Verb Collocations for Natural Disasters: a Contrastive Study of Mexican and Peninsular Spanish

Míriam Buendía Castro | Pamela Faber
miriam.buendia@uclm.es | pfaber@ugr.es
University of Castilla-La Mancha | University of Granada

Recibido: 08/12/2016 | Revisado: 06/02/2017 | Aceptado: 07/07/2017

Abstract

This research shows how to identify the diatopic verb phraseological differences between Mexican and peninsular Spanish in environmental texts, and more specifically, within the domain of natural disasters. The study is performed by analyzing five Mexican dictionaries as well as a specialized corpus. The ENVIRONMENTAL EVENT (EE) and the sematic category of natural disasters were organized based on the premises of Frame-based Terminology (FBT) (Faber 2009, 2011, 2012). In FBT, the EE represents and configures the most generic categories within the field of environmental science. Semantic categories in FBT are generalizations of a set of terms that have a similar semantic and syntactic behavior. To detect diatopic variants an integrated top-down and bottom-up approach was followed. As such, all the potential members of the semantic category NATURAL DISASTER in Spanish were searched in various dictionaries (top-down approach). Subsequently, these terms were extracted and analyzed in corpus texts (bottom-up approach) to find the most frequent verb collocations and argument patterns. This research highlights that phraseological diatopic varieties exist at a morphosyntactic, morphological, and lexical level in specialized discourse. The conclusion is that specialized dictionaries and other terminographic resources should incorporate these varieties so that users can become aware of them and use them when needed.

Keywords: phraseology; diatopic varieties; specialized language; Mexican Spanish; Peninsular Spanish

Resumen

Colocaciones verbales en el subdominio de los desastres naturales: un estudio contrastivo del español mexicano y peninsular

El presente estudio describe una metodología para identificar, mediante el análisis de cinco diccionarios mexicanos y un corpus especializado, las diferencias diatópicas verbales entre el español peninsular y el español mexicano en el dominio del medio ambiente y, más concretamente, en el subdominio de los desastres naturales, Para ello, se configuró el EVENTO MEDIOAMBIENTAL (EM) y las categorías semánticas de los desastres naturales en consonancia con los principios de la Terminología basada en Marcos (TBM) (Faber 2009, 2011, 2012). En la TBM, el EM representa y configura las categorías más generales del dominio de las ciencias ambientales. Las categorías semánticas en la TBM son generalizaciones para un conjunto de términos que presentan un comportamiento semántico y sintáctico similar. Con el fin de poder detectar variantes diatópicas, se siguió tanto un enfoque *top-down* como *bottom-up*. De esta forma, se buscaron en los diccionarios todos los posibles candidatos a

término de la categoría semántica de DESASTRE NATURAL en español (*top-down*). Seguidamente, se analizaron esos términos en los textos del corpus (*bottom-up*) para encontrar las colocaciones verbales más frecuentes y sus argumentos. Esta investigación pone de manifiesto que las variedades diatópicas existen a nivel morfosintáctico, morfológico y léxico en el discurso especializado y concluye con la idea de que los diccionarios especializados y otros recursos terminográficos deberían incluir estas variedades diatópicas para que sus usuarios pudieran aprender a usarlas adecuadamente en función de sus necesidades.

Palabras clave: fraseología; variedades diatópicas; lenguaje especializado; español mexicano; español peninsular

1. Introduction

Spanish is one of the most widely spoken languages in the world. There are currently 470 million people whose mother tongue is Spanish, which makes Spanish the second most important mother tongue by number of speakers, after Chinese. In addition, it has become the second most studied language after English, and the third most frequent language on the Internet after Chinese and English. In other words, 7.8% of the 2045 million of Internet users all over the world communicate in Spanish. Spanish is the official language of 21 countries, which in descending order of population are the following: Mexico, Colombia, Spain, Argentina, Peru, Venezuela, Chile, Ecuador, Guatemala, Cuba, Dominican Republic, Bolivia, Honduras, Paraguay, El Salvador, Nicaragua, Costa Rica, Puerto Rico, Uruguay, and Equatorial Guinea. In fact, Mexico alone has nearly 120 million Spanish speakers, and Colombia and Spain, approximately 47 million each[1].

This huge number of speakers along with the diversity and size of the countries where Spanish is spoken makes Spanish an extraordinarily rich language with a wide range of diatopic or geographic varieties. In many Spanish-speaking countries, part of the population speaks indigenous languages[2], and this influence is also reflected in Spanish phraseology (Mogorrón Huerta 2014b: 87). Coseriu (1999: 301-302) accurately affirmed that no historical language is composed of only one system, but rather is the result of diatopic (dialectal) differences, diastratic (sociolectal) differences, and diaphasic (register) differences.

Generally speaking, diatopic varieties of language are the geographical varieties of a language. To date, the vast majority of studies of diatopic varieties have focused on morphosyntactic, morphological, and lexical differences of general language that designate everyday entities, attributes, and processes (Blanco 2011, 2015; Mogorrón Huerta 2010, 2014b; Reig Alamillo 2009, *inter alia*). However, considerably less attention has been paid to the geographic variation of specialized and semi-specialized knowledge units, and even less to specialized phraseology in Latin America and Spain. Gallego Hernández (2016) is one of the few studies on diatopic variations in specialized phraseology within the financial domain. In addition, in many phraseological studies, the focus is on the description and analysis of noun-noun or noun-

adjective collocations to the exclusion of other grammatical categories of language, such as verbs, despite the fact that verbs are regarded as the most important category of language (L'Homme 1998).

The objective of this study was to discover whether there are diatopic differences in Mexican Spanish in comparison to peninsular Spanish, and explore whether this type of terminological variants should be included in terminographic resources. This would have the advantage of making users (translators, linguists, students, experts in the field, etc.) aware of these varieties. They would thus be better able to recognize them and decide whether to use them in texts. It is our assertion that language varieties enhance the fluency and naturalness when producing a text.

