Argentinian Journal of Applied Linguistics
Vol. 7, No. 1, May 2019, 109-130

High-density knowledge rich contexts

Pilar León-Araúz* - University of Granada, Spain
Arianne Reimerink - University of Granada, Spain

(Received 11/01/19; final version received 25/03/19)

ABSTRACT

Knowledge Rich Contexts (KRCs) are one of the usual data categories contained in terminological knowledge bases. In this paper we show how to extract KRCs that combine various items of knowledge to facilitate the work of terminographers as well as user knowledge acquisition. For this reason, a new knowledge pattern-based sketch grammar was designed within the corpus analysis tool Sketch Engine and applied to the EcoLexicon Environmental Corpus. After compiling the corpus with the sketch grammar, KRCs can be extracted through customized semantic word sketches (WS), which provide access to concordances where the item queried is related to others through one or several semantic relations. Then new queries are performed, reusing these new WS in order to collect high-density KRCs, which combine different semantic relations in a single sentence. In this paper we provide a characterization of high-density KRCs based on the amount of knowledge, concept types, conceptual depth, and number and type of conceptual relations codified in them.
Keywords: knowledge rich context; knowledge pattern; sketch grammar; EcoLexicon Environmental Corpus.

RESUMEN

Los Contextos Ricos en Conocimiento (CRC) son un campo habitual en las bases de conocimiento terminológicas. Los CRC que combinan varios elementos de conocimiento facilitan la labor de los terminógrafos y la adquisición del conocimiento de los usuarios. Por este motivo, se diseñó una nueva gramática colocacional *sketch grammar* para la herramienta de análisis de corpus Sketch Engine que, además de basarse en relaciones gramaticales como los *sketch grammar* por defecto, se basa en patrones de conocimiento y se aplicó al corpus medioambiental EcoLexicon English Corpus. Después de compilar el corpus con la gramática, se pueden extraer CRC utilizando los *word sketches* (WS) personalizados que relacionan el elemento de búsqueda con otros a través de una o varias relaciones semánticas. Después se reutilizan estos WS para realizar nuevas búsquedas y extraer CRC de alta densidad, aquellos que combinan varias relaciones semánticas diferentes en una sola oración. En este artículo se presenta una caracterización de los CRC de alta densidad basada en: la cantidad de conocimiento, los tipos de conceptos, la profundidad conceptual y el número y tipo de relaciones codificados en ellos.
Palabras clave: contexto rico en conocimiento; patrón de conocimiento; *sketch grammar*; EcoLexicon Environmental Corpus

* Corresponding author, email: pleon@ugr.es

KNOWLEDGE RICH CONTEXTS (KRCs) (Meyer, 2001) are one of the usual data categories contained in terminological knowledge bases (TKBs), such as EcoLexicon, a multilingual and multimodal TKB on the environment (ecolexicon.ugr.es; Faber et al., 2014; Faber et al., 2016, San Martín et al., 2017). KRCs are conceptually valuable contexts because they contain a term of interest in a particular domain that is semantically related to other terms. KRC extraction is thus essential for terminographic research. One of the most common approaches to find such contexts is to search for the terms in each entry in combination with knowledge patterns (KPs) in corpora. KPs are the linguistic and para-linguistic patterns that convey a specific semantic relation in real texts. For instance, examples of generic-specific KPs are *such as*, *is a kind of*, *and other*, etc.

KPs are considered one of the most reliable methods for the extraction of semantic relations (Barrière, 2004; Bowker, 2003; Condamines, 2002; Marshman, 2002; to cite a few). They have been applied in many terminology-related projects that have led to the development of knowledge extraction tools, such as Caméléon (Aussenac-Gilles and Jacques, 2008) and TerminoWeb (Barrière and Agbago, 2006).

However, no user-friendly application allowing terminologists to find KRCs in their own corpora is publicly available. The corpus query system, Sketch Engine (Kilgarriff et al., 2004), provides the Word Sketch (WS) function that rather than looking at an arbitrary window of text around the headword – as occurs in previous corpus tools –is able to look for each grammatical relation that the word participates in (Kilgarriff et al., 2004). WS are automatically extracted based on a sketch grammar previously defined with which corpora are compiled. These grammars, based on regular expressions and POS tags, identify and annotate in the corpus different structures of interest. For example, the following grammar rule enables the system to show, in the form of WS, statistically significant structures where a noun is modified by other nouns, adjectives and/or adverbs: 2:"(JJ|N).*" [tag="JJ.?"|tag="RB.?"|word=","]{0,3} "N.*"{0,2} 1:"N.*" [tag!="N.*"]. The default word sketches provided by Sketch Engine represent different relations, such as verb-object, modifiers or prepositional phrases. However, they do not represent conceptual relations. For this reason, in León-Araúz et al. (2016), a KP-based sketch grammar for Sketch Engine was developed, thanks to which a list of KRCs can be obtained when querying any item in a corpus through the word sketch functionality.

For contexts to be regarded as KRCs, they should indicate at least one item of domain knowledge that could be useful for conceptual analysis, whether it is an attribute or relation (Meyer, 2001, p. 281). However, obtaining as many items of domain knowledge in as few KRCs as possible would greatly facilitate both the work of terminographers and user knowledge acquisition. For instance, the following fragments (see Table 1) are high-density KRCs because they combine different KPs (in italics) conveying different semantic relations (i.e. hyponymy, meronymy and causality) between many different concepts (in bold):

Nuées ardentes *are but one type of* **pyroclastic flow**, *which include a variety of mixtures of* **volcanic blocks**, **ash**, **gas**, and **lapilli** *that produce* **volcanic rocks** *called* **ignimbrites**. **Contact metamorphism** *of* **carbonate rocks** *produces* **skarn deposits** *containing* **minerals** *such as* **wollastonite**, **tremolite** and **grossular garnet**, **spessartine garnet** and **andradite garnet**.

Table 1. High-density KRCs

The aim of this paper is to show how, thanks to a KP-based sketch grammar, such KRCs can be collected and analyzed. The focus of this paper is therefore on the characterization of high-density KRCs. Measuring the precision and recall of our sketch grammar is out of the scope of this paper.

The remainder of this paper is organized as follows: "Methodology: Improved KRC Extraction" explains our methodology for improved KRC extraction with our customized sketch grammar; in "Characterizing High-Density KRCs" we apply further queries by reusing the annotated semantic WS to collect different types of high-density KRC and we provide a detailed description of the characteristics of the KRCs collected; finally, some conclusions are drawn and ideas for future work are described in "Conclusions and Future Work".

Methodology: Improved KRC Extraction

The methodology for improved KRC extraction is based on the EcoLexicon Semantic Sketch Grammar (ESSG) (León-Araúz and San Martín 2018; León-Araúz et al., 2016) developed within Sketch Engine and applied to the EcoLexicon English Corpus (EEC) (León-Araúz et al., 2018)[1].

