Pattern-based Word Sketches for the Extraction of Semantic Relations

Pilar León-Araúz
University of Granada
Department of Translation and Interpreting
Buensuceso, 11
18001 Granada (Spain)
pleon@ugr.es

Antonio San Martín
Maynooth University
Department of Spanish and Latin American Studies
Arts Building, North Campus
Maynooth, Co. Kildare (Ireland)
antonio.sanmartin@nuim.ie

Pamela Faber
University of Granada
Department of Translation and Interpreting
Buensuceso, 11
18001 Granada (Spain)
pfaber@ugr.es

Abstract

Despite advances in computer technology, terminologists still tend to rely on manual work to extract all the semantic information that they need for the description of specialized concepts. In this paper we propose the creation of new word sketches in Sketch Engine for the extraction of semantic relations. Following a pattern-based approach, new sketch grammars are developed in order to extract some of the most common semantic relations used in the field of terminology: generic-specific, part-whole, location, cause and function.

1 Introduction

Terminological work is mostly based on corpus analysis because it is in texts where experts express knowledge and make it accessible (Bourigault and Slodzian 1999). The most basic way of using a corpus is by manually reading concordance lines containing a given term. However, this is time-consuming and inefficient, which has led to the development of new corpus-based methods and applications to analyze and extract information.

One of the most common approaches for the efficient extraction of information from a corpus is to search for knowledge-rich contexts (KRCs). A KRC is "a context indicating at least one item of domain knowledge that could be useful for conceptual analysis" (Meyer 2001). In order to find KRCs in corpora, knowledge patterns (KPs) are used, which are the linguistic and paralinguistic patterns that convey a specific semantic relation (Meyer 2001).

KPs have been successfully applied in many terminology-related projects that have led to the creation of knowledge extraction tools, such as Caméléon (Aussenac-Gilles and Jacques 2008) and TerminoWeb (Barrière and Agbago 2006). However, to the best of our knowledge, currently there are no user-friendly publicly available applications allowing terminologists to find KRCs in their own corpora with ready-made KPs. For this reason, terminologists still tend to rely on manual work to extract all the semantic information that they need for the description of specialized concepts.

In order to fill this void, we propose the creation of KP-based sketch grammars in the well-known corpus query system, Sketch Engine (Kilgarriff et al. 2004). This allows users to generate new word sketches that could be exploited by any terminologist, lexicographer or translator interested in the extraction of semantic relations.

Word sketches are automatic corpus-derived summaries of a word's grammatical and collocational behavior (Kilgarriff et al. 2004). Rather than looking at an arbitrary window of text around the headword – as occurs in previous corpus tools – Sketch Engine is able to look for each grammatical relation that the word participates in (Kilgarriff et al. 2004). The default word sketches provided by Sketch Engine represent different relations, such as verb-object, modifiers or prepositional phrases. However,

Proceedings of the 5th International Workshop on Computational Terminology,
pages 73–82, Osaka, Japan, December 12 2016.

with the exception of the recently implemented generic-specific word-sketches, they only represent linguistic relations. Therefore, we believe that the development of new sketch grammars focusing on the extraction of semantic relations is a timely contribution to the field of terminology, since a summary of the semantic behavior of concepts in the form of word sketches would allow terminologists to perform a more efficient conceptual analysis of any corpus uploaded to Sketch Engine.

The new sketch grammars presented in this paper have been developed for the extraction of some of the most common semantic relations used in the field of terminology, namely: generic-specific, part-whole, location, cause and function. Section 2 briefly reviews previous work on KPs and semantic relations; Section 3 shows the methodology followed to derive and formalize KPs; Section 4 presents our preliminary results; and Section 5 provides the conclusions derived from this work.

2 Semantic relations and knowledge patterns

The extraction of semantic relations from specialized corpora constitutes one of the most important tasks in terminology work, since many other tasks depend on them (i.e. conceptual modeling, definition elaboration). From a user perspective, the visualization of semantic relations is essential to comprehend how a concept interacts with others in a specialized domain (Faber, León-Araúz and Prieto Velasco 2009).