Since about 80% of the words in discourse are chosen according to the co-selection principle rather than for purely syntagmatic or grammatical reasons (Sinclair 2000: 197), this study focused on verbal collocations within the environmental domain, and more specifically, within the subdomain of natural disasters. As is well known, the Earth is currently experiencing an increasing number of disasters due to both natural hazards and human-induced accidents. Consequently, there is a rising interest and concern for the environment. However, to date, there have been few terminological studies specifically centered on natural hazards. The objective was to expand the scope of the environmental knowledge base, EcoLexicon (ecolexicon.ugr.es), by including combinatorial information regarding Spanish language varieties in environmental texts. This would increase the explanatory capacity of EcoLexicon and increase its value for linguists and users interested in Latin American Spanish. The conclusions derived from this study could also be extended to other terminographic resources within other domains.

This paper is organised as follows. Section 2 provides an overview of the environmental knowledge base EcoLexicon and describes how verb collocations are encoded in it. Section 3 deals with the search of diatopic variants by applying a top-down and bottom-up approach. Section 4 presents some conclusions and future lines of research.

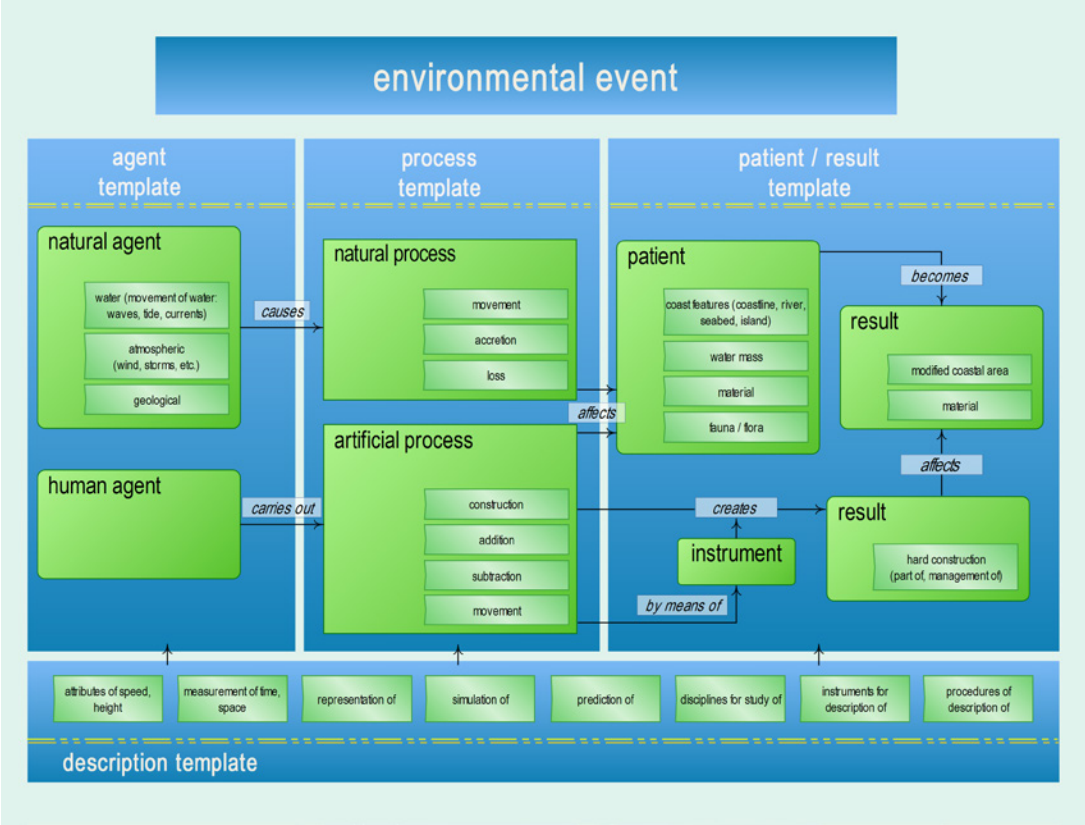
2. The environmental knowledge base EcoLexicon

EcoLexicon (ecolexicon.ugr.es) represents the conceptual structure of the specialized domain of the Environment in the form of a visual thesaurus in which concepts are configured in semantic networks. It is based on the theoretical premises of Frame-based Terminology (Faber 2009, 2011, 2012).

Frame-based Terminology (FBT) uses a modified version of Frame Semantics (Fillmore 1982, 1985; Fillmore & Atkins 1992), along with premises of Corpus Linguistics, and the Lexical Grammar Model (Faber & Mairal 1999) to configure specialized domains on the basis of definitional templates and situated representations for specialized knowledge concepts. Each knowledge area can be said to have its own event template and can be represented accordingly (Grinev & Klepalchenko 1999).

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 2011) and configures the most generic or base-level categories within the field of environmental science. In other 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).

Figure 1. The ENVIRONMENTAL EVENT



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. AGENT, PATIENT, and RESULT 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.

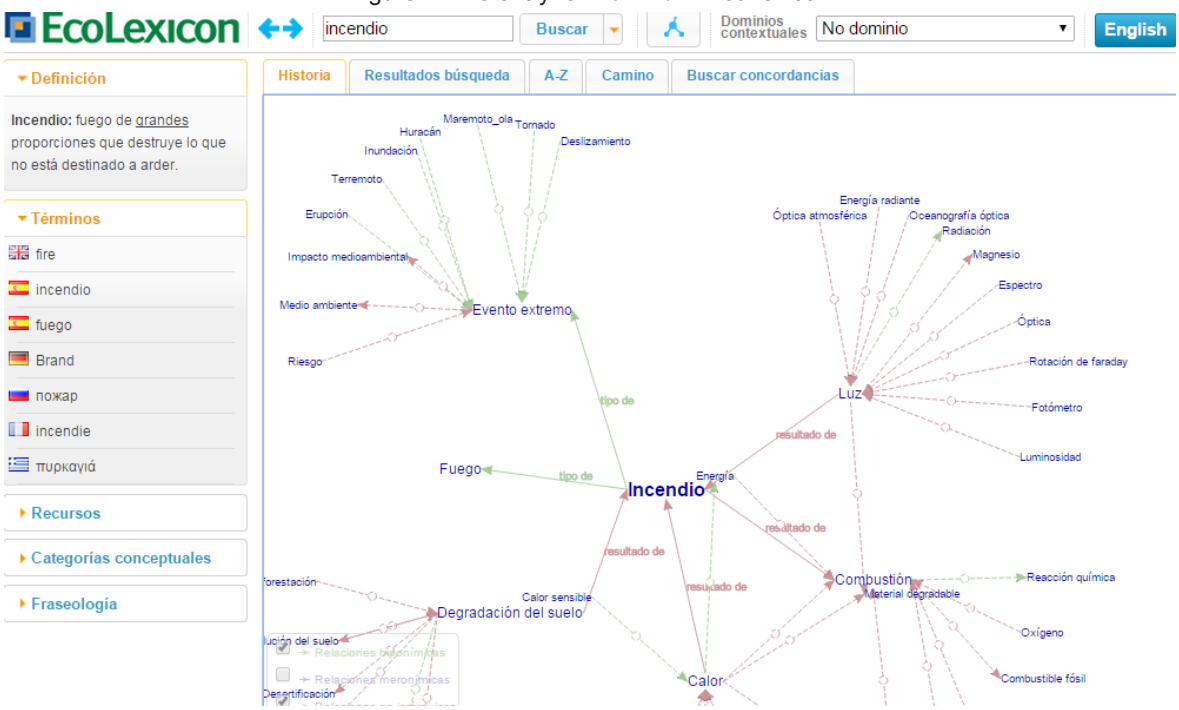
As previously mentioned, the practical application of FBT is EcoLexicon, a visual thesaurus in which the environmental domain is configured in semantic networks and conceptual subdomains. It currently contains more than 3,633 concepts and 20,334 associated terms in six languages, namely, Spanish, English, German, French, Russian, and Modern Greek. In EcoLexicon users are assumed to be familiar with scientific language and its usage in English or Spanish since these are the interface languages.