The EEC is a 23.1-million-word corpus of contemporary environmental texts. Each text in the EEC is tagged with a set of XML-based metadata, some of which are based on the Dublin Core Schema[2], while others have been included to meet our own needs. Corpus metadata permit users to constrain corpus queries based on pragmatic factors, such as environmental domains and target reader. Thus, for instance, the use of the same term in different contexts can be compared. Tags are based on the following main parameters:

- Domain: the EEC encompasses all the domains and subdomains of environmental studies (e.g., Biology, Meteorology, Ecology, Environmental Engineering, Environmental Law, etc.).
- User: the corpus includes texts for three types of user, depending on level of expertise (i.e. expert, semi-expert, general public), based on parameters such as the sender (expert in the field, journalist, governmental institution, etc.), type of document (specialized paper, manual, brochure, etc.), place of publication (specialized journal, general website, etc.) and specific mention of intended audience.

- Geographical variant: it comprises American, British, and Euro English (official documents from the European Union).
- Genre: it covers a wide variety of text genres (e.g. journal articles, books, websites, lexicographical material, etc.).
- Editor: it distinguishes texts edited by scholars/researchers, businesses, government bodies, etc.
- Year: it includes texts from 1973 to 2016.

The EEC was compiled with the Penn Treebank tagset (TreeTagger version 3.3) and with the ESSG, a CQL-based (Corpus Query Language) customized sketch grammar also containing the default sketch grammar. So far, the ESSG contains 64 sketch grammar rules[3], focused on the extraction of conceptual relations, which expands the functionality of word sketches to summarized representations of semantic behavior, namely, KRCs. This new sketch grammar for the English language includes some of the most common conceptual relations used in Terminology: generic-specific, part-whole, location, cause, and function.

In the development of the ESSG, issues specific to each relation and pattern had to be taken into account. For instance, a single sentence can produce more than one term pair because of the enumerations that are often found on each side of the pattern (e.g. *x, y, z and other types of w*). This entails performing greedy queries in order to allow any of the enumerated elements fill the target term. However, this may also cause endless noisy loops. Sometimes it is necessary to limit the number of possible words on each side of the pattern. For example, enumerations are more often found on the side of hyponyms, parts, and effects than on the side of hypernyms, wholes, and causes. Consequently, the loops were constrained accordingly in the latter case. Table 2 shows a summarized and simplified version of the patterns included for each semantic relation in the ESSG.

<u>Generic-specific (18 sketch grammars):</u> HYPONYM ,|(|:|is|belongs (to) (a|the|…) type|category|… of HYPERNYM // types|kinds|… of HYPERNYM include|are HYPONYM // types|kinds|… of HYPERNYM range from (…) (to) HYPONYM // HYPERNYM (type|category|…) (,|() ranging (…) (to) HYPONYM // HYPERNYM types|categories|… include HYPONYM // HYPERNYM such as HYPONYM // HYPERNYM including HYPONYM // HYPERNYM ,|(especially|primarily|… HYPONYM // HYPONYM and|or other (types|kinds|…) of HYPERNYM // HYPONYM is defined|classified|… as (a|the|…) (type|kind|…) (of) HYPERNYM // classify|categorize|… (this type|kind|… of) HYPONYM as HYPERNYM // HYPERNYM is classified|categorized in|into (a|the|…) (type|kind|…) (of) HYPONYM // HYPERNYM (,|() (is) divided in|into (…) types|kinds|… :|of HYPONYM // type|kind|… of HYPERNYM (is|,|() known|referred|… (to) (as) HYPONYM // HYPONYM is a HYPERNYM that|which|… // define HYPONYM as (a|the|…) (type|category|…) (of) HYPERNYM // HYPONYM refers to (a|the|…) (type|category|…) (of) HYPERNYM // (a|the|one|two…) (type|category|…) (of) HYPERNYM: HYPONYM

<table>
<tr><td><u>Part-whole (17 sketch grammars):</u> WHOLE is comprised|composed|constituted (in part) of|by PART // WHOLE comprises PART // PART composes WHOLE // PART is|constitutes (a|the|…) part|component|… of WHOLE // WHOLE has|includes|possesses (…) part|component|… (,|() (:|such as|usually|namely|…) PART // WHOLE has|includes|possesses (a|the|…) fraction|amount|percent… of PART // WHOLE part|component|… (,|() such as PART // part|component|… of WHOLE (,|() (:|such as|usually|namely|…) PART // (a|the|one|two| some|…) part|component|… of WHOLE is PART // (a|the|one|two|some|…) part|component|… of WHOLE (is) called|referred|… (to) (as) PART // PART (,|() (a|the|…) part|component|… of WHOLE // WHOLE is divided in| into (two|some|…) parts|components|… (,|() (:|such as|usually|namely|…) PART // WHOLE is divided in|into PART // WHOLE (is|,|() made|built|… (up) of|from|with PART // WHOLE contains PART // PART (is) contained in WHOLE // WHOLE consists of PART</td></tr>
<tr><td><u>Cause (10 sketch grammars):</u> CAUSE (is) responsible for EFFECT // CAUSE causes|produces|… EFFECT // CAUSE leads|contributes|gives (rise) to EFFECT // CAUSE-driven|-induced|-caused EFFECT // EFFECT (is) caused|produced|… by|because|due (of|to) CAUSE // EFFECT derives|results from CAUSE // cause of EFFECT is CAUSE // CAUSE (is) (a|the|…) cause of EFFECT // CAUSE (,|() (a|the|…) cause of EFFECT // EFFECT is|,|(forms|formed by|from CAUSE</td></tr>
<tr><td><u>Location (4 sketch grammars):</u> ENTITY (is) connected|delimited to|by PLACE // ENTITY (is) found|built|… in|on|… PLACE // ENTITY (is) formed|forms in|on|… PLACE // ENTITY (is) extended|extends (out) into| parallel|… (of|to) PLACE</td></tr>
<tr><td><u>Function (7 sketch grammars)</u>: ENTITY (has|provides|…) (a|the|…) function|role|purpose of FUNCTION // ENTITY is (built|designed|…) for|to FUNCTION // ENTITY is (useful|effective|…) for|to FUNCTION // ENTITY is (a|the|…) (…) built|designed|… for|to FUNCTION // ENTITY is (a|the|…) (…) used|employed|… for|as FUNCTION // use|employ|… ENTITY for|as|to FUNCTION // function|role|purpose of ENTITY is FUNCTION</td></tr>
</table>

Table 2. Simplified version of the patterns included in each grammar

Each grammar rule is the formalization of KPs in the form of regular expressions combined with POS tags. As an example, Table 3 shows the actual CQL representation of one of the generic-specific KP-based rules, followed by an explanation and three natural language examples of concordances matched with the grammar.

