Specifying Hyponymy Subtypes and Knowledge Patterns: A Corpus-based Study

Juan Carlos Gil-Berrozpe, Pilar León-Araúz, Pamela Faber

University of Granada

Department of Translation and Interpreting, Buensuceso 11, 18071 Granada, Spain

E-mail: juancarlosgb@correo.ugr.es, pleon@ugr.es, pfaber@ugr.es

Abstract

The organization of a terminological knowledge base (TKB) relies on the identification of relations between concepts. This involves making an inventory of semantic relations and extracting these relations from a corpus by means of knowledge patterns (KPs). In EcoLexicon, a multilingual and multimodal TKB on the environment, 17 semantic relations are currently being used to link environmental concepts. These relations include six subtypes of meronymy, but only one subtype of hyponymy (*type_of*). However, a recent pilot study (Gil-Berrozpe et al., in press) showed that the generic-specific relation could also be subdivided. Interestingly, these preliminary results indicated that hyponymy subtypes were constrained by the ontological nature of concepts, depending on whether they were entities or processes. The new proposal presented in this paper expands the scope of our preliminary research on hyponymy subtypes to include concepts belonging to a wider range of semantic categories, and examines the behavior of knowledge patterns used to extract hyponymic relations. In this research, corpus analysis was used to explore the correlation of concepts in many different categories with KPs as well as with hyponymy subtypes. Thanks to these constraints, it was possible to formulate a more comprehensive inventory of generic-specific relations in the environmental domain.

Keywords: hyponymy subtypes; knowledge patterns; corpus analysis; concept nature

1. Introduction

In recent years, the study of terminology and specialized language has been undergoing a ‘cognitive shift’ (Faber, 2009: 111), which places a greater focus on conceptual representation and knowledge organization. In this line, descriptive theories of terminology (Cabré, 1999; Temmerman, 2000; Faber, 2009) now reflect dynamic phenomena (such as variation or multidimensionality) and emphasize the importance of hierarchical and non-hierarchical relations.

A crucial factor in the organization of a terminology knowledge base (TKB) lies in the relations between its terms (Barrière, 2004a). These semantic relations can be discovered through corpus analysis and the use of knowledge-rich contexts (KRC). Such contexts are highly informative since they provide conceptual information and domain knowledge (Meyer, 2001), and usually codify semantic relations in the form of knowledge patterns (KPs) (Meyer, 2001; Condamines, 2002; Barrière, 2004b; Agbago & Barrière, 2005; León-Araúz, 2014).

In recent years, much research has targeted the development of semi-automatized procedures for extracting KRCs (Jacquemin & Bourigault, 2005; Bielinskiene et al., 2012; Schumann, 2012), especially for hyponymic term pairs. Although recent work has focused on other conceptual relations, such as meronymy, function, and causality (Marshman, 2002; Girju et al., 2003; León-Araúz et al., 2016), hyponymy is a complex relation that requires a more in-depth study. As the backbone of hierarchical organization, it entails both categorization and property inheritance (Barrière, 2004a). Moreover, it is characterized by a variety of nuances and dimensions that should be further exploited (Gil-Berrozpe & Faber, 2016).

To explore the viability of our proposal, a pilot study (Gil-Berrozpe et al., in press) was conducted to ascertain whether the generic-specific relation could be subdivided in EcoLexicon[1] (Faber et al., 2014, 2016), a multilingual and multimodal TKB on environmental science. For this purpose, the EcoLexicon English Corpus[2] was processed with Sketch Engine (Kilgarriff et al., 2004), where the Word Sketch (WS) module was used. WSs are automatic corpus-derived summaries of a word's grammatical and collocational behavior (Kilgarriff et al., 2004). In this pilot study, we reconstructed the taxonomies of ROCK (an entity) and EROSION (a process). The resulting hierarchies were based on the analysis of (i) the default *modifier* WS, from which hyponymy can be extracted by analyzing the composition of multiword terms; (ii) a customized WS based on hyponymic KPs, where hyponymy was explicitly conveyed in the texts. The results showed that hyponymy subtypes were based on the semantic category of the concept, and were constrained by the nature of the concept, namely, whether it was an entity or a process.

This paper presents the results of a new study on hyponymy subtypes that includes concepts belonging to a wider range of semantic categories (e.g. activities, chemical elements, landforms, etc.), and analyzes the behavior of the knowledge patterns used to extract hyponymic relations. Accordingly, corpus analysis was used to explore the correlation of concepts in a variety of different categories with KPs as well as with hyponymy subtypes. These constraints led to a more comprehensive inventory of generic-specific relations in the environmental domain, as well as to a more accurate way of extracting them.

The rest of this article is organized as follows. Section 2 briefly presents the EcoLexicon TKB and explains how hyponymy refinement can enhance its conceptual networks. Section 3 explains the materials used and the methods followed to analyze semantic categories in relation to hyponymic KPs and hyponymy subtypes. In Section 4, the results of our research are presented and discussed. Section 5 highlights the conclusions that can be derived from this study and outlines plans for future research.

---

1 http://ecolexicon.ugr.es/

2 Part of this corpus (23 million words) is now available in Sketch Engine's Open Corpora (https://the.sketchengine.co.uk/open/).

The bibliography cited is followed by three appendices in which semantic categories, hyponymic knowledge patterns, and hyponymy subtypes are defined and exemplified.

2. Hyponymy refinement in EcoLexicon

EcoLexicon is a TKB on environmental science that is based on the theoretical premises of Frame-Based Terminology (Faber, 2012, 2015). Its objective is to facilitate user knowledge acquisition through different types of multimodal and contextualized information, in order to respond to cognitive, communicative, and linguistic needs. This resource is available in English and Spanish, although five more languages (German, Modern Greek, Russian, French and Dutch) are currently being added. To date, EcoLexicon has a total of 3,601 concepts and 20,212 terms.

EcoLexicon has a visual interface with different modules for conceptual, linguistic, and graphical information (Figure 1). Once a concept has been selected, it is represented in the center of an interactive map. Also displayed are the multilingual terms for that concept, as well as different conceptual relations between all the concepts belonging to the same network.