Potential users should also possess a good command of any of the six languages in EcoLexicon, as well as a minimum of scientific knowledge (López, Buendía & Aragón 2012: 62). Concepts are designated by terms in the six languages contained. 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 (Buendía, Montero & Faber 2014), 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 the Lexical Grammar Model (LGM) lexical domains, and the predicate-argument analysis of phraseological constructions such as verb collocations.

EcoLexicon provides an array of conceptual and linguistic information. Conceptual information is reflected in EcoLexicon in three ways: (i) the semantic network for each concept, which is based on a closed inventory of conceptual relations[3] and which is also lexically represented in the definition; (ii) conceptual categories activated by each entry, which makes it possible to access the classes of the ontology to which the search concept belongs; (iii) graphical resources that are pictorial representations of the concept.

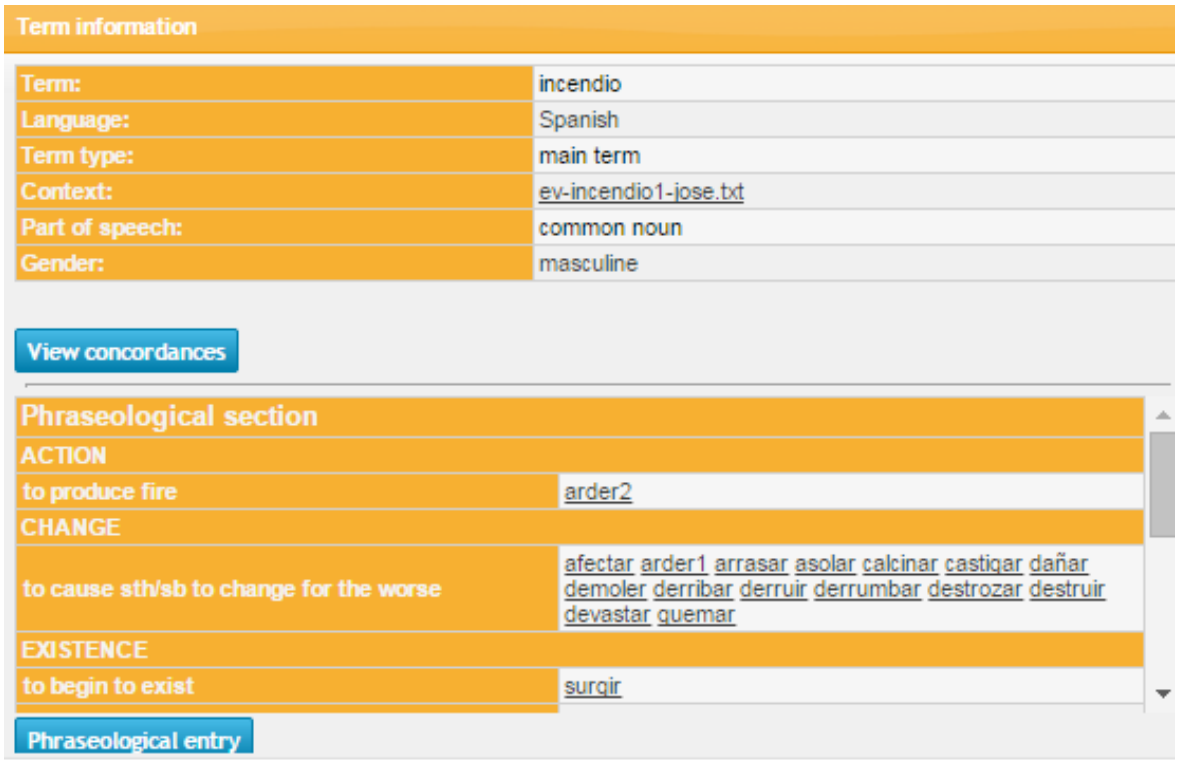
Figure 2 displays the entry for INCENDIO (fire) in EcoLexicon. The right hand side of the screen displays the semantic network of INCENDIO, which links INCENDIO to all related concepts. The menu on the left allows users to access the definition, resources, conceptual categories, and the terms associated with each concept in the six languages, i.e., *fire*, *incendio*, *fuego*, *Brand*, пожар, *incendie*, πυρκαγιά.