<table>
<tr><td colspan="2">1:"N.*" [word=",|\("]? [tag="IN/that|WDT"]? "MD"* [lemma="be|,|\("] "RB.*"* [word="classified| categori.ed"] ([word="by"] [tag!="V.*"]+)? [word="in|into"] [tag!="V.*"]* [lemma="type|kind| example| group|class| sort|category|family|species|subtype| subfamily|subgroup| subclass|subcategory|subspecies"]? [tag!="V.*"]* 2:[tag="N.*" & lemma!="type|kind|example| group|class| sort|category|family|species| subtype|subfamily|subgroup| subclass|subcategory|subspecies"]</td></tr>
<tr><td>1:"N.*"</td><td>The hypernym is a noun.</td></tr>
<tr><td>[word=",|\("]?</td><td>An optional comma or bracket.</td></tr>
<tr><td>[tag="IN/that|WDT"]?</td><td>Optionally “that” or “which”.</td></tr>
<tr><td>"MD"*</td><td>Any modal verb from zero to infinite times.</td></tr>
<tr><td>[lemma="be|,|\("]</td><td>Lemma “be” or a comma or a bracket.</td></tr>
<tr><td>"RB.*"*</td><td>Any adverb from zero to infinite times.</td></tr>
<tr><td>[word="classified|categori.ed"]</td><td>Classified, categorised, or categorized.</td></tr>
<tr><td>([word="by"] [tag!="V.*"]+)?</td><td>Optionally, “by” followed by anything from one to infinite times that does not contain a verb.</td></tr>
<tr><td>[word="in|into"]</td><td>In or into.</td></tr>
<tr><td>[tag!="V.*"]*</td><td>Anything from zero to infinite times that does not contain a</td></tr>
</table>

	verb.
[lemma="type|kind| example|group|class|sort|category|family|species|subtype|subfamily|subgroup|subclass|subcategory|subspecies"]?	Optionally any of the lemmas "type", "kind", "example", "group", "class", "sort", "family", etc.
[tag!="V.*"]*	Anything from zero to infinite times that does not contain a verb.
2:[tag="N.*" & lemma!="type|kind|example| group|class|sort|category|family|species|subtype|subfamily| subgroup|subclass|subcategory|subspecies"]	The hyponym is any noun other than "type", "kind", "example", "group", "class", "sort", "family", etc.
Stony-iron meteorites are classified into pallasites and mesosiderites. Modern reefs are classified into several geomorphic types: atoll, barrier, fringing, and patch. Littoral materials are classified by grain size in clay, silt, sand, gravel, cobble, and boulder.	

Table 3. CQL representation of a generic-specific KP-based rule with its explanation

Once all the rules in the ESSG are applied, different semantic word sketches can be automatically derived, as shown in Figure 1 for the term *mineral* (only hyponymic and meronymic sketches are displayed).

mineral *(noun)* Alternative PoS: adjective (freq: 127)
EcoLexicon English (Environment) freq = 5,252 (183.53 per million)

"mineral" is the generic of...		21.42
quartz	38	9.96
gold	26	9.40
mica	24	9.35
feldspar	24	9.35
carbonate	22	9.12
iron	24	9.06
calcite	19	9.04
mineral	23	8.98
copper	17	8.74
clay	17	8.62
salt	16	8.54
amphibole	13	8.46
sulfide	13	8.41
calcium	13	8.35
pyroxene	11	8.28
sulfur	12	8.19
zinc	11	8.19
uranium	11	8.17
magnetite	10	8.14
coal	13	8.11
oxide	11	7.91
garnet	8	7.83
silver	8	7.78
ore	8	7.77
amount	9	7.72

"mineral" is part of...		10.36
rock	79	10.65
soil	15	8.71
magma	7	8.63
mineral	9	8.54
melt	6	8.46
jade	6	8.46
peridotite	5	8.18
silt	5	8.13
crust	6	8.08
meteorite	6	8.07
limestone	5	7.97
planet	5	7.97
type	7	7.91
earth	7	7.83
deposit	5	7.70
material	6	7.60
sand	5	7.56
sediment	6	7.48
beach	5	7.09
water	7	7.07
group	5	7.02

"mineral" is a type of...		6.80
mineral	23	8.98
rock	12	8.03
resource	10	7.99
substance	7	7.85
earth	7	7.63
material	14	7.51
compound	7	7.47
soil	6	7.20
feature	8	6.93
chemical	5	6.83
factor	6	5.75

"mineral" has part...		4.70
silicon	11	10.14
co3	7	9.68
carbonate	11	9.61
oxygen	14	9.61
calcium	8	9.27
anion	5	9.25
magnesium	6	9.23
iron	8	9.17
aluminum	5	9.08
silica	5	8.94
ion	6	8.93
mineral	9	8.54

Figure 1. Semantic word sketches of *mineral*

When clicking on the number next to each item, concordances unfold as depicted in Figure 2, where *quartz* is shown as a type of *mineral* through different KPs (*such as*, *is a*, *and/or other*, *typically*, *including*).

in a regular three-dimensional pattern to form minerals such as *quartz* (silicon dioxide) or calcite (
' planes of weakness. *Quartz* is a common mineral that lacks cleavage , however. Hardness is
component too , such as *quartz* , pyrite and other minerals , and evaporite minerals , especially
17.18. Continental crust is composed mostly of minerals such as feldspar and *quartz* , which form
reportedly consists primarily of brittle minerals such as calcite , dolomite , albite feldspar ,
being replaced by calcite , *quartz* , or other minerals during burial and diagenesis. Fossils are
conditions fibrous minerals or other minerals , typically *quartz* or calcite , may fill the
depending on the strain rate. Most other common minerals , including *quartz* and calcite , stop behaving

Figure 2. Concordances of QUARTZ *type_of* MINERAL

In Sketch Engine, these concordances can also be shown in the form of whole sentences, thus complying with the idea of KRCs (Figure 3).

Most of the earth 's materials have their molecules arranged in a regular three-dimensional pattern to form minerals such as *quartz* (silicon dioxide) or calcite (calcium carbonate).

Quartz is a common mineral that lacks cleavage , however.

There can be a noncarbonate component too , such as *quartz* , pyrite and other minerals , and evaporite minerals , especially gypsum-anhydrite , can be found in association with limestones.

Continental crust is composed mostly of minerals such as feldspar and *quartz* , which form less-dense , lighter-colored granitic rocks.

The shale matrix reportedly consists primarily of brittle minerals such as calcite , dolomite , albite feldspar , ankerite , *quartz* as well as significant rutile and pyrite.

To be preserved life-forms become mineralized after they die , with organic tissues typically being replaced by calcite , *quartz* , or other minerals during burial and diagenesis.