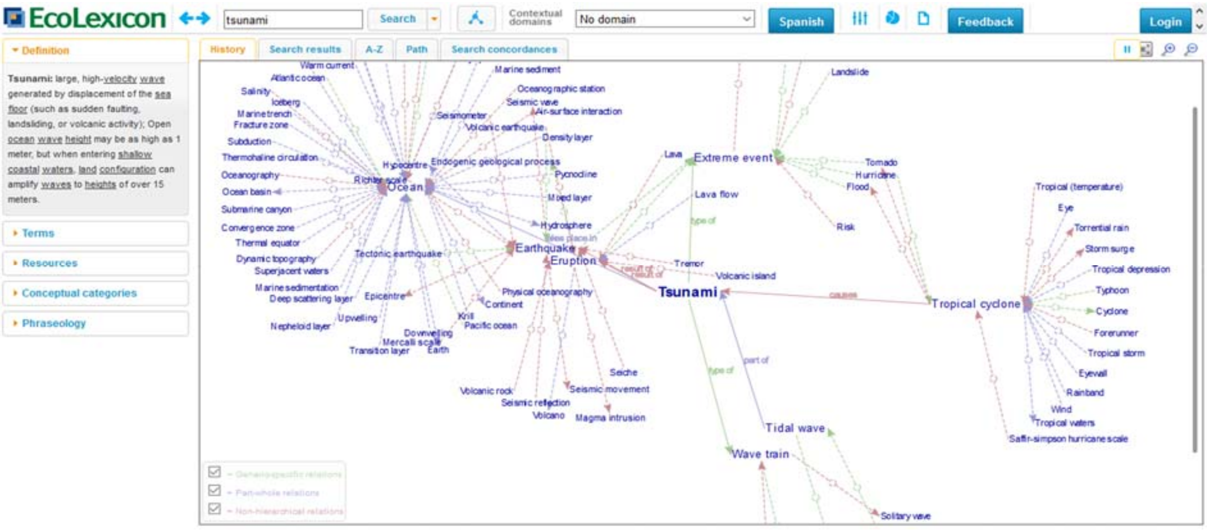

Figure 1: Visual interface of EcoLexicon (conceptual network of TSUNAMI).

The conceptual relations in EcoLexicon are classified as follows: (i) generic-specific relation (1 type); (ii) part-whole relations (6 types); (iii) non-hierarchical relations (10 types). Evidently, the generic-specific or hyponymic relation, which only has one subtype, would benefit from a more fine-grained representation since this would enhance its informativity and help to eliminate noise, information overload, and redundancy in the conceptual network (Gil-Berrozpe & Faber, 2016). Hyponymy is a semantic relation of inclusion whose converse is hyperonymy (Murphy, 2006: 446), and it can be refined by specifying subtypes (Murphy, 2003) or by establishing ‘facets’ and/or ‘microsenses’ (Cruse, 2002: 4-5).

Our pilot study (Gil-Berrozpe et al., in press) based hyponymy refinement on the following criteria: (i) the correction of property inheritance according to concept definitions; (ii) the creation of umbrella concepts; (iii) the decomposition of hyponymy into subtypes. As previously mentioned, our results indicated that hyponymy subtypes were based on whether the concept was an entity (ROCK) or a process (EROSION). For example, natural entities, such as ROCK, were found to have different sets of hyponyms based on formation (e.g. SEDIMENTARY ROCK, IGNEOUS ROCK), composition (SILTSTONE, SANDSTONE), and location (PLUTONIC ROCK, VOLCANIC ROCK).

3. Materials and methods

Our study analyzed hyponymic KPs as well as hyponymy subtypes. In both cases, the main information source was the EcoLexicon English corpus (67,903,384 words), which was uploaded to Sketch Engine. Apart from the default options, the system also permitted the creation of customized word sketches by storing CQL queries in new sketch grammars.

The corpus was thus compiled by implementing hyponymic sketch grammars developed by León-Araúz et al. (2016). These grammars are based on the KPs that generally reflect hyponymy in real texts. Simple examples of such KPs are *HYPERNYM such as HYPONYM*, *HYPONYM is a kind of HYPERNYM*, *HYPONYM and other HYPERNYM*, etc. These patterns were formalized as regular expressions combined with POS-tags, which resulted in 18 hyponymic sketch grammars. Table 1 shows a summarized version of the KPs.

1. HYPONYM ,|(|:|is|belongs (to) (a|the|...) type|category|... of HYPERNYM // **2.** types|kinds|... of HYPERNYM include|are HYPONYM // **3.** types|kinds|... of HYPERNYM range from (...) (to) HYPONYM // **4.** HYPERNYM (type|category|...) (,|() ranging (...) (to) HYPONYM // **5.** HYPERNYM types|categories|... include HYPONYM // **6.** HYPERNYM such as HYPONYM // **7.** HYPERNYM including HYPONYM // **8.** HYPERNYM ,|(especially|primarily|... HYPONYM // **9.** HYPONYM and|or other (types|kinds|...) of HYPERNYM // **10.** HYPONYM is defined|classified|... as (a|the|...) (type|kind|...) (of) HYPERNYM // **11.** classify|categorize|... (this type|kind|... of) HYPONYM as HYPERNYM // **12.** HYPERNYM is classified|categorized in|into (a|the|...) (type|kind|...) (of) HYPONYM // **13.** HYPERNYM (,|() (is) divided in|into (...) types|kinds|... :|of HYPONYM // **14.** type|kind|... of HYPERNYM (is|,|() known|referred|... (to) (as) HYPONYM // **15.** HYPONYM is a HYPERNYM that|which|... // **16.** define HYPONYM as (a|the|...) (type|category|...) (of) HYPERNYM // **17.** HYPONYM refers to (a|the|...) (type|category|...) (of) HYPERNYM // **18.** (a|the|one|two...) (type|category|...) (of) HYPERNYM: HYPONYM

Table 1: Hyponymic knowledge patterns (León-Araúz et al., 2016)

3.1 Hyponymic KPs and semantic categories

When the customized hyponymic sketch grammars were applied to the English EcoLexicon corpus, this created a filtered subcorpus, which was only composed of hyponymic concordances. This was accomplished by applying the CQL query *[ws(".*-n","\"%w\" is the generic of...",".*-n")]*. The resulting subcorpus contained a total of 938,386 potential hyponymic concordances (Figure 2).