Thus, not surprisingly, the automatic retrieval of related term pairs has been explored for many years and from different perspectives. One of them is based on KPs, which are considered one of the most reliable methods for the extraction of semantic relations (Condamines 2002; Marshman, Morgan, and Meyer 2002; Marshman 2002; Barrière 2004; Bowker 2003; L'Homme and Marshman 2006; Cimiano and Staab 2005; Auger and Barrière 2008; Lefever, Kauter, and Hoste 2014; Marshman 2014; Lafourcade and Ramadier 2016). The term KP was coined by Meyer (2001) to refer to the lexico-syntactic patterns between the terms encoded in a proposition in real texts, but they were introduced much earlier by Hearst (1992). Since then, much has been written about them. Nevertheless, despite their popularity, KPs are still far from being fully studied and exploited, especially in specialized domains. Furthermore, as observed by Bowker (2003), there are still major problems with regard to noise and silence, pattern variation, anaphora, domain and language dependency, etc. Moreover, not all relations have been analyzed in the same depth.

Patterns conveying hyponymic relations are the most commonly studied since they play an important role in categorization and property inheritance (Barrière 2004). Some of the simplest examples of such KPs are *x is a kind of y*, *As include Bs, Cs and Ds* (Meyer 1994) and *comprise(s)*, *consist(s)*, *define(s)*, *denote(s)*, *designate(s)*, *is/are*, *is/are called*, *is/are defined as*, *is/are known as* (Pearson 1998).

Meronymy, or part-whole relations, have also been previously researched (Berland and Chamiak 1999; Girju, Badulescu, and Moldovan 2003) and common patterns include *part of*, *constituent of*, *constituted by*, *made of*, *composed of*, *contains*, etc. These relations may be codified by prepositional phrases, possessives, and partitive verbs, but one of their main features is the fact that many KPs can be polysemic. For instance, *including* expresses both hyponymy and meronymy; and *formed by* expresses meronymy and causality (León Araúz 2014; León Araúz and Reimerink 2010).

Although to a lesser extent, other non-hierarchical relations have also been studied and implemented as KPs, each has their own peculiarities. For instance, unlike certain fairly clear-cut hierarchical relations, such as generic-specific, *cause* has many different subtypes (Marshman 2002). All studies dealing with causality affirm that there are many ways to express causation since it can be expressed by passive, active, subject-object, nominal or verbal propositions. Moreover, causes and effects have very diverse syntactic representations. More specifically, causation is not only expressed by constructions such as *due to* or *because of*, but also by causative nouns (*cause* or *consequence*) and verbs. Although there are many causative verbs (e.g. *cause*, *generate*, *lead*, *produce*, etc.), their syntactic behavior can also vary. As a result, one single grammar would not be sufficient to formalize their complementation structures (León Araúz and Faber 2012).

The above-mentioned patterns are only a simplification of what is actually found in a corpus. For instance, when formalizing the pattern *is a type of* we should also take into account all of its possible variants. The verb *to be* may be in its plural form or substituted by a comma; if it is in the plural, various hyponyms will be enumerated to the left of the pattern; the verb *to be* may be preceded by a modal

verb; the word *type* may be preceded by an adjective and an adverb; and it may be substituted by other synonyms such as *kind*, *sort*, *example*, *group*, etc. This is in line with the types of KP identified by Meyer (2001): lexical patterns (literal strings); grammatical patterns (taking into account POS tags); and paralinguistic patterns (punctuation).

Therefore, these patterns can be useful as they are when manually querying a corpus. However, formalizing them in grammars requires finding the balance between precision and recall, and efficiency and complexity. This entails having to decide the number of possible paths that the same grammar may cover, how many elements will be optional or compulsory, whether the anchor points should be literal words or lemmas, POS tags or punctuation marks, while taking into account negative adverbs (*not*, *never*, *hardly*) that would give a false positive, etc.

3 Materials and methods

For this study, we used the English EcoLexicon corpus[1], which currently contains over 59 million words in English and is focused on the environmental domain. Although KPs have been tested in a domain-specific corpus, we believe that most of them could also be applied to other domains. Except for patterns such as *built for* or *built with*, which would only be activated in construction related domains, most of them are not domain-specific.

