbonell i Cortés, Reyes Albarrán, Blanca García Riaza & Miriam Pérez-Veneros (eds.), *Empiricism and Analytical Tools for 21st Century Applied Linguistics*, 711–724. Salamanca: Ediciones de la Universidad de Salamanca.

Montero Martínez, Silvia & Pamela Faber. 2009. Terminological Competence in Translation. *Terminology* 15(1). 88–104.

Montero Martínez, Silvia, Pedro Antonio Fuertes Olivera & Mercedes García de Quesada. 2001. The translator as "Language Planner": Syntactic Calquing in an English-Spanish Translation of Chemical Engineering. *META* 46. 687–98.

Montero Martínez, Silvia, Mercedes García de Quesada & Pedro Antonio Fuertes Olivera. 2002. Terminological Phrasemes in OntoTerm: A new Theoretical and Practical Approach. *Terminology* 8(2). 177–206.

Olcina Cantos, Jorge. 2006a. *¿Riesgos Naturales? Vol.1*. Barcelona: Davinci.

Olcina Cantos, Jorge. 2006b. *¿Riesgos Naturales? Vol.2*. Barcelona: Davinci.

Pawley, Andrew. 2001. Phraseology. Linguistics and the Dictionary. *International Journal of Lexicography* 14(2). 122–134.

Prieto Velasco, Juan Antonio. 2009. *Traducción e imagen: la información visual en textos especializados*. Granada: Tragacanto.

Prieto Velasco, Juan Antonio & Clara Inés López Rodríguez. 2009. *Managing Graphic Information in Terminological Knowledge Bases*. *Terminology* 15(2). 179–213.

Reimerink, Arianne, Mercedes García de Quesada & Silvia Montero Martínez. 2010. Contextual Information in Terminological Knowledge Bases: A Multimodal Approach. *Journal of Pragmatics* 42(7). 1928–1950.

Reimerink, Arianne, Mercedes García De Quesada & Silvia Montero Martínez. 2012. Contextual Selection for Term Entries. In Pamela Faber (ed.), *A Cognitive Linguistics View of Terminology and Specialized Language*, 207–223. Berlin & Boston: Mouton De Gruyter.

Roberts, Roda. 1994/1995. Identifying the Phraseology of LSPs. *ALFA* 7(8). 61–73.

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

Seibel, Claudia & Catalina Jiménez Hurtado. 2002. La pragmática de la terminología: en busca del perfil del usuario. In Pamela Faber & Catalina Jiménez Hurtado (eds.), *Investigar en Terminología,* 91–123. Granada: Comares.

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

Sinclair, John. M., Susan Jones & Robert Daley. (1970/2004). *English collocation studies: The OSTI report* (R. Krishnamurthy, ed.). London: Continuum.

Tercedor Sánchez, María Isabel. 1999*: La Fraseología en el Lenguaje Biomédico: Análisis desde las Necesidades del Traductor*. Granada: University of Granada PhD Thesis. http://elies.rediris.es/elies6 (accessed 10 January 2014).

Tercedor Sánchez, María Isabel, Pamela Faber & Clara Inés López Rodríguez. 2012. Working with Words: Research Approaches in Translation-oriented Lexicographic Practice. *TTR: Traduction, Terminologie, Rédaction.* XXV(1). 181–214.

United Nations. 2009. *UNISDR Terminology on Disaster Risk Reduction*. Geneva: United National International Strategy for Disaster Reduction.

Vaamonde, Gael, Fita González & Jose M. García. 2010. ADESSE. A Database with Syntactic and Semantic Annotation of a Corpus of Spanish. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner & Daniel Tapias (eds.), *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10),* 1903–1910. Valletta: European Language Resources Association (ELRA).

Van Valin, Robert D.Jr. 2005. *The Syntax-semantics-pragmatics Interface: an Introduction to Role and Reference Grammar*. Cambridge: Cambridge University Press.

Van Valin, Robert D.Jr. & Randy LaPolla. 1997. *Syntax: Structure, Meaning and Function*. Cambridge: Cambridge University Press.

Wisner, Ben, Piers M. Blaikie, Terry Cannon & Ian Davis. 2004. *At Risk. Natural Hazards, People's Vulnerability and Disasters*. Second Edition. London and New York: Routledge.

# Appendix 1: General characteristics of the English and Spanish corpus

| | English corpus | Spanish corpus |
|---|---|---|
| **Number of tokens** | 779, 995 | 449,416 |
| **Number of types** | 26,285 | 31,230 |
| **Type/token ratio** | 3.71 | 7.24 |
| **Standardized TTR** | 40.61 | 41.28 |
| **Number of texts** | 45 | 44 |
| **Medium** | Written | |
| **Topic/domain** | Extreme event | |
| **Publication date** | from 1996 to 2012 | |
| **Language** | English | Spanish |
| **Source of text** | **Specialized journals**<br>– *Natural Hazards*<br>– *Natural Hazards Review* | **Conference proceedings**<br>– *Riesgos naturales, ordenación del territorio y medio ambiente*. VI Congreso Nacional y Conferencia Internacional de Geología Ambiental y Ordenación del Territorio (Chacón and Irigaray 1996)<br>– *Riesgos naturales y antrópicos en geomorfología*. Libro de actas de los simposios desarrollados durante la VIII Reunión Nacional de Geomorfología (Benito and Díez 2004) |
| | **Monographs**<br>– McGuire et al. (2002)<br>– Wisner et al. (2004)<br>– Bryant (2005)<br>– United Nations (2009)<br>– European Environment Agency (2010) | **Monographs**<br>– Olcina (2006a)<br>– Olcina (2006b)<br>– Keller and Blodget (2007)<br>– United Nations (2009) |
| | **Scientific magazines**<br>– *National Geographic*<br>– *ScienceDaily* | **Scientific magazines**<br>– *National Geographic*<br>– *Muy Interesante* |