Under some deformation conditions fibrous minerals or other minerals , typically *quartz* or calcite , may fill the spaces between the boudins.

Most other common minerals , including *quartz* and calcite , stop behaving brittlely and deform ductily at 570oF-1,470oF (300oCâ-800oC).

Figure 3. KRCs for QUARTZ *type_of* MINERAL

As previously mentioned, these can be regarded as KRCs because they indicate, at least, one item of knowledge. However, some KRCs are more useful than others. For instance, the second example in Figure 5 only contains one conceptual proposition[4] (QUARTZ-*type_of*-MINERAL), whereas the fifth activates eight (CALCITE-*type_of*-MINERAL; DOLOMITE-*type_of*-MINERAL; ALBITE-*type_of*-MINERAL; FELDSPAR-*type_of*-MINERAL; ANKERITE-*type_of*-MINERAL; QUARTZ-*type_of*-MINERAL; RUTILE-*type_of*-MINERAL; PYRITE-*type_of*-MINERAL). In these two examples, *type_of* is the only conceptual relation that can be extracted. However, the fourth KRC conveys two: *made_of* (CONTINENTAL CRUST-*made_of*-MINERAL; GRANITIC ROCK-*made_of*-MINERAL) and *type_of* (FELDSPAR-*type_of*-MINERAL; QUARTZ-*type_of*-MINERAL). Therefore, we can assume that, among KRCs, there are poorer KRCs and high-density KRCs. High-density KRCs can be defined as those that contain two or more conceptual propositions.

A high-density KRC can be useful both for users when querying a TKB and for terminographers when building a TKB. Users may activate a whole conceptual network in their brains if in a single sentence they are confronted with multiple concepts and relations. In the same way, terminographers will populate their TKB in a much more efficient way if a single sentence provides them with multiple conceptual propositions. Moreover, another advantage of high-density KRCs is their reusability, since they can be used in as many term entries as conceptually related terms are found in the KRC.

Characterizing High-Density KRCs

KRCs have been previously characterized in a study conducted by Condamines et al. (2013), where the authors differentiate between conceptually rich and linguistically rich contexts based on their usefulness for translators. In this paper, the focus is on further characterizing high-density KRCs, which can be regarded as a kind of conceptually rich contexts. Three different sets of this kind of KRCs were collected through different queries reusing the semantic WS annotated through the ESSG: (1) KRCs where three different conceptual relations could be found; (2) KRCs codifying two conceptual relation types but several conceptual propositions; and (3) KRCs codifying only one conceptual relation type but several conceptual propositions.

For extracting them, the following query (Figure 4) was first applied to the corpus after excluding the documents of the corpus annotated with Genre "Lexicographic material". This way we avoided confusing definitions from dictionaries and thesauri with actual KRCs in text.

```
([ws(".*-n","\"%w\" is a type of...",".*-n")] []* [ws(".*-n","\"%w\" has part...",".*-n")] []*[ws(".*-n","\"%w\" is the cause of...",".*-n")] | [ws(".*-n","\"%w\" is a type of...",".*-n")] []*[ws(".*-n","\"%w\" is the cause of...",".*-n")] []*[ws(".*-n","\"%w\" has part...",".*-n")] | [ws(".*-n","\"%w\" has part...",".*-n")] []*[ws(".*-n","\"%w\" is a type of...",".*-n")] []*[ws(".*-n","\"%w\" is the cause of...",".*-n")] | [ws(".*-n","\"%w\" has part...",".*-n")] []*[ws(".*-n","\"%w\" is the cause of...",".*-n")] []*[ws(".*-n","\"%w\" is a type of...",".*-n")] |[ws(".*-n","\"%w\" is the cause of...",".*-n")] []*[ws(".*-n","\"%w\" is a type of...",".*-n")] []*[ws(".*-n","\"%w\" has part...",".*-n")] | [ws(".*-n","\"%w\" is the cause of...",".*-n")] []*[ws(".*-n","\"%w\" has part...",".*-n")] []*[ws(".*-n","\"%w\" is a type of...",".*-n")])
within <s/>
```

Figure 4. CQL query for the extraction of high-density KRCs

This query reuses the annotations of three types of semantic word sketches (e.g. [ws(".*-n","\"%w\" is a type of...",".*-n")] means any item annotated as a hyponym) and combines them with any other item or items ([]*) that may appear within the same sentence (within <s/>). In this way, we first collected a sample of high-density KRCs that activate at least three different relations: hyponymy, meronymy and causality. The selection of these relations is based on the grammar rules that are more extensively developed so far in the ESSG.

High-density KRCs in Terminology can be assimilated to what "good dictionary examples" have traditionally been in Lexicography. According to Kilgarriff et al. (2008), a good example must be: (1) typical, exhibiting frequent and well-dispersed patterns of usage; (2) informative, helping to elucidate the definition; and (3) intelligible to learners, avoiding high sentence length, difficult lexis and structures, puzzling or distracting names, anaphoric references, etc.

These parameters were outlined with a collocation dictionary in mind. In our case, typicality is based on the kind of concepts related in the KRCs, depending on whether they refer to typical environmental entities and/or processes or not; informativity depends on the number of conceptual propositions and different relations conveyed in the same KRC; and intelligibility is understood in the same way, although the complexity that can be expected in a domain-specific resource is inevitably higher than in a learner environment. In fact, after performing the query, we manually filtered out all KRCs that: (1) were too long for the scarce amount of knowledge obtained proportionally (see for example Table 4); (2) showed anaphora, as in Table 5, where the particle "such" indicates that not all soils contain calcium ions; (3) contained too many named entities, thus not relating proper concepts (Table 6); or (4) were not actual KRCs because KPs were noisy, as in Table 7, where *part of* does not convey actual meronymy ("another part of the problem"); or where *consist of* is not used as a meronymic KP but rather as a hyponymic one.

They may also authorise, specifying the conditions for: -injections of **water** *containing* **substances** *resulting from* the operations for **exploration and extraction of hydrocarbons or mining activities**, and injection of water for technical reasons, into geological formations from which **hydrocarbons** *or other* **substances** have been extracted or into geological formations which for natural reasons are permanently unsuitable for other purposes.

Table 4. Filtered KRC: high length and low content

Such **soils** are usually *rich in* **calcium ions**, and are often *derived from* **rocks** *such as* **limestone** or **chalk**, which are mostly *composed of* **calcium carbonate**.

Table 5. KRC with anaphora (*such soils*)

Flooding *caused by* **Katrina** was particularly severe *in* the **polder** that *comprises part of* **St. Bernard and Orleans Parishes in New Orleans**, Louisiana, *referred to here as* the **St. Bernard Polder**.

Table 6. KRC containing too many named entities

Another *part of* the **problem** was the **difficulty** in separating anthropogenic changes from natural fluctuations in fish availability, and in **phenomena** *such as* **coastal erosion** and **flooding** which could be *caused by* **human interference or natural processes**. **Managed retreat schemes** will *result in* **loss of land** which in many cases may *consist of* **grazing** or other **agricultural land**.