Figure 2. The entry for INCENDIO in EcoLexicon



By clicking on each term in the term section *(Términos)* on the left, users can access the following linguistic information (Figure 3): (i) kind of term (main term, synonym, and acronym); (ii) contexts of use; (iii) morphosyntactic information regarding grammatical category (noun, verb, adjective or adverb); (iv) gender (masculine, feminine, and neuter); (v) phraseological section which contains verb collocations; (vi) access to the complete specialized corpus on the environment (by means of the section *Buscar concordancias*), (vii) the complete phraseological entry of the term through the section *Fraseología.*

Figure 3. The term entry of *incendio*



The EcoLexicon corpus was manually compiled by the members of the Lexicon research group, and is the result of more than ten years of work. To date, the corpus contains about 50 million tokens in all of the languages included in EcoLexicon, although the English and Spanish corpora are by far the largest in number. EcoLexicon has developed its own tagging template for each of the texts that conform the corpus in order to improve the quality of searches for all users (see Figure 4). The tagging template takes into account the date of the text, the country of origin, the contextual domain, the level of specificity, the kind of text (article, thesis, book, etc.), the language, its language variant, etc. However, even though language variants are included in the tagging template, the Spanish terms are peninsular Spanish.

Figure 4. Tagging template used in EcoLexicon

2.1. The semantic category of NATURAL DISASTER

Any typology of semantic categories is a topic of debate. There have been many initiatives in general language, such as WordNet (Fellbaum 2006), Gross's (1994) *classes d'objets*, meaning-based resources such as ADESSE (Vaamonde, González & García 2010) or VerbNet (Kipper 2005), and ontologies such as SUMO (Niles & Pease 2001). The frame elements proposed by FrameNet (Fillmore 1982, 1985; Fillmore & Atkins 1992; Ruppenhofer et al. 2010) can also be regarded as categories on the basis of which it is possible to make generalizations about arguments. However, the fact that this inventory is open-ended and relies exclusively on the intuition of the annotator means that the role set in FrameNet lacks descriptive adequacy. In FBT and in EcoLexicon, semantic categories are generalizations for a set of terms that are assumed to have a similar semantic and syntactic behavior.

In specialized language, verb meaning is more restrictive 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 context of a specialized event.

In this study, categories were established based on the following: (i) the semantic relations of the concepts expressed by the linguistic realizations in the corpus; (ii) their verification by means of tests based on Gross's (1994) *classes d'objets.* In EcoLexicon, the first term in the definition of concepts, such as EARTHQUAKE or FIRE, is always *natural disaster,* which reflects their category membership. NATURAL DISASTER, which is thus regarded as a category, is defined in EcoLexicon as follows (Table 1):

Table 1. Definition of NATURAL DISASTER in EcoLexicon

NATURAL DISASTER: adverse event that affects the environment in a relatively short space of time, and which causes human, material, economic or environmental losses, which exceed those affected to cope with it.

The top-level concept NATURAL DISASTER is characterized by the following properties (Table 2):

Table 2. Properties of the concept NATURAL DISASTER

A natural disaster *causes* human/economic/material losses.
A natural disaster *affects* the environment.
A natural disaster *occurs* in a short period of time.

In this regard, whenever a concept fulfils these three entry conditions, it belongs to the category of natural disaster. This is verified with corpus information and pattern-based searches. As such, it can be seen that the referents of the terms 'earthquake' 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 appear in a verb's argument structure, the verb is usually one of change (to make something worse), or existence (to cause something not to exist anymore).

In line with this, the main terms that instantiate the category of NATURAL DISASTER in Spanish include the following: *avalancha, sequía, terremoto, seísmo, sismo, ciclón, ciclón tropical, huracán, tifón, tsunami, maremoto, deslizamiento de tierra, corrimiento de tierra, erupción, volcán, inundación, fuego, incendio, desastre natural, alud, depresión tropical, tormenta tropical,* and *desastre natural.*

2.2. Verb collocations in EcoLexicon

FBT and EcoLexicon take a broad approach to the concept of *collocation.* As such, verb collocations refer to frequent combinations of two or more words following a noun-verb or verb-noun pattern, where the noun is the *base* and the verb is the *collocate.* In this approach, the meaning of the collocate (the verb) is imposed to a certain

extent by the meaning of the base (the noun), but at the same time, the collocate also constrains the arguments that appear with it (Buendía, Montero & Faber 2014: 73). In other words, users normally select a collocate to go with a certain base term, but at the same time, a predicate imposes restrictions on its arguments. Consequently, base and collocate retain their meaning to a certain extent. As such, in the collocation, 'fire burns', the predicate 'burn' only admits noun phrases designating combustible entities. In addition, 'fire' requires a verb designating a combustion process ('burn').

Verb collocations in EcoLexicon are classified according to their meaning. Therefore, they are first classified in terms of their lexical domains (i.e. nuclear meaning) based on the lexical domains of the lexical grammar model (LGM) (Faber & Mairal 1999), and then according to their subdomains (i.e. meaning dimension). As is well known, the LGM divides the lexicon into twelve lexical domains, each of which has one or two generic verbs or superordinates in terms of which all the verbs belonging to the domain can be defined. Table 3 displays the lexical domains within the LGM (in square brackets), along with their superordinate verbs (italics):

Table 3. Lexical domains in the verbal lexicon

(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/stay/put* [POSITION]

Lexical subdomains can be further divided into subdomains. Each subdomain pertains to a certain area of meaning and thus focuses on a different part of its content. Consequently, verbs belonging to the same subdomain share the same nuclear meaning and syntax. Since the object of study was natural disasters, it was not surprising that the lexical domains most prototypically activated were those of EXISTENCE, ACTION, POSITION, and CHANGE, since a natural disaster normally begins to exist, and lasts over time until it ceases to exist [EXISTENCE]. During its existence, it produces a strong impact [ACTION] at a certain location and causes a negative effect, which is damage [CHANGE] (Buendía 2012: 154-157).

Figure 3 (above) displays the partial phraseological subentries for *incendio* [fire]. As shown, one of the lexical domains that fire activates is that of CHANGE, and more specifically, it instantiates the subdomain or dimension *to cause to change for the worse* (lower case) with the nuclear meaning of CHANGE (upper case). The verbs (hyperlinks) in this dimension include *afectar, arder, arrasar, asolar, calcinar, castigar, dañar, demoler, derribar, derruir, derrumbar, destrozar, destruir, devastar* and *quemar.*

By clicking, for example, on *calcinar*, the user has access to four usage examples as well as a note section with information about meaning restrictions (Table 4). In this case, the note states that the NATURAL FORCE is always a fire or an extreme heat entity, and that the PATIENT is normally a construction, human being, area, or plant. SITUATION/EXPERIENCE can also be included as well as LOCATION, TIME and MANNER.