Query .*-n, , is the generic of... 938,386 > Positive filter minerals 3,274 (38.55 per million)

Page 1 of 164 Go Next | Last

file429289...	Rivers also carry small rock fragments and minerals , including *clays* , which are produced
file4292891.	feldspar, mica, and, occasionally, heavy minerals such as *zircon* , tourmaline, and hornblende
file429289...	feldspar, mica, and, occasionally, heavy minerals such as zircon, tourmaline, and *hornblende*
file429289...	feldspar, mica, and, occasionally, heavy minerals such as zircon, *tourmaline* , and hornblende
file429289...	feldspar, mica, and, occasionally, heavy minerals such as zircon, tourmaline, and hornblende
file429289...	shape and generally belong to a group of minerals known as the *aluminosilicates* . These are
file429289...	*shape* and generally belong to a group of minerals known as the aluminosilicates. These are
file429289...	recombining the more reactive constituent minerals , such as micas and feldspars, while the
file429289...	recombining the more reactive constituent minerals , such as micas and *feldspars* , while the
file429289...	recombining the more reactive constituent minerals , such as *micas* and feldspars, while the
file429289...	recombining the more reactive constituent minerals , such as micas and feldspars, while the
file429289...	recombining the more reactive constituent minerals , such as micas and feldspars, while the
file429289...	.7, whereas *iron* oxides and other heavy minerals may be twice as dense. For all these reasons
file429289...	.7, whereas iron *oxides* and other heavy minerals may be twice as dense. For all these reasons
file429289...	sense, *clay* refers to a particular group of minerals , many of which occur in the clay fraction
file429289...	, clay refers to a particular group of minerals , many of which occur in the clay fraction
file429289...	in diameter. *Clay* minerals: A group of minerals found in the soil's clay fraction, generally
file429289...	in diameter. Clay *minerals* : A group of minerals found in the soil's clay fraction, generally
file429289...	regular three-dimensional pattern to form minerals such as quartz (*silicon* dioxide) or calcite
file429289...	regular three-dimensional pattern to form minerals such as quartz (silicon dioxide) or calcite

Page 1 of 164 Go Next | Last

Figure 2: Concordances retrieved from the hyponymic subcorpus

However, after filtering the hyponymic concordances in the EcoLexicon corpus with the customized word sketch, a manual process of data extraction was required. Since the customized word sketch was composed of 18 grammars describing a wide range of permutations and paraphrases of the hyponymic KPs, it was necessary to manually collect and analyze a representative sample of this information. Furthermore, the hyponymic subcorpus contained various identical sentences (since multiple hypernym-hyponym pairs in the same concordance were shown several times). There were also false positives that had to be eliminated from the results.

A randomized portion of the hyponymic subcorpus was examined, from which a set of 3,133 positive hyponymic concordances were selected to be the basis of the KP analysis. The extracted information was subsequently classified for analysis (Figure 3).

No.	Hypernym(s) [HYPER]	Hyponym(s) [HYPO]	Activated semantic category	Hyponymic pattern	Hyponymic pattern type
2635.	Acacia	Acacia tortilis, Capparis decidua	lifeform	types of HYPER, mainly HYPO	selection
1.	academic field	geography, architecture, psychology	domain	HYPO and other HYPER such as HYPO	itemization + exemplification
1585.	acid	H2SO4	element	HYPER such as HYPO	exemplification
1584.	acidic species	H2SO4, HCl, HF	element	HYPER such as HYPO	exemplification
2714.	acidic surface oxide	strong carboxylic, weak carboxylic	element	# types of HYPER, namely HYPO	enumeration + selection
692.	acidification	episodic acidification	process	HYPER *be* classified into # types: HYPO	enumeration + classification
2495.	acidification	episodic acidification	process	HYPER, especially HYPO	selection
1722.	acrylamide	N-alkylacrylamide	element	HYPER such as HYPO	exemplification
1064.	acrylic acid	alkyl acrylate, methacrylate	element	HYPER such as HYPO	exemplification
2904.	active region	swash zone	location	HYPER, such as HYPO	exemplification
1378.	active substance	clay, charcoal, diatomaceous earth	substance	HYPER such as HYPO	exemplification
405.	active volcano	Mount Spur	landform	HYPER, such as HYPO	exemplification
414.	active volcano	Mount Erebus	landform	HYPER, such as HYPO	exemplification

Figure 3: Extract of the hyponymic KP table

As shown in Figure 3, the hyponymic KP table contained the following categories: (i) ID number of the concordance; (ii) hypernym in the concordance; (iii) hyponym(s) in the concordance; (iv) semantic category of the hypernyms/hyponyms; (v) hyponymic KP expressing the generic-specific relation; (vi) type of hyponymic KP. A list of semantic categories and a list of pattern types were also formulated in order to classify and filter the information. As previously mentioned, our research objective was to examine the correlation between hyponymic KPs and the semantic category of concepts. It was thus necessary to create an inventory of semantic categories (Section 4.1).

3.2 Hyponymy subtypes and semantic categories

In the KP study (Section 3.1), the compilation of hypernym-hyponym pairs was performed by filtering KPs, rather than by focusing on semantic categories. However, in the case of hyponymy subtypes, emphasis was placed on selecting different concept types so as to generate a list of hyponymy subtypes that was as comprehensive as possible. Since our previous results seemed to indicate that hyponymy subtypes depended on the nature of the concept (Gil-Berrozpe & Faber, 2016), we wished to confirm this hypothesis by using more fine-grained semantic categories (e.g. *activity, landform, chemical element,* etc.).

It was thus necessary to perform a second compilation of hypernym-hyponym pairs, though this time with a greater focus on semantic categories. For this reason, we extracted 109 hypernyms of concepts belonging to a wide range of semantic categories: 32 natural entities, 32 artificial entities, 21 natural processes, 17 artificial processes, and seven hybrid processes (which could be considered natural or artificial depending on their respective agents or methods). These 109 hypernyms were then analyzed using the default *modifier* word sketch in Sketch Engine. This gave us a set of hyponyms characterized by their modifier (Figure 4).

landform (noun) Alternative PoS: adjective (2)
enviro_normal wsketches freq = 2,912

modifier	1,902	65.32
depositional depositional landforms ,	80	9.51
glacial + glacial landforms	170	9.04
erosional erosional landforms	38	8.55
distinctive distinctive landforms	36	8.34
lateroglacial lateroglacial landforms .	16	8.00
karst of karst landforms	18	7.87
paraglacial paraglacial landforms	16	7.87
glaciotectonic glaciotectonic landforms	14	7.83
glaciofluvial glaciofluvial landforms	14	7.81
periglacial of periglacial landforms	18	7.78
famous . These include famous landforms such as the	16	7.40
solutional	10	7.40
characteristic characteristic landforms	26	7.34
ice-marginal	10	7.31
asymmetrical and other asymmetrical landforms are made when	10	7.26
constructive mountain or other constructive landform built by a	10	7.18

vehicle (noun)
enviro_normal wsketches freq = 8,560

modifier	5,440	63.55
motor + motor vehicles	622	11.52
light-duty + light-duty vehicles and	142	9.67
electric + electric vehicles	192	9.38
hybrid + hybrid vehicles	136	9.21
clean-fuel clean-fuel vehicles	96	9.14
heavy-duty heavy-duty vehicles	72	8.72
fuel + fuel vehicles	142	8.62
underwater underwater vehicles	80	8.53
personal personal vehicles	70	8.30
duty duty vehicles	48	8.11
nonroad or nonroad vehicles	46	8.08
diesel diesel vehicles .	48	8.00
autonomous autonomous underwater vehicle	44	7.98
off-road off-road vehicles	40	7.89
passenger passenger vehicles	36	7.67

Figure 4: *Modifier* word sketches of LANDFORM and VEHICLE

Furthermore, it was necessary to manually select the relevant information in order to avoid matches that were not necessarily terms (e.g. FAMOUS LANDFORM, seen in the *modifier* word sketch of LANDFORM in Figure 4). A total of 1,912 hypernym-hyponym pairs were extracted and inserted in a classification table (Figure 5).