Table 7. Noisy KPs

Finally, approximately 40 high-density KRCs were collected (examples are shown in Tables 8-15). The number of these KRCs involving hyponymy, meronymy and causality in the EEC is relatively low, considering the size of the corpus. This may result from the fact that the ESSG still needs to be refined in order to accommodate new KPs, or the fact that the combination of hyponymy, meronymy, and causality within the same sentence may have a constraining effect on the concepts related and not be as frequent as initially thought. This would actually explain the concept types involved in the vast majority of the KRCs analyzed, which were chemical substances and matter in the case of entities, and formation and transformation processes in the case of events.

The low frequency may also convey that this combination of conceptual relations, found simultaneously in the same sentence, is atypical in the domain. These would therefore be less useful as user-aimed examples in a TKB. They would, however, be useful for a terminographer who wants to build the underlying conceptual system.

Other features that characterize high-density KRCs are the number of conceptual propositions involved and the depth of the hierarchy. In the KRCs analyzed, the number of propositions goes from 3 (as required by the starting query) to 11; and hierarchy depth goes from 1 to a maximum of 3 hierarchical nodes.

For the sake of clarity in our description and characterization of high-density KRCs, we have included numerous examples extracted from the corpus. Tables 8-19, 23, and 26-31 show how KRCs are dissected according to conceptual propositions and hierarchical levels. This qualitative analysis was carried out manually. For instance, the KRC in Table 8 activates 3 propositions and only one hierarchical level, since all other concepts are related to the same starting concept (*fire plume*).

Fire plumes *and other* **biomass fires** *contain* **CO** and *can contribute to* **O3 formation**.		
fire plume	*type_of*	biomass fire
	made_of	CO
	cause	O3 formation

Table 8. High-density KRC: 3 propositions and 1 hierarchical level

In contrast, the KRC in Table 9 shows up to 10 propositions and 3 hierarchical levels. *Photochemical smog* is surrounded by its different components, and all these materials are also related to what they are, what they are made of and what they cause.

Photochemical smog *contains* **compounds** *such as* **aldehydes** (*compounds containing* the **-COOH group** *joined directly to* **another carbon atom**), **ketones** (*compounds containing* the **CO.CO.C group**), and **formaldehyde** (or methanol, HCHO), which impart a characteristic odor, and **nitrogen dioxide** (NO2) and **solid particles** that *cause* a **brownish haze**.

photochemical smog	*made_of*	compound	*generic_of*	aldehyde	*made_of*	-COOH group
						carbon atom
				ketone		C.CO.C group
				formaldehyde		
				nitrogen dioxide		
				solid particles	*cause*	brownish haze

Table 9. High-density KRC: 10 propositions and 3 hierarchical levels

This does not necessarily mean that the number of propositions correlates with the number of hierarchical levels, since there are KRCs where only 3 propositions may activate 3 hierarchical levels, as shown in Table 10. This builds a conceptual chain starting from a single concept.

Phytoplankton *comprises* at least four thousand **species of plants** that, as on land, use sunlight in the process of photosynthesis to *generate* **sugars** *and other* **high-energy organic compounds**.

phytoplankton	*made_of*	plant	*causes*	sugar	*type_of*	high-energy organic compound

Table 10. High-density KRC: 3 propositions and 3 hierarchical levels

In some of the KRCs analyzed, we found other relations not explicitly searched for, such as *affects* (Table 11) – not as yet formalized in the ESSG but included as a conceptual relation in the environmental TKB EcoLexicon – or *has_location* (Table 12 and 13) and *has_function* – both formalized within the ESSG. This means that in a single sentence we can find even more than 3 different relation types.

In addition, **non-CO2 climate forcers** *(defined as* any **gaseous or particulate compound** that *contributes to* **climate change** *including* **O3**, **CH4**, **nitrous oxide**, **F-gases** *(***gases** *containing* **fluorine**) *as well as* **PM**) *exert influence on* the **Earth's energy balance** and on **climate**.

non-CO2 climate force	*type_of*	gaseous compound		
	causes	climate change		
	generic_of	O3		
		CH4		
		nitrous oxide		
		PM		
		F-gas	*made_of*	fluorine
	affects	Earth's energy balance		
		climate		

Table 11. High-density KRC: other relations (*affects*)

PM is emitted from many sources, and *is a* complex heterogeneous *mixture comprising* both **primary and secondary PM**; primary PM *is the fraction of* PM that is emitted directly into the atmosphere, whereas **secondary PM** *forms in* the **atmosphere** *following* the **oxidation and transformation of precursor gases** *(mainly* **SOX**, **NOX**, **NH3** and some **volatile organic compounds** (VOCs)).

PM	*made_of*	primary PM		
		secondary PM	*has_location*	atmosphere
			caused_by	oxidation and transformation of precursor gases
precursor gas	*generic_of*	SOX		
		NOX		
		NH3		
		VOCs		

Table 12. High-density KRC: other relations (*has_location*)

Ozone *is a* **secondary pollutant** *formed in* the **troposphere**, the lower *part of* the **atmosphere**, *from* complex **chemical reactions** *following* emissions of **precursor gases** *such as* **NOX** and **non-methane VOC** (NMVOC).

ozone	*type_of*	secondary pollutant				
	has_location	troposphere	*part_of*	atmosphere		
	caused_by	chemical reactions	*caused_by*	precursor gas	*generic_of*	NOX
						non-methane VOC

Table 13. High-density KRC: other relations (*has_location*)

The *affects* relation in Table 11 is conveyed through the KP *exert influence on*, which is a KP that has not yet been formalized in the ESSG. This is an example of how the analysis of KRCs can help to extract new KPs to improve the ESSG through the refinement of grammar rules or the formalization of new relations. In the case of the *has_location* relation in Tables

12 and 13, it is conveyed through the KP *form(ed) in*, which is already stored in the ESSG; but in other KRCs, new location-related KPs have been found that might be integrated into the grammar, as is the case with the example in Table 14 (*are prevalent in*).

Still another *constituent of* the **clay traction** *is a group of* **hydrous oxides** *of* **iron** and **aluminum** *known as* **sesquioxides**, which *are prevalent mainly in* the **soils of tropical and subtropical regions** and *are responsible for* the predominantly **reddish or yellowish hue** of these soils.				
clay traction	*made_of*	sesquioxide	*type_of*	hydrous oxide
			made_of	iron
				aluminum
			has_location	tropical region
				subtropical region
			causes	reddish or yellowish hue

Table 14. High-density KRC: other relations (*has_location*)

Nevertheless, we must be cautious when integrating new KPs in the ESSG, since they can be more noisy than productive. In this sense, *following* (Tables 12 and 13) can be polysemic, since it can point to a causal relation, as in this case, or a time-related one.