Table 4. Verb examples of *calcinar*

1. El incendio **calcina** al menos siete naves industriales de Azuqueca.
2. 200 hectáreas **calcinadas** en el incendio.
3. Varias personas **calcinadas** en el incendio.
4. Controlado el incendio que ha **calcinado** varias plantas protegidas.
NOTE: The NATURAL FORCE is always a fire or an extreme heat entity. The PATIENT is normally a construction, human being, area, or plant. SITUATION/EXPERIENCE can also be included as well as LOCATION, TIME and MANNER.

In addition to the information in phraseological subentries, the user can display the complete phraseological information by means of the hyperlink *Fraseología (Phraseological entry)* at the bottom of the term entry for *incendio* (cf. Figure 3).

3. Methodology: in search of diatopic variants

An integrated top-down and a bottom-up approach was followed to detect diatopic variants. Accordingly, all the terms initially regarded as members of the category NATURAL DISASTER in Spanish were searched in various dictionaries (top-down approach). Then, these terms were extracted and analyzed in corpus texts (bottom-up approach) to find the most frequent verb collocations and argument patterns.

3.1. Dictionary analysis

As previously mentioned, terms were searched in various dictionaries, not only by lemmas, but also in the context examples given in the entries. The dictionaries used were the following: (1) *Diccionario del español usual en México* (1996) (DEUM), (2) *Diccionario breve de mexicanismos de Guido Gómez de Silva, Academia Mexicana de la Lengua* (2000) (DBM), (3) *Diccionario de Mexicanismos, Academia Mexicana de la Lengua* (2010) (DM), (4) *Diccionario de americanismos, Asociación de Academias*

de la Lengua Española, 2010 (DAMER), and (5) *Diccionario de hispanoamericanismos no recogidos por la Real Academia: (formas homónimas, polisémicas y otras derivaciones morfosemánticas)* (1997) (DEHISP).

Table 5 summarizes the terms included in each dictionary either as a lemma with its entry, or as a context use example, which was helpful for the extraction of the candidate verbs. As shown, an asterisk is included when the specific term appeared in the dictionary, but was activated with a different meaning from the one within the conceptual category of natural disaster.

Table 5. Natural disaster terms in Mexican dictionaries

	DEUM	DBM	DM	DAMER	DEHISP
avalancha					
sequía	√				
terremoto	√	√			
seísmo					
sismo	√				
huracán	√				
tifón					
ciclón	√				
tsunami					
maremoto	√				
corrimiento	√				√*
deslizamiento					
erupción	√				
volcán	√				
inundación	√				
fuego	√	√*			√
incendio	√		√*		√*
desastre	√				
alud					
depresión					
tormenta	√				

The main problem seems to lie in the fact that most of these dictionaries focus on the specific words used in these countries, and thus ignore general terms used in Spanish (Mogorrón 2014a: 141):

> Los diccionarios hispanoamericanos incluyen solamente creaciones usuales en sus países dejando de lado otras expresiones que pertenecerían a un español común por lo que a menudo no se sabe si esas UF que pertenecen al español estándar son conocidas y usadas por los hablantes hispanoamericanos.
> [Latin American dictionaries only include expressions generally used in their countries to the exclusion of, other expressions belonging to general Spanish. Therefore, very often, it is difficult to know whether those phraseological units typical of standard Spanish are known and used by Latin American speakers.]

Of the dictionaries analyzed, the DEUM was the one that provided the most useful information. Our analysis found that many of the verbs included in EcoLexicon as verb collocates of natural disaster terms were also used in Mexican Spanish. This was the case of *arder* el bosque, *extinguir* o *sofocar* un fuego, *abrasar* o *quemar* un terreno, *azotar* la sequía, *asolar* el terremoto, *consumir* el fuego, *destruir* el fuego, *dañar* el fuego, *devorar* el fuego, *escupir* el volcán, *incendiar* el bosque, *propagar* fuego, *extenderse* el fuego, *derrumbar* el terremoto, *atacar* el ciclón, *impactar* el huracán, *desatarse* una tormenta, *remitir* la tormenta, and *estremecer* el terremoto, *inter alia*.

Although this analysis identified variants in the verb collocates in peninsular and Mexican Spanish, it also identified terminological variants. For example, *terremoto* [earthquake] can also be *temblor* (DEUM) (DBM) in Mexico, which in peninsular Spanish is the action and effect of *temblar* [shake]. The differences found were divided into three categories: (i) morphosyntactic differences (i.e. differences in the use of prepositions or determiners); (ii) morphological differences (e.g. singular, plural, suffixes); (iii) lexical differences (i.e. change of the verb used to activate the same meaning). Polysemic diatopic variants were also identified, i.e. identical verb collocates, but whose meaning in Mexico differs from their meaning in Spain. This kind of polysemy, known as *geopolysemy* (Mogorrón 2014b), does not facilitate user comprehension when users are not familiar with these meanings.

Regarding morphosyntactic differences, the DEUM offered the collocate *barrer con* for the term *terremoto* "El terremoto *barrió con* varias ciudadades", where *barrer con* is defined as "acabar con algo o con alguien, no dejar nada de alguna cosa" [to finish completely with something or somebody; to not leave anything of something or somebody]. This would be a variant with regards to peninsular Spanish. In Spain, *barrer* is used with same meaning, but it is not followed by *con* (1):

(1) Mexican Spanish: "El terremoto **barrió con** varias ciudades" (DEUM)
Peninsular Spanish: "El terremoto **barrió** varias ciudades"

This is something to be included in the EcoLexicon database. So far, the verb *barrer* is displayed as a collocation of all the terms that activate the category of natural disaster: *el huracán barrió, el terremoto barrió, el tsunami barrió*, etc. However, now the following step would be to include the variant *barrer con* and specify that it is a construction typical of Mexican Spanish.

Morphological changes were also acknowledged, as evidenced by the collocation *vernirse la tormenta* (DEUM), explained as *suceder algo con intensidad, rápida o repentinamente* (DEUM) [to happen suddenly], e.g. *se vino la tormenta* [the storm happened suddenly]. This is a non-idiomatic phrase in peninsular Spanish. Instead of the reflexive *venirse,* it would be more natural to use *sobrevenir,* which is derived from *venir,* and adds specificity to *to occur* suddenly (2).