For corpus querying and the generation of word sketches, we employed Sketch Engine. Corpus querying in Sketch Engine is based on an extension of the CQL formalism (Schulze and Christ 1996), allowing for the formalization of grammar patterns in the form of regular expressions combined with POS-tags. CQL expressions in Sketch Engine can be used as one-time queries (giving access to matching concordance lines) or stored in a sketch grammar, which will produce word sketches.

As previously stated, the only semantic relation included in the default English sketch grammar so far is the hyponymic word sketch. Table 1 shows the resulting word sketch when querying *earthquake* in the general publicly available English Web 2013 (enTenTen13) corpus:

... is a "earthquake"		
	534	0.00
body	57	5.01
mind	46	5.15
event	41	3.92
heart	27	5.25
example	26	3.61

"earthquake" is a ...		
	1,295	0.00
disaster	80	7.72
event	75	5.01
result	74	3.65
part	57	1.04
time	48	2.20

Table 1.... *is a* word sketch in Sketch Engine.[2]

The results in Table 1 would not be satisfactory for a terminologist. However, more sophisticated hyponymic KPs will soon be implemented in Sketch Engine in order to extract definitions (Kovář, Močiariková, and Rychlý 2016). Also, from a terminology perspective, Baisa and Suchomel (2015) have already explored hyponymy extraction by using sketch grammars in a specialized Czech corpus on the domain of land surveying. In line with our view, they acknowledge that apart from the term extraction function, terminologists need a function for placing the extracted terms in a tree structure.

Nonetheless, apart from placing terms in a tree structure, terms also need to be linked to others by means of other semantic relations. In what follows we explain our methodology for the extraction of generic-specific, part-whole, location, cause, and function relations.

Besides collecting the patterns mentioned by other authors (see Section 2), we also added our own based on our experience during the construction of EcoLexicon. All approaches seem to agree that the use of KPs for knowledge extraction involves a series of complementary steps. Nevertheless, the order of the steps differs depending on research objectives (e.g. identification of term pairs, discovery of

[1] This corpus was compiled by the LexiCon Research Group for the development of EcoLexicon (http://ecolexicon.ugr.es), a terminological knowledge base on the environment.

[2] The second column shows the number of occurrences and the third one the collocation strength score as calculated by Sketch Engine.

new KPs, searching for known KPs to discover new term pairs, etc.). In our case, we followed the following steps:

1. *Collection of KPs*: this first stage only includes the collection of patterns in plain English (no formalism or encoding language used).
   a. Patterns referenced by other authors.
   b. Patterns already known.
   c. Recursive method: term pairs linked by already known semantic relations are searched for to find new patterns. Then these patterns are used to find new term pairs, and so on.

2. *CQL encoding*: This second stage consists of translating the KPs collected during the first stage into CQL sketch grammars.
   a. Splitting or lumping: Some KPs collected in the first stage can be lumped into a single CQL sketch grammar, while others collected as a single KP need to be split.
   b. Addition of adverbs, punctuation, modal verbs, relative phrases, adjectives, determiners, etc.

3. *Validation, enrichment, refining*
   a. CQL patterns are validated trying to keep the balance between noise and silence.
   b. Enrichment: Testing the CQL patterns with additional optional elements to spot new variations of the pattern (for instance, the possibility of an adverb in a place where it was not previously accounted for). Validation of the new addition.
   c. Refining: Detection of erroneous concordance lines obtained with the CQL patterns. Analysis of the source of the error, and determination of whether it is appropriate to change the CQL pattern.

In the development of our sketch grammars (a total of 56), we have considered different issues that are specific to each relation. For instance, there are certain patterns that always take the same form and order (e.g. *such as*), whereas others show such a diverse syntactic structure that the directionality of the pattern must also be accounted for. We also had to take into account the fact that a single sentence could 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 non-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. In this sense, we observed that 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 in each grammar according to the semantic relation conveyed.

Generic-specific (18 sketch grammars): 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

Part-whole (17 sketch grammars): 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
Cause (10 sketch grammars): 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
Location (4 sketch grammars): 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
Function (7 sketch grammars): 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

Table 2. Simplified summary of knowledge patterns and semantic relations.