In the same way, prepositions such as *in*, *from* or *for* are very ambiguous. *In* may express location, time, manner, etc.; *from* may convey function, direction, duration, etc.; and *of* may mean part, time, possession, etc. For this reason, they are never included as part of grammar rules unless they are accompanied by a disambiguating element (as in *form(ed) in*), even though prepositions are some of the most important indicators of event-knowledge (Barrière, 2004). However, when encountered within high-density KRCs, their ambiguity can be reduced based on the type of relations and concepts activated around them.

Besides prepositions, there are also equally noisy verbs, such as *have*, which can express meronymy (as in Table 15) but should only be employed as a KP in very controlled scenarios, where its ambiguity can be reduced based on the relations or concepts types activated nearby.

Venus *has* **clouds** *containing* **sulfuric acid** and an **atmosphere** *of* **carbon dioxide** that *produces* a strong **greenhouse effect**, the **Earth** *is* the only **planet** that *has* all three forms of **water** on its surface.						
Venus	*made_of*	cloud	*made_of*	sulfuric acid		
		atmosphere		carbon dioxide	*causes*	greenhouse effect
Earth	*type_of*	planet				
	made_of	water				

Table 15. High-density KRC: polysemic verb

Since the number of high-density KRCs collected at this stage was relatively low probably due to the number of relations searched for, the next logical step was to extract high-density KRCs where only two relation types were found. These KRCs are much more numerous in the corpus and their combinatorial patterns provide further insights into how these KRCs behave.

The results obtained amounted to: 1,392 KRCs showing the combination of *type_of* and *part_of* relations; 3,057 for *type_of* and *caused_by* relations; and 759 for *part_of* and *caused_by* relations. Again, some manual filtering had to be done before analyzing the KRCs, since the same problems reported for the KRCs above were found (e.g. noise, anaphora, KP polysemy, etc.). Then, a random selection was again manually analyzed from a qualitative perspective, since the analysis of the whole set goes beyond the scope and extension of this paper.

Regarding concept types, again chemical substances and matter are the concepts that appear to be most often simultaneously linked through hyponymy and meronymy, although organisms and landforms are also prevalent. In some of the *type_of* +*part_of* KRCs other relations can be found, such as *affects, has_location, delimited_by* or *has_function*, but in most of them only *type_of* and *part_of* appear. In this case, KRCs tend to show deeper hierarchical levels, as shown in Table 16, where a fourth hierarchical level appears.

The typical **aluminosilicate clay minerals** appear as laminated microcrystals, *composed mainly of* two basic **structural units**: a **tetrahedron** *of* four **oxygen atoms** surrounding a central **cation**, *usually* **Si^{4+},** and an **octahedron** *of* six **oxygen atoms** surrounding a somewhat larger **cation** of lesser valency, *usually* **Al^{3+}** or **Mg^{2+}.**

Al. clay mineral	*has_part*	structural unit	*has_type*	tetrahedron	*made_of*	oxygen atom		Si^{4+}
				ocahedron		cation	*type_of*	Al^{2+}
								Mg^{2+}

Table 16. High-density KRC: 8 propositions and 4 hierarchical levels (type_of+part_of)

Type+cause KRCs are the most numerous of the two-relation combinations, prototypically activating both processes and entities related to organisms, landforms, coastal structures, substances, wave action, meteorological phenomena, and chemical reactions and diseases, which is only natural, since causality involves the description of processes and their participants. For example, Tables 17-19 show three KRCs where only *type_of* and *caused_by* are activated.

Non-electrical energy sources, *such as* **natural gas furnaces**, also *produce* **GHGs**.

non-electrical energy source	*has_type*	natural gas furnace	*causes*	GHG

Table 17. High-density KRC: 2 propositions and 2 hierarchical levels (type_of+cause)

Water *(mainly* **acidic water**) and **gases** in air or dissolved in water *(mainly* **oxygen** and **carbon dioxide**) *can cause* **chemical weathering**.				
chemical weathering	*caused_by*	acidic water	*type_of*	water
		oxygen		gas
		carbon dioxide		

Table 18. High-density KRC: 5 propositions and 2 hierarchical levels (type_of+cause)

Climatic hazards *such as* **strong winds** and **heavy rains**, **storms** and **hurricanes** frequently *give rise to* **landslides**, **sediment flows**, and **water floods**.				
strong wind	*type_of*	climatic hazard	*causes*	landslide
heavy rain				sediment flow
storm				water flood
hurricane				

Table 19. High-density KRC: 7 propositions and 2 hierarchical levels (type_of+cause)

In contrast to *type_of+part_of* combinations, these KRCs seem to reflect less deep hierarchical levels (i.e. only two in the previous examples), no matter the number of conceptual propositions reflected (e.g. from 2 to 7 in the previous examples), at least when only *type_of* and *caused_by* are activated. However, causality and hyponymy are found very often in combination with other relations not yet developed in the ESSG, such as *affects*. Only in these cases deeper hierarchical levels emerge. Furthermore, most KRCs of this kind are not as straightforward as the ones shown in Tables 17-19. The expression of causality often includes entire clauses that are difficult to merge in a single concept, as in Tables 20 and 21, where "variations in the force of gravity", "relative movements of heavenly bodies", "the moon revolving around the Earth", "movements of marine boundaries", etc. would hinder the construction of concept systems. Therefore, despite the fact that these contexts contain unambiguous KPs pointing to useful semantic relations, their use as KRCs may be restricted to terminology users rather than terminographers when building a TKB.

The **astronomical tide** *refers to* the **regular oscillations of the sea or ocean surface**, *due to* **variations in the force of gravity** *caused by* the **relative movements of heavenly bodies**, *mainly* **those of the moon revolving around the Earth, and those of the Earth, revolving around the Sun**.

Table 20. High-density KRC: clauses

Tidal waves or **tsunamis** *are* **long waves**, *generated by* **movements of marine boundaries**, *such as* the **sea bottom** *because of* a **submarine earthquake**, or of a **sea slope** *because of* a **sediment slump** or **landslide** above or underneath the water.

Table 21. Table 20. High-density KRC: clauses

As for *part_of+caused_by* KRCs (Tables 22 and 23), the concepts involved are related to the parts and materials of the same concept types related in *type_of+caused_by* KRCs in addition to other categories of concepts such as tools and instruments or physical and mathematical concepts. Again in the case of this type of KRCs, the hierarchical depth seems to be lower except when other relations apply (Table 23; *affects*).