ID	Hypernym [HYPER]	General semantic category	Hyponym [HYPO]	Specific semantic category	Hyponymy subtype
NE10	acid	natural entity	abscisic acid	element	**effect-based hyponymy**
NE02	element	natural entity	abundant element	element	**amount-based hyponymy**
HP02	contamination	hybrid process	accidental contamination	phenomenon	**method-based hypoynymy**
NE10	acid	natural entity	acetic acid	substance	**composition-based hyponymy**
NP11	precipitation	natural process	acid precipitation	phenomenon	**patient-based hyponymy**
NE16	soil	natural entity	acid soil	substance	**composition-based hyponymy**
HP04	reaction	hybrid process	acid-base reaction	process	**agent-based hyponymy**
NE03	compound	natural entity	acidic compound	element	**composition-based hyponymy**
NP19	absorption	natural process	active absorption	process	**method-based hypoynymy**
NE23	dune	natural entity	active dune	mass of matter	**activity-based hyponymy**
AP09	management	artificial process	adaptive management	activity	**method-based hypoynymy**
NP20	radiation	natural process	adaptive radiation	process	**method-based hypoynymy**
HP04	reaction	hybrid process	addition reaction	process	**method-based hypoynymy**
NP08	melting	natural process	adiabatic melting	change of state	**method-based hypoynymy**
NE21	continent	natural entity	adjacent continent	mass of matter	**location-based hyponymy**
NE22	land	natural entity	adjacent land	mass of matter	**location-based hyponymy**

Figure 5: Extract of the hyponymy subtype table

The hyponymy subtype table in Figure 5 has the following categories: (i) ID number of the hypernym; (ii) hypernym; (iii) general semantic category of the hypernym; (iv) hyponym; (v) semantic category of the hyponym; (vi) hyponymy subtype derived from the hypernym-hyponym pair. As in the corpus study, our objective was to explore the correlation between hyponymy subtype and concept type, expressed in the form of semantic categories. For this reason, it was necessary to create an inventory of semantic classes (Section 4.2).

4. Results and discussion

As part of this research, two sets of hypernym-hyponym pairs were analyzed: (i) 3,133 pairs extracted from the corpus with customized hyponymic grammars; (ii) 1,912 pairs extracted from word sketch data using the default *modifier* word sketch. In both cases, concepts were classified in semantic categories. Although most of the semantic categories coincided in both data sets, there were certain categories exclusive to each set.

4.1 Hyponymic KP analysis: general results

Figure 6 shows the distribution of the 3,133 concepts extracted for hyponymic KP analysis. As can be observed, 21 semantic categories were found. (See Appendix A for the description and typical examples of each category.)

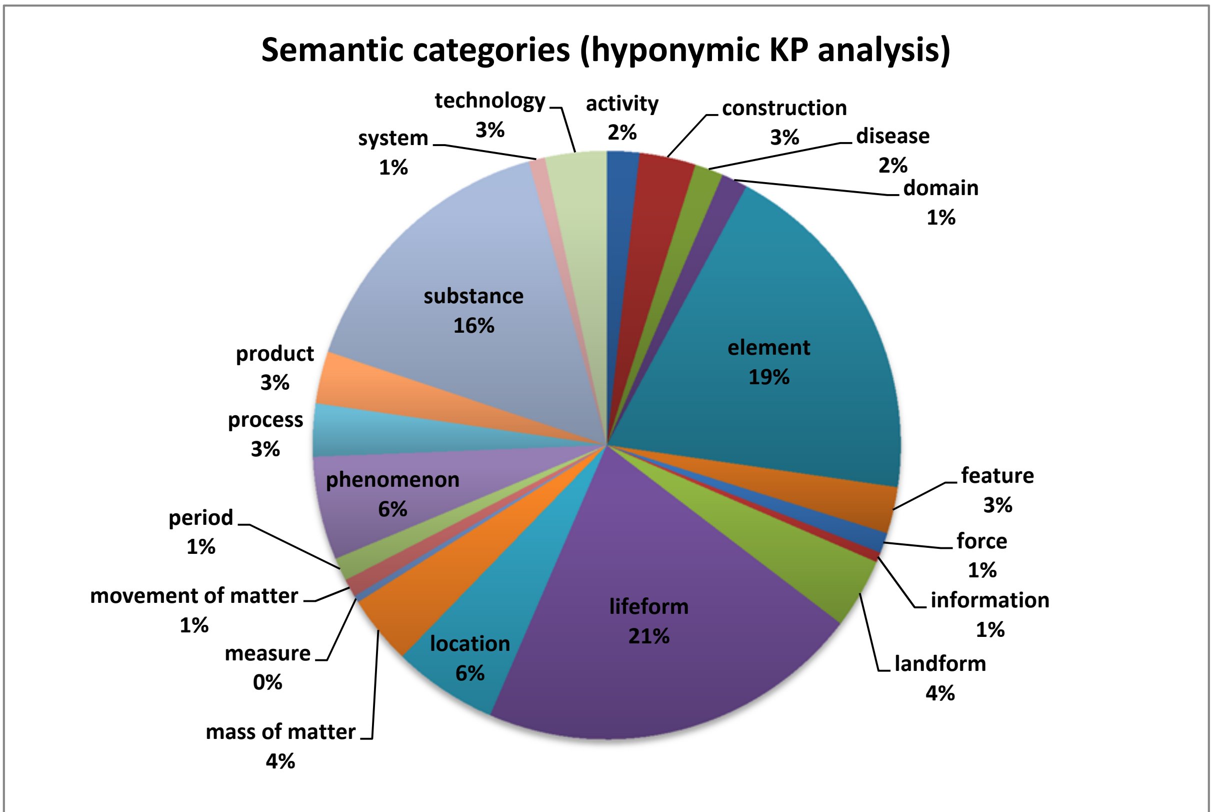

Figure 6: Semantic categories of the concepts of the hyponymic KP analysis

The results of our study showed that the semantic categories of the main concept types were lifeform, chemical element and substance, whose percentages were significantly higher than those of the other categories.

In regard to hyponymic KPs, 125 patterns were identified. KPs that expressed hyponymy in a similar way were placed in the same category. Figure 7 shows the distribution of these 125 patterns in 10 categories. (See Appendix B for a description of each knowledge pattern with examples.)