By way of example, Tables 3 and 4 show the actual CQL representation of a generic-specific KP and a causal KP respectively, followed by an explanation.

2:"N.*" [tag!="V.*"]{0,5} "MD"? [word!="not"]? [lemma="be|,|\("] [word!="not"]? [word="defined|classified|categori.ed|regarded"] [word="as"] "DT.*|RB.*|JJ.*"* ([lemma="type|kind|example|group| class| sort|category|family|species|subtype|subfamily|subgroup| subclass|subcategory|subspecies"] [word="of"])? [tag!="V.*"]{0,2} 1:[tag="N.*" & lemma!="type|kind|example|group|class|sort|category|family|species|subtype|subfamily|subgroup|subclass|subcategory|subspecies"]	
2:"N.*"	The hyponym is a noun.
[tag!="V.*"]{0,5}	From 0 to 5 words that are not verbs. This allows to capture enumerations and allows for the presence of adverbs, prepositions, etc.
"MD"?	An optional modal verb
[word!="not"]?	Optional word that is not *not*. This filters out negative sentences.
[lemma="be|,|\("]	The lemma *be,* comma or opening parenthesis.
[word!="not"]?	Optional word that is not *not*. This filters out negative sentences.
[word="defined|classified|categori.ed|regarded"]	The words *defined*, *classified*, *categorized*, *categorised* or *regarded*.
[word="as"]	The word *as*.
"DT.*|RB.*|JJ.*"*	From 0 to infinite determiners, adverbs or adjectives. This allows for phrases such as “the most important”, “a very special”, etc.

([lemma="type|kind|example|group|class|sort|category|family|species|subtype|subfamily|subgroup|subclass|subcategory|subspecies"] [word="of"])?	The lemma *type, kind, example, group, class, sort, category, family, species, subtype, subfamily, subgroup, subclass, subcategory* or *subspecies* followed by the word *of* (both optional).
[tag!="V.*"]{0,2}	From 0 to 2 words that are not verbs. This allows for the presence of determiners, adjectives, adverbs, etc.
1:[tag="N.*" & lemma!="type|kind|example|group|class|sort|category|family|species|subtype|subfamily|subgroup|subclass|subcategory|subspecies"]	The hypernym is a noun that does not have *type, kind, example, group, class, sort, category, family, species, subtype, subfamily, subgroup, subclass, subcategory* or *subspecies* as lemma.

Table 3. CQL representation of a generic-specific KP followed by its explanation.

2:"N.*" [tag!="V.*"]{0,7} [lemma="be|,|\("]? [tag="RB.*" & word!="not|never"]* [word="caused|produced|generated|provoked|induced|triggered|originated"] "RB.*"* ([word="by"]|[word="because"][word="of"] | [word="due"] [word="to"]) [tag!="V.*"]{0,7} 1:"N.*"	
2:"N.*"	The effect is a noun.
[tag!="V.*"]{0,7}	From 0 to 7 words that are not verbs. This allows to capture enumerations and allows for the presence of adverbs, prepositions, etc.
[lemma="be|,|\("]?	The lemma *be,* comma or opening parenthesis.
[tag="RB.*" & word!="not|never"]*	From 0 to infinite adverbs except *not* or *never*.
[word="caused|produced|generated|provoked|induced|triggered|originated"]	The word *caused*, *produced*, *generated*, *provoked*, *induced*, *triggered* or *originated*.
"RB.*"*	From 0 to infinite adverbs.
([word="by"]|[word="because"] [word="of"] | [word="due"] [word="to"])	The word *by*, the phrase *because of* or the phrase *due to*.
[tag!="V.*"]{0,7}	From 0 to 7 words that are not verbs.
1:"N.*"	The cause is a noun.

Table 4. CQL representation of a causal KP followed by its explanation.