Ground-level O3 and **black carbon**, *a constituent of* **PM**, *contribute to* **global warming**.				
particulate matter	*has_part*	black carbon ground-level O3	*causes*	global warming

Table 22. High-density KRC: 4 propositions and 2 hierarchical levels (part_of+cause)

In temperate climates, **beaches** typically *consist of* **quartz** and **feldspar grains** *derived from* the **weathering** *of* **terrestrial rocks**.						
beach	*has_part*	quartz grain feldspar grain	*caused by*	weathering	*affects*	terrestrial rock

Table 23. High-density KRC: 5 propositions and 3 hierarchical levels (part_of+cause)

In this case it is also very common to find causal clauses, as shown in Table 24 (e.g. "exposure of GAC surface to such effluents").

Since the **effluent** of primary treatment *contains* a relatively high amount of **biodegradable substances**, the **exposure of GAC surface to such effluents** *leads to* rapid **colonization by microorganisms**.

Table 24. High-density KRC: clauses

Finally, since causality seemed to be widely represented in the corpus (even though causal grammars are not as developed as others, such as hyponymy), a different query was performed in order to extract high-density KRCs where different propositions of the same kind (i.e. causality) could be extracted. In this case the analysis was more focused on the type, amount and diversity of causal KPs involved in these KRCs (i.e. *cause*, *caused by*, *produce*, *generated by*, *give rise to*, *derived from*, *trigger*, etc.). The initial set of the extracted contexts amounted to more than 7,000. After a manual filtering and a random selection, a set of 260 was retained and classified according to the KPs contained. Most of them only contained two KPs, but 38 of them contained 3 or more KPs. The maximum number of KPs encountered in the same KRC was 7, as shown in the KRC in Table 25. However, again in this case not all propositions could be used in a concept system due to the nodes expressed in the form of clauses instead of self-contained concepts.

> There are four primary meteorologic causes that should be recognized and the data segregated accordingly: (1) **thunderstorm type events** where the *resulting* **flood** is *caused by* high intensity, short duration, **rainfall** that *produces* **high peak discharges** and relatively low volumes (2) **general rain type events** where the *resulting* **flood** *is caused by* moderate intensity, long duration, **rainfall** (3) **snowmelt floods** *resulting from* the **melting of an accumulated snow pack** and (4) **floods** *resulting from* a combination of **rain falling on a melting snowpack**.

Table 25. High-density KRC: causality through 7 KPs

Some of these KRCs include several conceptual propositions through the same KP, such as the one shown in Tables 26 (i.e. *produce*), 27 (i.e. *cause*), 28 (i.e. *produce*) and 29 (i.e. *induced by*). It usually occurs in enumerations (i.e. Tables 26 and 27), when the causes and/or effects belong to the same conceptual category, especially co-hyponyms (i.e. Table 28), or when the same cause or effect participates in several propositions, usually acting as the subject or object of the sentence (i.e. Table 29).

> Typical process-form inferences include: **creep processes**, which *produce* an expanding **upper convexity** on a slope; **erosion by overland flow**, especially by gullying, which *produces* an increasing **lower concavity**; **uniform solution**, which *produces* a parallel **downwearing**; and shallow **landslides**, which *produce* parallel **slope retreat**, associated with a lower concavity in which the slide debris accumulates.

creep process		upper convexity
erosion by overland flow	*causes*	lower concavity
uniform solution		downwearing
landslide		slope retreat

Table 26. High-density KRC: 4 propositions and 1 hierarchical level (cause+cause)

> However , **sulfates** *cause* other **environmental problems** *such as* **acid rain**, **carbon black** *causes* **human health problems**, and **dimming** *causes* **ecological problems** *such as* **changes in evaporation and rainfall patterns**, with **droughts** and **increased rainfall** both *causing* **problems for agriculture**.

environmental problem	*has_type*	acid rain	*caused_by*	sulfate
carbon black	*causes*	human health problem		
dimming	*affects*	evaporation	*affects*	Agriculture
		rainfall		
drought				

Table 27. High-density KRC: 8 propositions and 2 hierarchical levels (cause+cause)

> **Steep pressure gradients** (tightly packed isobars) *produce* **strong pressure gradient forces** and **high winds**; **gentle pressure gradients** (widely spaced isobars) *produce* **weak pressure gradient forces** and **light winds**.

steep pressure gradient		strong pressure gradient force
	causes	high wind
gentle pressure gradient		weak pressure gradient force
		light wind

Table 28. High-density KRC: 4 propositions and 1 hierarchical level (cause+cause)

Water level changes can be locally *induced by* **winds** blowing across the bay or they can be *induced by* **flow of water** through inlets from surges generated on the open coast.		
water level change	*caused by*	wind
		flow of water

Table 29. High-density KRC: 3 propositions and 1 hierarchical level (cause+cause)

In constrast, there are other KRCs where causality is expressed with different KPs each time. In these cases the number of conceptual propositions and the complexity of causality tend to increase (Table 30).

Along the studied beaches, the persistent **longshore transport gradient** can *result from* **longshore transport divergence** *induced by* **wave refraction** over an ebb-tidal shoal, **flood currents** along the beach proximal to a tidal inlet, **increased wave energy** *due to* a nearshore **dredge pit**, and total **littoral blockage** *by* structured inlet with a minimal ebb-tidal shoal.				
longshore transport gradient	*caused_by*	longshore transport divergence	*caused_by*	wave refraction
		flood current		
		increased wave energy	*caused_by*	dredge pit
		littoral blockage	*caused_by*	structured inlet

Table 30. High-density KRC: 3 propositions and 1 hierarchical level (cause+cause)

In the case shown in Table 27, despite the KP repeatedly used (i.e. *cause*), not only causality is actually codified, since some of the effects can be understood in terms of the *affect* semantic relation. For example, the clause "causing problems for agriculture" should be understood as "something *affects* agriculture", since "problems for agriculture" is not a concept itself. Also, "causes ecological problems such as changes in evaporation and rainfall" can be conceptually rephrased as "something *affects* evaporation and rainfall".

As can be observed, in causal KRCs, multiple conceptual propositions may be extracted but few hierarchical levels are activated by most of them. We can encounter KRCs showing several different causes linked to their effects through a single hierarchical level (e.g. Table 26 or 28) or a central concept from which a causal chain is deployed (e.g. Table 31), but not more than 2 levels have been found. In the first case, there will be more chances of finding the same KP conveying causality, whereas in the second, causality will be conveyed through a greater variety of KPs.

These types of **earthquakes** also frequently *cause* large **submarine** (underwater) **landslides** or **slumps**, which also *generate* **tsunamis**.				
earthquake	*causes*	submarine landslide	*causes*	tsunami
		submarine slump		

Table 31. High-density KRC: 4 propositions 2 hierarchical levels (cause+cause)

Conclusions and Future Work

In this paper we have shown how the combination of word sketches with KPs can provide a reliable user-friendly method for the extraction of KRCs. KRC extraction can be further improved by agglutinating different semantic word sketches in the same query, thus giving rise to high-density KRCs. Three different sets of this kind of KRCs were collected: (1) KRCs where three different semantic relations could be found; (2) KRCs codifying two semantic relation types but several conceptual propositions; and (3) KRCs codifying only a semantic relation type but several conceptual propositions.