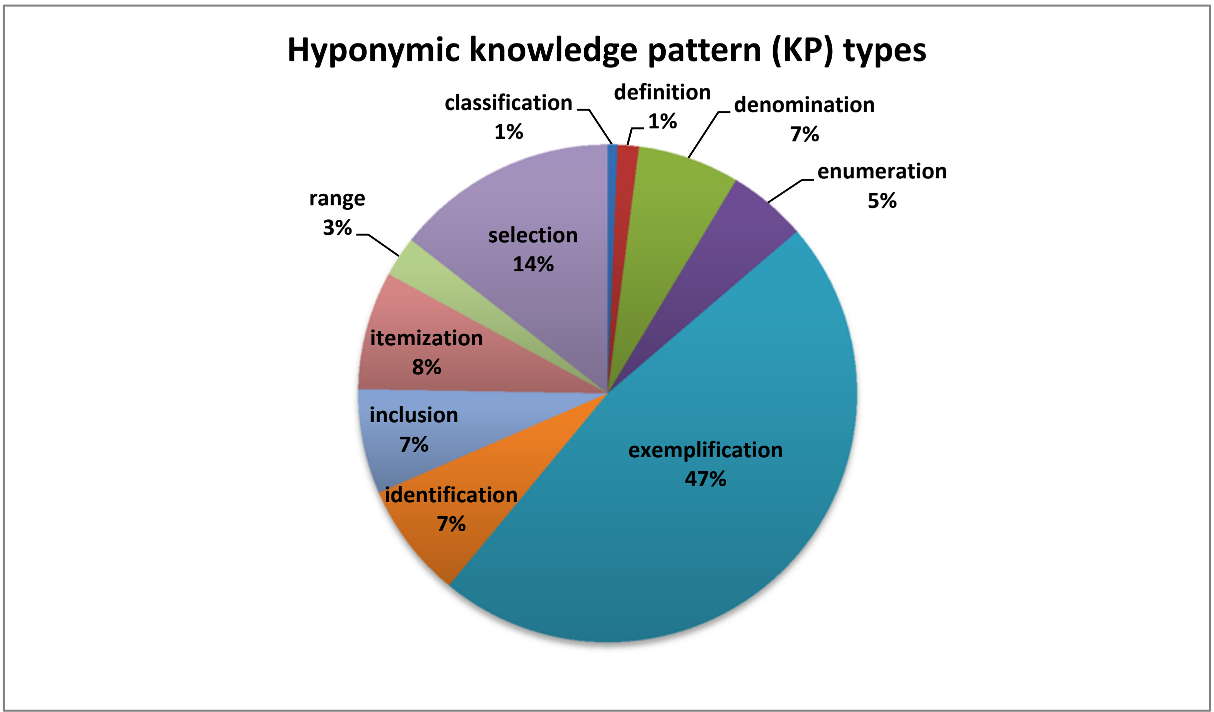

Figure 7: Hyponymic knowledge patterns

As reflected in our results, the most frequent hyponymic pattern types were exemplification KPs, selection KPs, and itemization KPs, though patterns expressing any sort of exemplification were clearly the most predominant.

4.1.1 Correlations between hyponymic KPs and semantic categories

Exemplification KPs (Figure 8), by far the most frequent pattern, comprised almost half of the sample analyzed. Because of the quantity of information in these patterns, they were typical of the most common semantic categories, namely: chemical element, lifeform, and substance. The second most significant group of categories included location, phenomenon, landform, and construction. The other semantic categories were found in significantly fewer concordances.

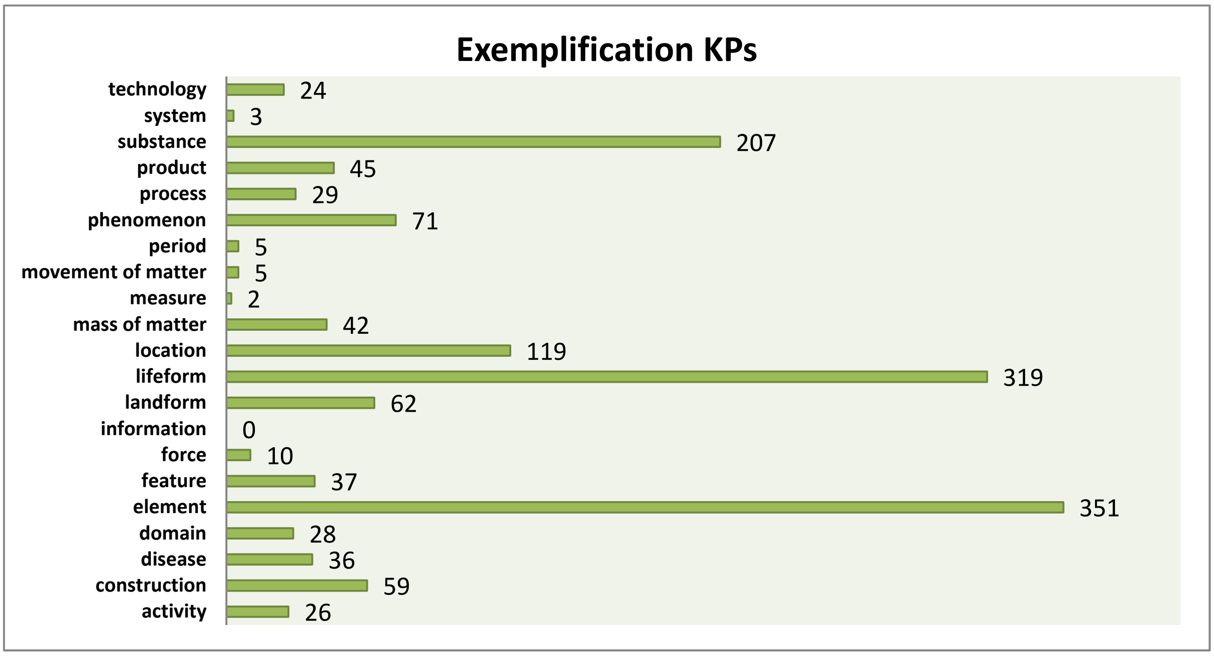

Figure 8: Exemplification KPs per semantic category

Since exemplification KPs were the most common, the only conclusion that can be derived is that the occurrences of exemplification KPs per semantic category are proportional to the ratios of semantic categories shown in Figure 6.

As for selection KPs (Figure 9), itemization KPs (Figure 10), and inclusion KPs (Figure 11), lifeform, chemical element, and substance were also the most prominent semantic categories.