These grammars combine our previously retrieved KPs, which act as anchor points, with certain constraints imposed by POS tags, punctuation or operators (i.e.?, *, {0,5}), which means that they include all types of KPs (lexical, grammatical and paralinguistic). Tables 4 and 5 show a sample of the concordances that can be extracted with several of our generic-specific and causal grammars:

bacteria , viruses, protozoans worms and other types of agents
Bacteria and protozoa are the major groups of microorganisms
bacteria are the main types of organisms
Clouds are classified into four families: high clouds, middle clouds, low clouds
materials are classified by grain size into clay, silt, sand, gravel, cobble, and boulder
Cumulonimbus is classified as a low cloud
weather phenomena such as local storms, tornadoes, hurricanes, or extra-tropical and tropical cyclones
sediment , usually sand but occasionally silt or clay
structures , namely headland breakwaters, nearshore breakwaters, and a groin field
sea stars, urchins, sea cucumbers, and other creatures

Table 5. Concordances extracted with generic-specific grammars.

earthquakes can trigger massive landslides
flooding causes many deaths and much damage
Pesticides and commercial inorganic fertilizers cause air, water, and soil pollution
cancers caused by air pollution
radiation can lead to cancer
erosion results from storms
damage caused by severe winds
tsunami causes massive destruction

Table 6. Concordances extracted with causal grammars.

4 Results

The combination of our sketch grammars with the statistics used in the Sketch Engine system has yielded encouraging results. To show the potential of this initial approach, we have selected different concepts showing word sketches for all types of relation and their inverse in Table 7. The results are sorted by frequency. Because of space constraints, only the first few results of each word sketch are shown.

"bacterium" is a type of...	1,007	0.12
organism	158	10.00
microorganism	88	10.92
micro-organism	28	9.64
agent	18	8.09
decomposer	15	8.83

"bacterium" is the generic of...	1,028	0.12
coli	17	8.94
plant	14	6.85
Pseudomonas	10	8.24
Escherichia	10	8.22
fungus	9	7.60

"rock" has part...	3,029	0.09
mineral	213	10.54
quartz	65	9.17
fragment	47	8.79
feldspar	45	8.79
plagioclase	41	8.67

"rock" is part of...	2,055	0.06
crust	44	9.09
soil	34	7.97
belt	27	8.52
continent	23	8.30
part	22	7.96

"volcano" is located at...	318	0.04
plate	17	10.11
island	14	9.42
boundary	11	9.38
Pacific	8	8.71
margin	7	8.87

"volcano" is the location of...	71	0.01
cone	7	11.10
ocean	3	8.23
type	3	6.74
area	3	6.59
precursor	2	9.66

"tsunami" is the cause of...	196	0.04
damage	18	7.54
destruction	12	8.74
erosion	7	6.70
devastation	6	9.08
death	6	6.67

"tsunami" is caused by...	1,057	0.20
earthquake	177	11.31
landslide	68	10.73
eruption	36	9.34
water	33	7.70
movement	23	8.65

"energy" has function...	2,151	0.03
water	57	8.71
produce	41	8.83
make	33	8.45
process	22	7.99
electricity	21	8.17

"energy" is the function of...	999	0.02
fuel	23	8.96
carbon	14	8.12
biomass	13	8.44
waste	13	8.20
light	12	8.28

Table 7. Examples of different word sketches obtained with our sketch grammars

There are several issues that still need to be dealt with in order to improve the outcome of these grammars. For instance, (1) there is still noise because the grammars need to be refined, especially that of function, where target terms may be nouns or verbs, and verbs are not always semantically relevant or self-contained (i.e. *make*, *produce*) and need an object to constitute a meaningful proposition; (2) most false positives (i.e. *fungus* or *plant* as a type of *bacterium*, or *type* as something located at a *volcano*, as shown in Table 5) are due to the imprecision of certain grammars or even to some mistakes derived from the POS tagger; (3) there is also pattern ambiguity that could only be solved by adding semantic constraints on the type of entities being linked (as done by Girju, Badulescu, and Moldovan 2003); (4) and semantic relations are also ambiguous, for example in the sense that distinguishing parts from locations is not always an easy task (i.e. *cone* could be a part of a *volcano* or be located in a *volcano*). Furthermore, of these issues there is one related to the processing of multiword expressions. So far, these word sketches only retrieve one-word terms, which is one of the causes of noise and lack of precision. This can be solved relatively easily (Kilgarriff et al. 2012), but poses the challenge of differentiating between multiword terms and usual collocations. For instance, in sentences (1) and (2) only *shield volcano* should take the role of the hyponym, whereas *huge* would only qualify as a simple modifier of *volcano*.