As the number of KRCs with three different semantic relations (hyponymy, meronymy and causality) was relatively low, the next logical step was to extract high-density KRCs where only one or two relation types were found. These KRCs were much more numerous in the corpus and their combinatorial patterns provided further insights into how these KRCs behave.

High-density KRCs were characterized based on the amount of knowledge conveyed, the concept types involved, the conceptual depth activated and the number and type of conceptual relations codified in them. For instance, in KRCs combining *type_of* and *part_of* relations, deeper hierarchical structures can be found compared to the rest.

Not all the high-density KRCs collected were equally useful for different types of users. For example, KRCs that express causality often included entire clauses that are difficult to merge in a single concept. Therefore, despite the fact that these contexts contained unambiguous KPs pointing to useful conceptual relations, their use as KRCs may be restricted to terminology users rather than terminographers when building a TKB. In contrast, KRCs where too many conceptual propositions are activated may be more useful for terminographers than for terminology users.

The analysis of KRCs can help to extract new KPs to improve the ESSG through the refinement of grammar rules or the formalization of new relations. Nevertheless, we must be cautious when integrating new KPs in the ESSG, since they can be more noisy than productive, as is the case of prepositions such as *in* and *from*, and verbs such as *have*.

As future work, we plan to expand this analysis to KRCs beyond one sentence and to KRCs where hyponymy and meronymy are combined with the *has_location* or *has_function* relations, which are already developed in the ESSG. A more thorough and quantitative analysis will be conducted on those KRCs where only two relations apply but from which several propositions may be extracted.

Furthermore, the extraction of high-density KRCs will be more productive as soon as the ESSG is refined and improved. Finally, following the approach of Condamines et al. (2013), it would be interesting to explore the usefulness of high-density KRCs as well as the cognitive effort they require for end users such as translators.

Notes

1. The EEC is freely available in the.sketchengine.co.uk/open and the ESSG can be downloaded and reused with any other corpus following the instructions on http://ecolexicon.ugr.es/essg
2. More details on the Dublin Core Metadata Initiative can be found on http://dublincore.org/specifications/dublin-core/
3. The latest version of the ESSG can be downloaded from http://ecolexicon.ugr.es/essg/
4. By conceptual proposition we mean the triple concept-relation-concept. QUARTZ *type_of* MINERAL would be a hyponymic proposition.

References

Aussenac-Gilles, N., & Jacques, M.-P. (2008). Designing and evaluating patterns for relation acquisition from texts with Caméléon. *Terminology, 14*(1), 45-73. Retrieved from http://doi.org/10.1075/term.14.1.04aus

Barrière, C. (2004). Knowledge-Rich Contexts Discovery. In *Seventeenth Canadian Conference on Artificial Intelligence (AI'2004)*, Vol. 3060. London, Ontario: CSCSI, 187-201. Retrieved from http://doi.org/10.1007/978-3-540-24840-8_14

Barrière, C., & Agbago, A. (2006). TerminoWeb: a software environment for term study in rich contexts. In *Conference on Terminology, Standardisation and Technology Transfer (TSTT 2006)*. Beijing.

Bowker, L. (2003). Lexical Knowledge Patterns, Semantic Relations, and Language Varieties: Exploring the Possibilities for Refining Information Retrieval in an International Context. *Cataloging & Classification Quarterly, 37*(1–2), 153-171.

Condamines, A., Josselin-Leray, A., Fabre, C., Lefeuvre, L., Picton, A., & Rebeyrolle, J. (2013). Using Comparable Corpora to Characterize Knowledge-Rich Contexts for Various Kinds of Users: Preliminary Steps. In *Proceedings of the 5th International Conference on Corpus Linguistics. Procedia-Social and Behavioral Sciences 95*, 581-586.

Faber, P., León-Araúz, P., & Reimerink, A. (2014). Representing environmental knowledge in EcoLexicon. In *Languages for Specific Purposes in the Digital Era. Educational Linguistics, 19*. Springer, 267-301.

Faber, P., León-Araúz, P., & Reimerink, A. (2016) EcoLexicon: new features and challenges. In I. Kernerman, I. Kosem Trojina, S. Krek, & L. Trap-Jensen (Eds.) *GLOBALEX 2016: Lexicographic Resources for Human Language Technology in conjunction with the 10th edition of the Language Resources and Evaluation Conference*, Portorož, 73-80.

Kilgarriff, A., Rychly, P., Smrz, P., & Tugwell, D. (2004). The Sketch Engine. In G. Williams & S. Vessier (Eds.) *Proceedings of the Eleventh EURALEX International Congress*. Lorient: EURALEX, 105-116.

Kilgarriff, A., Husák, M., McAdam, K., Rundell, M., & Rychlý, P. (2008) GDEX: Automatically Finding Good Dictionary Examples in a Corpus. In E. Bernal & J. DeCesaris (Eds.) *Proceedings of the XIII EURALEX International Congress*. Barcelona: Universitat Pompeu Fabra.

León-Araúz, P., San Martín, A. & Reimerink, A. (2018) The EcoLexicon English Corpus as an open corpus in Sketch Engine. In *Proceedings of the XVIII EURALEX International Congress*. Ljubljana, Slovenia.

León-Araúz, P. & San Martín, A. (2018). The EcoLexicon Semantic Sketch Grammar: from Knowledge Patterns to Word Sketches. In I. Kerneman, & S. Krek (Eds.) *Proceedings of the LREC 2018 Workshop "Globalex 2018 – Lexicography & WordNets"*. Miyazaki: Globalex, 94-99.

León-Araúz, P., San Martín, A. & Faber, P. (2016). Pattern-based Word Sketches for the Extraction of Semantic Relations. In *Proceedings of the 5th International Workshop on Computational Terminology*. Osaka, Japan, 73-82.

Marshman, E. (2002). The cause-effect relation in a biopharmaceutical corpus: English knowledge patterns. In *Proceedings of the 6th international Conference on Terminology and Knowledge Engineering*. Nancy, 89-94.

Meyer, I. (2001). Extracting knowledge-rich contexts for terminography: a conceptual and methodological framework. In D. Bourigault, C. Jacquemin, M.C. L'Homme, (Eds.) *Recent Advances in Computational Terminology*. John Benjamins, Amsterdam, 279-302.

San Martín, A., Cabezas-García, M., Buendía, M., Sánchez-Cárdenas, B., León-Araúz, P. & Faber, P. (2017) Recent Advances in EcoLexicon. *Dictionaries: Journal of the Dictionary Society of North America*, 38(1):96-115.