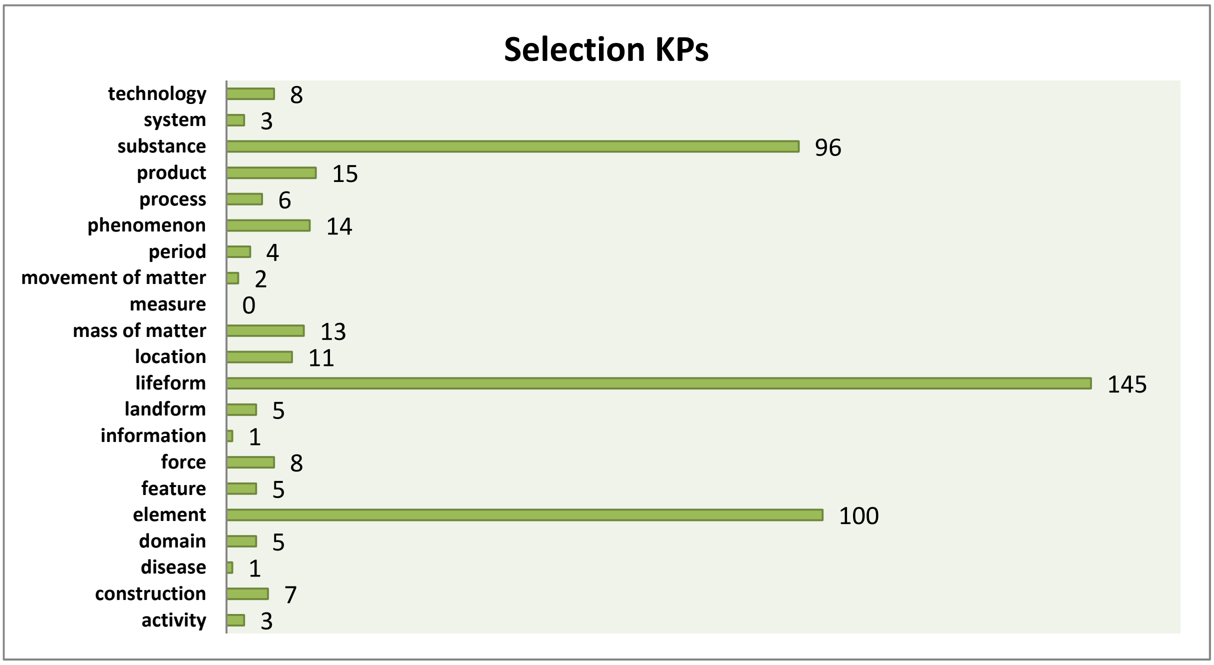

Figure 9: Selection KPs per semantic category

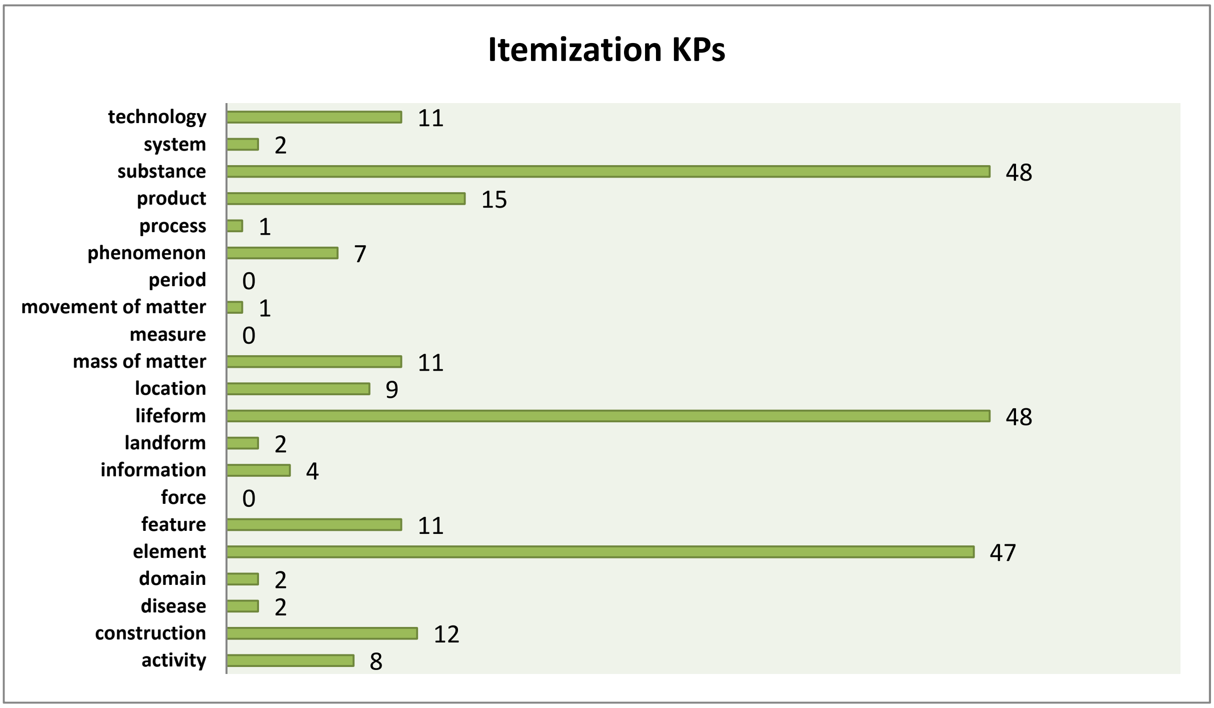

Figure 10: Itemization KPs per semantic category

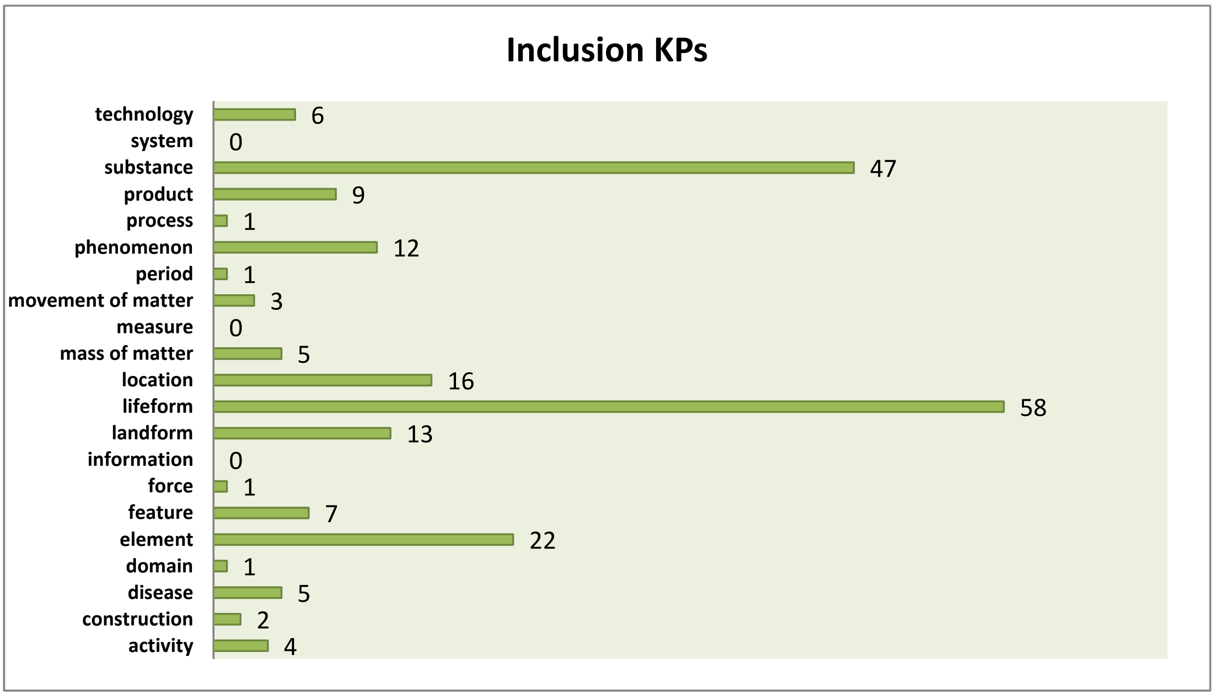

Figure 11: Inclusion KPs per semantic category

The predominance of these patterns could be a matter of statistics, since these concepts are the most frequent in the English EcoLexicon corpus. However, another possibility is that this phenomenon is related in some way to discourse type and function since most of the texts in the corpus are research articles, textbooks, and encyclopedias, whose functions are to facilitate the acquisition of specialized environmental knowledge.

With regard to identification KPs (Figure 12) and denomination KPs (Figure 13), the category of phenomenon held the second position, only surpassed by chemical element, and followed by lifeform and substance. In addition, the categories of process and technology also had a significant