(1) "…monogenetic volcanoes are smaller than **polygenetic volcanoes**, *such as* **shield volcanoes**..."
(2) "…with igneous and **tectonic features** *such as* **huge volcanoes** and rift valleys..."

5 Conclusions and future work

In this paper we have shown how KPs can be converted into sketch grammars to generate new word sketches showing semantic relations. The resulting word sketches can be of great value to terminologists during the conceptual modeling of any domain. However, much remains to be done. First of all, the sketch grammars should be refined as new patterns are found and extended to include multiword terms. New grammars will also be needed to include other semantic relations, especially those related to process concept types, such as temporal relations. Precision and recall studies will be performed in order to improve the grammars and find the right balance between noise and silence. Finally, pattern disambiguation techniques are also needed for polysemic KPs.

References

Auger, Alain, and Caroline Barrière. 2008. "Pattern-Based Approaches to Semantic Relation Extraction: A State-of-the-Art." *Terminology* 14 (1): 1–19. doi:10.1075/term.14.1.02aug.

Aussenac-Gilles, Nathalie, and Marie-Paule Jacques. 2008. "Designing and Evaluating Patterns for Relation Acquisition from Texts with Caméléon." *Terminology* 14 (1): 45–73. doi:10.1075/term.14.1.04aus.

Baisa, Vít, and Vít Suchomel. 2015. "Corpus Based Extraction of Hypernyms in Terminological Thesaurus for Land Surveying Domain." In *Ninth Workshop on Recent Advances in Slavonic Natural Language Processing*, 69–74. Brno: Tribun EU.

Barrière, Caroline. 2004. "Knowledge-Rich Contexts Discovery." In *Seventeenth Canadian Conference on Artificial Intelligence (AI'2004)*, 187–201. London, Ontario: CSCSI. doi:10.1007/978-3-540-24840-8_14.

Barrière, Caroline, and A Agbago. 2006. “TerminoWeb: A Software Environment for Term Study in Rich Contexts.” In *Conference on Terminology, Standardisation and Technology Transfer (TSTT 2006)*. Beijing. http://nparc.cisti-icist.nrc-cnrc.gc.ca/npsi/ctrl?action=rtdoc&an=8913210.

Berland, Matthew, and Eugene Charniak. 1999. “Finding Parts in Very Large Corpora.” In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, 57–64. Morristown, NJ: Association for Computational Linguistics. doi:10.3115/1034678.1034697.

Bourigault, Didier, and Monique Slodzian. 1999. “Pour Une Terminologie Textuelle.” *Terminologies Nouvelles* 19: 29–32.

Bowker, Lynne. 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): 153–71. doi:10.1300/J104v37n01_11.

Cimiano, Philipp, and Steffen Staab. 2005. “Learning Concept Hierarchies from Text with a Guided Agglomerative Clustering Algorithm.” In *Proceedings of ICML 2005. Workshop on Learning and Extending Lexical Ontologies with Machine Learning Methods*, edited by Chris Biemann and Gerhard Paas. Bonn. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.59.7546&rep=rep1&type=pdf.

Condamines, Anne. 2002. “Corpus Analysis and Conceptual Relation Patterns.” *Terminology* 8 (1): 141–62. doi:10.1075/term.8.1.07con.

Faber, Pamela, Pilar León Araúz, and Juan Antonio Prieto Velasco. 2009. “Semantic Relations, Dynamicity, and Terminological Knowledge Bases.” *Current Issues in Language Studies* 1: 1–23.

Girju, Roxana, Adriana Badulescu, and Dan Moldovan. 2003. “Learning Semantic Constraints for the Automatic Discovery of Part-Whole Relations.” In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - NAACL ’03*, 1:1–8. Morristown, NJ: Association for Computational Linguistics. doi:10.3115/1073445.1073456.

Hearst, Marti A. 1992. “Automatic Acquisition of Hyponyms from Large Text Corpora.” In *Actes de COLING-92*, 2:539–45. Morristown, NJ: International Committee on Computational Linguistics.