(2) Mexican Spanish: "**se vino** la tormenta" (DEUM)
Peninsular Spanish: "**sobrevino** una tormenta"

Regarding lexical changes, various differences were identified. For example, in relation to *huracán* [hurricane], the DUE provides the collocation *crecer el huracán,* with the meaning of "aumentar la intensidad de algo" [increase the intensity of something]. According to the *Diccionario de la Real Academia Española* (DRAE), *crecer* means "dicho de una cosa: Recibir aumento por añadírsele nueva materia: crecer el río, el montón" [said of something, to increase by receiving more material, e.g. a river], but, as can be inferred, it cannot be applied to the increase of the intensity of an atmospheric phenomenon. In this case, the equivalent in peninsular Spanish would simply be an explanation such as *aumentó su intensidad* (3):

(3) Mexican Spanish: "**crecer** el huracán" (DEUM)
Peninsular Spanish: "**aumentar la intensidad** el huracán"

Fuego [fire] also generates lexical differences. The DUE and the DEHISP highlight the phrase *agarrar el fuego*, with the meaning of "prender o arraigar cualquier proceso que comience" (DUE), "incendiarse, dícese especialmente cuando el fuego se propaga con rapidez" [to start to burn and extend quickly], e.g. *El jardín había agarrado fuego con ferocidad* (DEHISP) [The garden had begun to burn fiercely]. In peninsular Spanish, the equivalent would again be an explanatory phrase such as *propagar, extender* rápidamente (4):

(4) Mexican Spanish: "**agarrar** el fuego" (DEUM) (e.g. *El jardín había agarrado fuego con ferocidad*) (DEHISP).
Peninsular Spanish: "**propogarse/extenderse rápidamente** el fuego" (e.g. *El fuego* se había propagado rápidamente *con ferocidad en el jardín*).

Another remarkable difference is reflected in the collocation *romper el fuego*, defined as *comenzar algo de pronto* (DEUM) [to start something suddenly]. In Spain, the phrase would be *el fuego comenzó/se inició de pronto* (5):

(5) Mexican Spanish: "**romper** el fuego" (DEUM)
Peninsular Spanish: "El fuego **se inició/comenzó de pronto"**

Another difference to highlight is the collocation *dominar el incendio* (DEUM), e.g. *los bomberos dominaron el incendio* [the firemen controlled the fire]. Although according to the DRAE, the meaning of *dominar* [to dominate] is *tener dominio sobre algo o alguien,* [to have control over something or somebody], from a pragmatic perspective, a Spaniard would be more like to say *controlar* (6).

(6) Mexican Spanish: "**dominar** el incendio" (DEUM)
Peninsular Spanish: "**controlar** el incendio"

As reflected in our analysis of natural disasters, these dictionaries show more combinatorial differences for general language. Although this was not the focus of this research, we found differences that affect terms such as *fuego.* These include *arrebatar el fuego* (DEUM), *meter la mano en el fuego por alguien* (DEUM), *pegar el fuego* (DEUM), and *tatemar* (DBM). *Arrebatar el fuego* (DEUM) means "apresurar excesivamente el cocimiento de algún alimento poniéndolo en un fuego demasiado fuerte y provocando con ello un mal resultado como crudo por dentro y quemado por fuera" [to hurry the cooking of any food by heating it up a lot, and causing a negative result, such as burning it], e.g. *bájale el fuego al arroz, que lo vas a arrebatar* (DUE) [Lower the fire on the rice because you are going to burn it]. In peninsular Spanish, this meaning would be conveyed by "quemar" [burn].

Regarding the second expression, *meter la mano en el fuego por alguien* (DEUM) [put one's hand in the fire for someone], the difference lies in the use of the verb in this idiom. A Spaniard would have said *poner* instead of *meter*. As a synonym of *quemar* [burn], the DBA gives *tatemar*, which signifies burning something without meaning to, e.g. *si no se aleja se va a tatemar la mano con la parrilla* [if you do not step back, you are going to (accidentally) burn yourself with the grill]. In peninsular Spanish, "quemar" would have been used.

Finally, *pegar el fuego* (DEUM) was identified. This combination is an example of a polysemic diatopic variant or geopolysemic variant in that the same construction is used in peninsular Spanish, but with a different meaning. As such, in Spain, *pegar fuego* means *incendiar, quemar* [to burn], whereas in Mexico, *pegar el fuego* is a familiar expression that is used to denote success.

3.2. Corpus

A corpus of Spanish semi-specialized texts belonging to the subdomain of natural disasters was compiled for this research. It was not easy to find specialized and semi-specialized materials originally written in Spanish since most textbooks in machine-readable format are written in English. In this sense, most of the papers in Mexican specialized environmental journals on the environment are written in English as well. This is the case of journals such as *Atmósfera*[4], and *Revista Internacional de la Contaminación Ambiental*[5]. The corpus includes 100 texts totaling 279,995 tokens. Table 6 describes the source of the texts in the corpus and the URL from which they were extracted.

Table 6. Source of the texts in the corpus

CORPUS	
NAME OF THE PUBLICATION	URL
Boletín info Atmósfera from the *Centro de Ciencias de la Atmósfera* of the Universidad Nacional Autónoma de México (all the publications[9]	http://www.atmosfera.unam.mx/vinculacion/boletin.html
Servicio Sismológico Nacional de México (all the seismic reports)	http://www.ssn.unam.mx/ http://www2.ssn.unam.mx:8080/website/jsp/reportes.jsp
Atlas Nacional de Riesgos de México	http://www.atlasnacionalderiesgos.gob.mx/index.php/riesgos-geolgicos/sismos
Secretaría de Marina de la Armada de México	http://meteorologia.semar.gob.mx/
Newspaper *El Informador*	http://www.informador.com.mx/
Newspaper *Excélsior*	http://www.excelsior.com.mx/
Newspaper *El Universal*	http://www.eluniversal.com.mx/pre-home.html
Newspaper *CNN Expansion*	http://www.cnnexpansion.com