presence. As in the previous cases, this showed that identification KPs and denomination KPs are also activated by semantic categories in relation to the ratios shown in Figure 6. However, the significantly greater frequency of phenomenon, process and technology also indicates that these hyponymic KPs could be related to complex concepts that need an identifying or denominating structure (HYPO is a HYPER, a type of HYPER is a HYPO, types of HYPER are called HYPO) in order to better explain them.

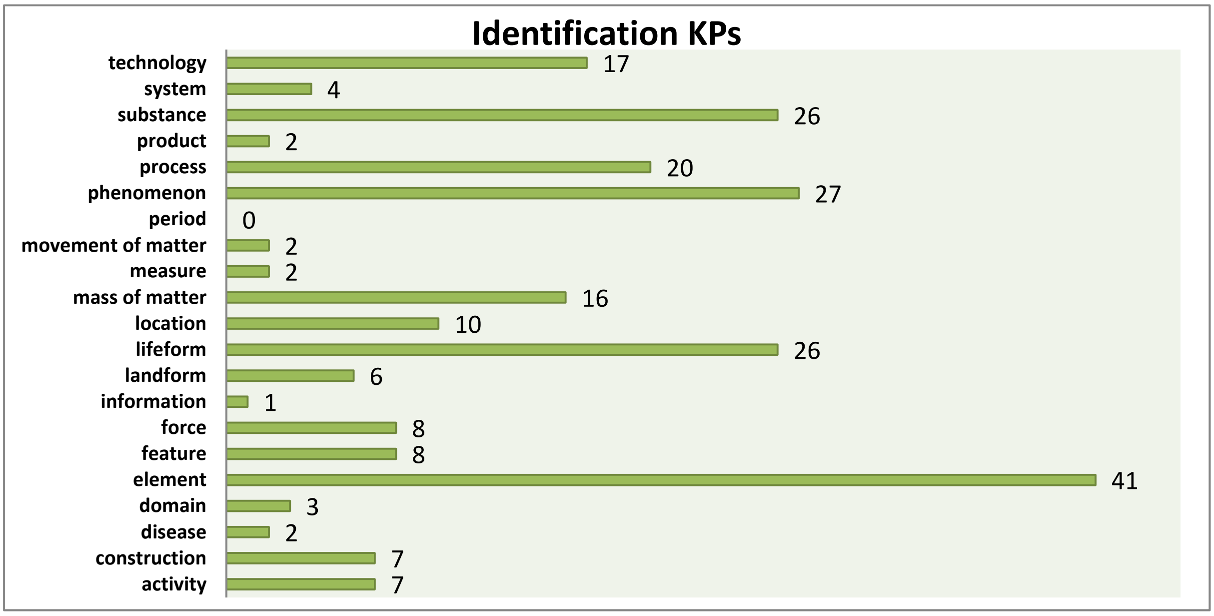

Figure 12: Identification KPs per semantic category

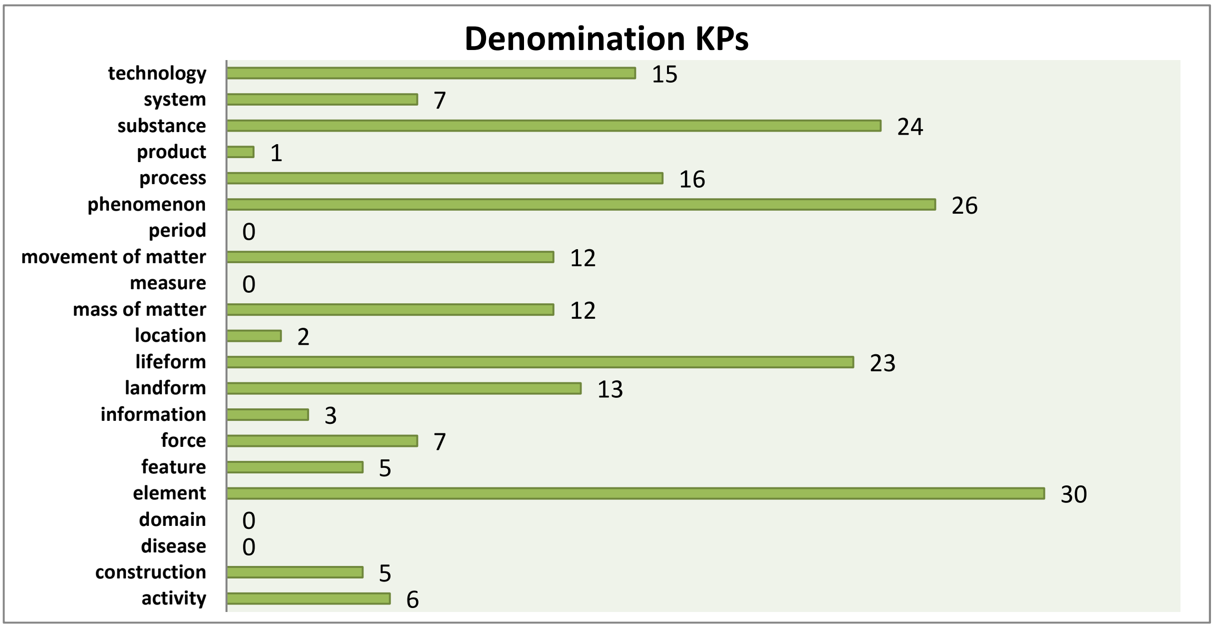

Figure 13: Denomination KPs per semantic category

This could also be true of definition KPs (Figure 14), where the categories of technology and phenomenon share second position, after substance. Once again, the KP expressions in this category specifically define a concept (HYPO: a HYPER, HYPO: a type of HYPER) in terms of its superordinate.