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, and Vít Suchomel. 2014. “The Sketch Engine: Ten Years on.” *Lexicography* 1 (1): 7–36. doi:10.1007/s40607-014-0009-9.

Kilgarriff, Adam, Pavel Rychlý, Vojtěch Kovář, and Vít Baisa. 2012. “Finding Multiwords of More Than Two Words.” *Proceedings of the 15th EURALEX International Congress*, 1–7.

Kilgarriff, Adam, Pavel Rychlý, Pavel Smrz, and David Tugwell. 2004. “The Sketch Engine.” In *Proceedings of the Eleventh EURALEX International Congress*, edited by Geoffrey Williams and Sandra Vessier, 105–16. Lorient: EURALEX.

Kovář, Vojtěch, Monika Močiariková, and Pavel Rychlý. 2016. “Finding Definitions in Large Corpora with Sketch Engine.” In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2016)*, edited by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis. Portorož, Slovenia: European Language Resources Association (ELRA).

L’Homme, Marie-Claude, and Elizabeth Marshman. 2006. “Terminological Relationships and Corpus-Based Methods for Discovering Them: An Assessment for Terminographers.” In *Lexicography, Terminology and Translation. Text-Based Studies in Honour of Ingrid Meyer*, edited by Lynne Bowker, 67–80. Ottawa: University of Ottawa Press.

Lafourcade, Mathieu, and Lionel Ramadier. 2016. “Semantic Relation Extraction with Semantic Patterns Experiment on Radiology Reports.” In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2016)*, edited by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, 4578–82. Portorož, Slovenia: European Language Resources Association (ELRA).

Lefever, E, M Van de Kauter, and V Hoste. 2014. “HypoTerm: Detection of Hypernym Relations between Domain-Specific Terms in Dutch and English.” *Terminology* 20 (2): 250–78. doi:10.1075/term.20.2.06lef.

León Araúz, Pilar. 2014. “Semantic Relations and Local Grammars for the Environment.” In *Formalising Natural Languages with NooJ 2013*, edited by Svetla Koeva, Slim Mesfar, and Max Silberztein, 87–102. Newcastle-upon-Tyne: Cambridge Scholars Publishing.

León Araúz, Pilar, and Pamela Faber. 2012. “Causality in the Specialized Domain of the Environment.” In *Proceedings of the Workshop “Semantic Relations-II. Enhancing Resources and Applications” (LREC’12)*, edited by Verginica Barbu Mititelu, Octavian Popescu, and Viktor Pekar, 10–17. Istanbul: ELRA.

León Araúz, Pilar, and Arianne Reimerink. 2010. “Knowledge Extraction and Multidimensionality in the Environmental Domain.” In *Proceedings of the Terminology and Knowledge Engineering (TKE) Conference 2010*. Dublin: Dublin City University.

Marshman, Elizabeth. 2002. “The Cause-Effect Relation in a Biopharmaceutical Corpus: English Knowledge Patterns.” In *Proceedings of the 6th International Conference on Terminology and Knowledge Engineering*, 89–94. Nancy.

———. 2014. “Enriching Terminology Resources with Knowledge-Rich Contexts: A Case Study.” *Terminology* 20 (2): 225–49. doi:10.1075/term.20.2.05mar.

Marshman, Elizabeth, Tricia Morgan, and Ingrid Meyer. 2002. “French Patterns for Expressing Concept Relations.” *Terminology* 8 (1): 1–29. doi:10.1075/term.8.1.02mar.

Meyer, Ingrid. 1994. “Linguistic Strategies and Computer Aids for Knowledge Engineering in Terminology.” *L’actualité terminologique/Terminology Update* 27 (4): 6–10.

———. 2001. “Extracting Knowledge-Rich Contexts for Terminography.” In *Recent Advances in Computational Terminology*, edited by Didier Bourigault, Christian Jacquemin, and Marie-Claude L’Homme, 279–302. Amsterdam/Philadelphia: John Benjamins.

Pearson, Jennifer. 1998. *Terms in Context*. Amsterdam/Philadelphia: John Benjamins.

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