The corpus was searched and analyzed with the terms that instantiate the natural disaster phenomenon, as well as the verbs that were identified with the previous dictionary analysis. This verb extraction confirmed the idiomatic use of the above-mentioned verbs, i.e. *barrer con, agarrar, venirse, crecer, romper,* etc. In addition, more striking differences between Mexican and Peninsular Spanish were highlighted, such as the ones associated with the support verb construction[6] *tocar tierra* [make landfall], when referred to a wind disaster, such as a hurricane. It was ascertained that *tocar tierra* was also used in Mexican Spanish as reflected in the corpus (e.g. *El*

huracán categoría III, tocó tierra a las 23:45 horas). However, two slight differences were also found, the lexical variant *impactar tierra* and the morphosyntactic variant *impactar en tierra* (7). All of them were just lexicalized by the combination *tocar tierra* in Peninsular Spanish:

(7) Mexican Spanish
a. "**tocar tierra**" "El huracán categoría III, tocó tierra a las 23:45 horas"
b. "**impactar tierra**" "El huracán INGRID impactó tierra al Noroeste de La Pesca, Tamps", "Este remanente de Iván se convirtió nuevamente en tormenta tropical el 23 de septiembre y por segunda ocasión impactó tierra firme sobre el extremo suroeste de Louisiana el 24 de septiembre"
c. "**impactar en tierra**" "El centro del huracán impactó en tierra con vientos máximos sostenidos de 260 km/h y rachas de 315 km/h"
Peninsular Spanish: "**tocar tierra**" (for all the examples provided)

Another difference corresponds to the support verb *dejar afectaciones* (e.g. *Huracán 'Marie' deja afectaciones en costas de Colima*). In Spain, the verbs *afectar, dañar* would be used instead (8):

(8) Mexican Spanish
a. "**dejar afectaciones**" "Huracán 'Marie' deja afectaciones en costas de Colima"
b. "**dañar**" "El huracán dañó seriamente los vestíbulos y fachadas de varios hoteles"
c. "**afectar**" "El huracán Marie afecta a seis Estados"
Peninsular Spanish: "**dañar, afectar**" (for all the examples provided)

The last difference corresponds to the combination *reportar un sismo* instead of *registrar un seísmo.* In this example two differences were observed. On the one hand, the use of the verb *reportar* instead of *registrar* to convey the meaning of register, and on the other hand, the use of the noun *sismo* instead of *seísmo* [earthquake]. As previously mentioned, *terremoto* was also referred to as *temblor* in Mexican Spanish. Now, it was also found that Mexicans prefer the use of *sismo* instead of *seísmo* as a synonym of *terremoto.*

(9) Mexican Spanish
a. "**reportar sismo**" "Más de 260 réplicas siguieron tras el primer sismo reportado a las 08:02"
b. "**registrar sismo/temblor**" "El temblor se registró a las 14:13 horas, con epicentro a 57 kilómetros al noreste de Petatlán"
Peninsular Spanish: "**registrar**" (for all the examples provided)

4. Conclusions

This study identified and analysed diatopic verb phraseological differences between Mexican and peninsular Spanish. It focused on verbal collocation diatopic varieties within the environmental domain, and more specifically, within the domain of natural disasters. The main limitations of this analysis were the following: (i) the fact that textbooks and texts in Mexican specialized journals on natural disasters were often written in English; (ii) the fact that most Mexican dictionaries include the words used in Mexico to the exclusion of general language words of standard Spanish; and (iii) geopolysemy (Mogorrón 2014b), i.e. when the same collocation is used in both countries, but with a different meaning.

As is well known, to date, the majority of studies dealing with diatopic differences have focused on the study of general or familiar languages. Despite the fact that the number of diatopic expressions is very low compared to the number of tokens of the corpus or the lemmas included in the various dictionaries, this research concludes that phraseological diatopic varieties exist in more specialized contexts. These varieties are given at a morphosyntactic, morphological and lexical level, and that lexical diatopic varieties are the most prototypical ones. Although this analysis was performed to identify possible variants in the verb collocates used in peninsular Spanish and Mexican Spanish, it also provided us with terminological variants.

This research highlights that specialized dictionaries and other terminographic resources should include these varieties so that their users, such as translators, linguists, students, specialists, etc. can use them, depending on their needs. As such, the results of this study can be used as a starting point for the development of a new research line within the knowledge base EcoLexicon*:* the representation of diatopic phraseological varieties in Spanish in order to make them accessible to a larger number of users. In a parallel way, this methodology can also be extended to other terminographic resources within other domains.

In future research, this methodology will be applied to other Spanish variants and to other subdomains within the general domain of the environment. When this is done, figures and percentages regarding the number of diatopic variants will be provided and an in-depth statistical analysis will be performed. The results will provide valuable insights into the multiple ways that term variants arise and will heighten awareness of their importance in the translation process.

5. Acknowledgements

This research was carried out within the framework of project FF2014-52740-P, *Cognitive and Neurological Bases for Terminology-enhanced Translation* (CONTENT) funded by the Spanish Ministry of Economy and Competitiveness.

6. References

- Barsalou, Lawrence W. (2003). Situated Simulation in the Human Conceptual System. *Language and Cognitive Processes* 18 (5-6), 513–562.
- Blanco, Xavier (2011). Carné de conducir vs licencia de manejar. Les locutions nominales en espagnol dans une perspective diatopique. In *Le figement linguistique: la parole entravée*. J.C. Anscombre & S. Mejri (eds.), 377-390. Paris: Honoré Champion.
- Blanco, Xavier (2015). Variation diatopique des pragmatèmes en espagnol. *Linguisticae Investigationes* 38 (2), 263-275.
- Buendía-Castro, Míriam (2012). Verb Dynamics. *Terminology* 18 (2), 149–166.
- Buendía-Castro, Míriam, Montero-Martínez, Silvia & Faber, Pamela. (2014). Verb collocations and Phraseology in EcoLexicon. In *Yearbook of Phraseology*. K. Kuiper (ed.), 57-94. Berlin: De Gruyter Mouton.
- Coseriu, Eugenio (1999). *Lecciones de Lingüística General* (2nd edition). Madrid: Gredos.
- Faber, Pamela & Mairal, Ricardo (1999). *Constructing a Lexicon of English Verbs*. Berlin & New York: Mouton de Gruyter.
- Faber, Pamela (2009). 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.
- Fellbaum, Christiane (2006). WordNet and Wordnets. In *Encyclopedia of Language and Linguistics*. K. Brown et al. (eds.), 665-670. Oxford: Elsevier.
- Fernández Vítores, David (2014). El español: una lengua viva. Informe 2014. Madrid: Instituto Cervantes. <http://eldiae.es/wp-content/uploads/2014/07/El-espa%C3%B1ol-lengua-viva-2014.pdf> [Accessed: 12 January 2015].
- Fillmore, Charles J. (1982). Frame Semantics. In *Linguistics in the Morning Calm*. The Linguistic Society of Korea (ed.), 111-137. Seoul: Hanshin.
- Fillmore, Charles J. (1985). Frames and the Semantics of Understanding. *Quaderni Di Semantica* 6, 222–254.
- Fillmore, Charles J. & Atkins, Beryl T. (1992). Toward a Frame-based Lexicon: The Semantics of RISK and its Neighbors. In *Frames, Fields and Contrasts: New Essays in Semantic and Lexical Organization*. A. Lehrer, & E. Feder Kittay (eds.), 75-102. Hillsdale: Erlbaum.
- Gallego Hernández, Daniel. (2016). Variation diatopique en phraséologie spécialisée dans le domaine financier. Étude comparative basée sur corpus. In *Computerised and Corpus–based Approaches to Phraseology: Monolingual and Multilingual Perspectives*. G. Corpas Pastor (ed.), 283-292. Geneva: Tradulex.

- Grinev, Sergei & Klepalchenko, Igor. (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.
- 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: 12 January 2015].
- L'Homme, Marie Claude. (1998). Le Statut Du Verbe En Langue De Spécialité Et Sa Description Lexicographique. *Cahiers De Lexicographie* 73 (2), 61–84. <http://www.ling.umontreal.ca/lhomme/docs/cahiers-lexico-98.PDF> [Accessed: 2 February 2012].
- López Rodríguez, Clara Inés, Buendía-Castro, Míriam & García Aragón, Alejandro (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.
- Mogorrón Huerta, Pedro. (2010). Estudio contrastivo lingüístico y semántico de las construcciones verbales fijas diatópicas mexicanas/españolas. *Quaderns de filología. Estudis lingüístics*, 15, 179–198.
- Mogorrón Huerta, Pedro. (2014a). Importancia (numérica) de las variantes diatópicas españolas y su tratamiento en los diccionarios. In *Yearbook of Phraseology*. K. Kuiper (ed.), 123-144. Berlin: De Gruyter Mouton.
- Mogorrón Huerta, Pedro. (2014b). Las expresiones fijas diatópicas argentinas y mexicanas. In *Didáctica y Traducción de las unidades fraseológicas*. Mª. I. González Rey (ed.), 77-98. Santiago de Compostela: Servizo de Publicacións e Intercambio Científico.
- Niles, Ian & Pease, Adam. (2001). Toward a Standard Upper Ontology. In *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems* (FOIS-2001). C. Welty & B. Smith (eds.).
- Reig Alamillo, Asela (2009). Cross-dialectal variation in propositional anaphora: Null objects and propositional lo in Mexican and Peninsular Spanish. *Language Variation and Change* 21, 381-412.
- Ruppenhofer, Josef, Ellsworth, Michael, Petruck, Miriam R. L., Johnson, Cristpher R. & Scheffczyk, Jan (2010). *FrameNet II: Extended Theory and Practice*. ICSI Technical Report. <https://framenet2.icsi.berkeley.edu/docs/r1.5/book.pdf> [Accessed: 12 January 2015].
- Sinclair, John. (2000). Lexical Grammar. *Darbai Ir Dienos*. <http://donelaitis.vdu.lt/publikacijos/sinclair.pdf> [Accessed: 12 January 2015].
- Subirats Rüggeberg, Carlos (2001). *Introducción a la sintaxis léxica del español*. Frankfurt: Vervuert, Iberoaméricana.
- Vaamonde, Gael, González Domínguez, Fita & García-Miguel, José M. (2010). ADESSE. A Database with Syntactic and Semantic Annotation of a Corpus of Span-

ish. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*. N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, & D. Tapias (eds.), 1903-1910. Valletta: European Language Resources Association (ELRA).

7. Dictionaries

- DAMER = *Diccionario de americanismos* (2010). Madrid: Santillana: http://www.asale.org/recursos/diccionarios/damer
- DBM= *Diccionario breve de mexicanismos de Guido Gómez de Silva* (2000). México: Academia Mexicana de la Lengua: http://www.academia.org.mx/universo:lema/obra:Diccionario-breve-de-mexicanismos-de-Guido-Gomez-de-Silva
- DEHISP= *Diccionario de hispanoamericanismos no recogidos por la Real Academia: (formas homónimas, polisémicas y otras derivaciones morfosemánticas),* R. Richard (coord.) (1997). Madrid: Cátedra.
- DEUM = *Diccionario del español usual en México* (1996) dirigido por Luis Fernando Lara. México: El Colegio de México: http://www.cervantesvirtual.com/servlet/SirveObras/35716130101359941976613/
- DM = *Diccionario de Mexicanismos* (2010). México: Academia Mexicana de la Lengua: http://www.academia.org.mx/DiccionarioDeMexicanismos
- DRAE = *Diccionario de la lengua española (23rd edition)* (2014). Madrid: Espasa.

Notes

1. Information extracted from *El español: una lengua viva. Informe 2014* (Fernández Vítores, 2014). Retrieved from: http://eldiae.es/wp-content/uploads/2014/07/El-espa%C3%B1ol-lengua-viva-2014.pdf

2. According to an UNESCO report, there are currently about 420 indigenous languages in Latin America: http://www.unicef.org/lac/lenguas_indigenas.pdf (Mogorrón Huerta, 2014b: 87).

3. The inventory of relations used in EcoLexicon is the following: AFFECTS, ATTRIBUTE_OF, CAUSES, COMPOSED_OF (MATERIAL), OPPOSITE_OF, DELIMITED_BY, STUDIES, PHASE_OF, MEASURES, PART_OF, REPRESENTS, RESULT_OF, EFFECTED_WITH, HAS_FUNCTION, TAKES_PLACE_IN, TYPE_OF, LOCATED_AT.

4. http://www.journals.unam.mx/index.php/atm

5. http://www.journals.unam.mx/index.php/rica

6. Support verbs refer to verbs semantically empty which are neither predicates nor arguments, but predicative marks whose meaning is fundamentally aspectual and dependent on the predicative noun, adjective or phrase they came with (Subirats 2001: 90).