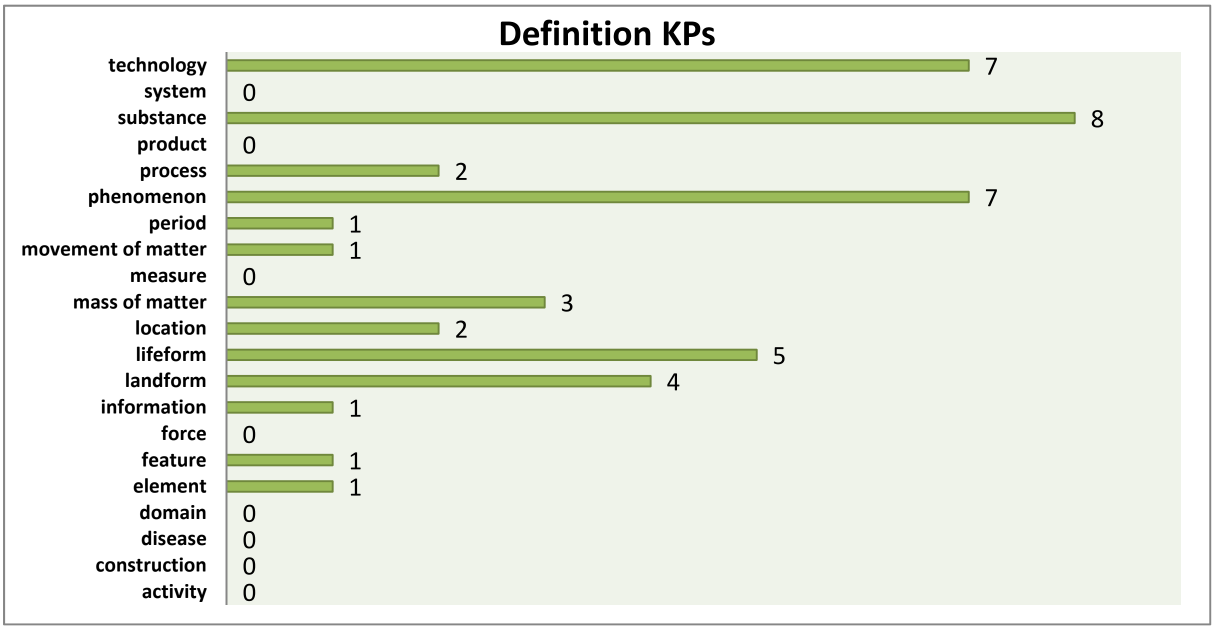

Figure 14: Definition KPs per semantic category

As for range KPs (Figure 15), a different semantic category held first position. The nature of this hyponymic KP makes it ideal for expressing time periods, scales, and degrees (HYPER ranging from HYPO to HYPO). Not surprisingly, the semantic category, measure, which had little or no relevance in the other patterns, frequently occurred in range KPs.

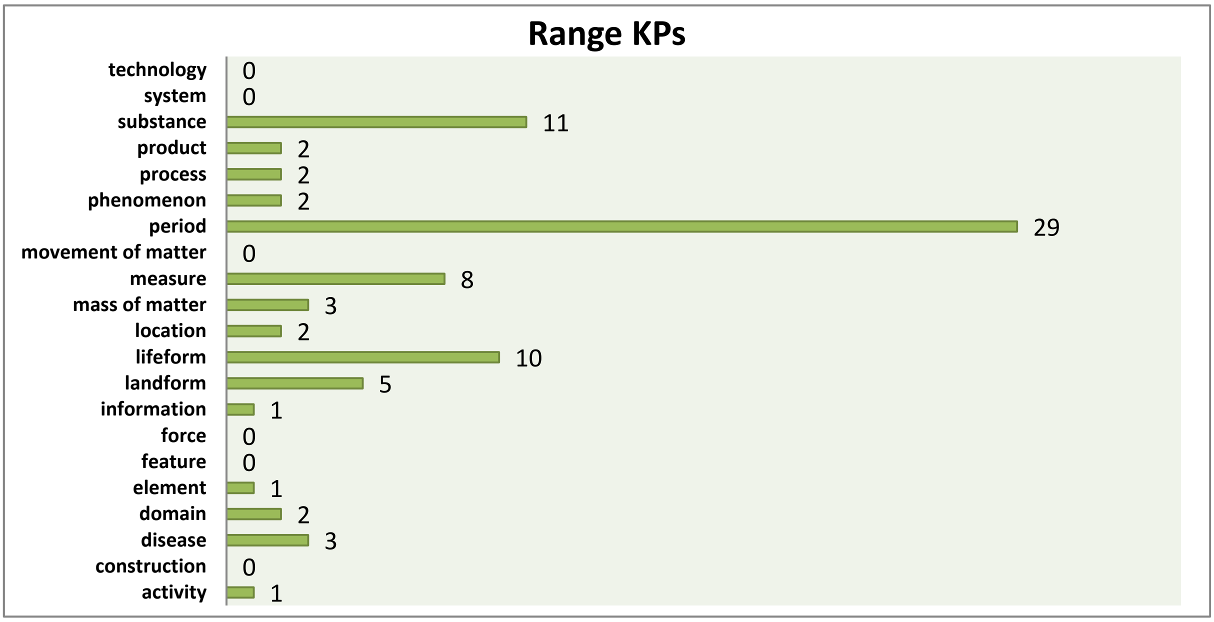

Figure 15: Range KPs per semantic category

Finally, in the case of enumeration KPs (Figure 16) and classification KPs (Figure 17), it was not possible to extract any specific correlation pattern. Our results showed that enumeration KPs, in the same way as exemplification KPs, were applicable to any concept type. Furthermore, the data for classification KPs was insufficient to draw any conclusions.

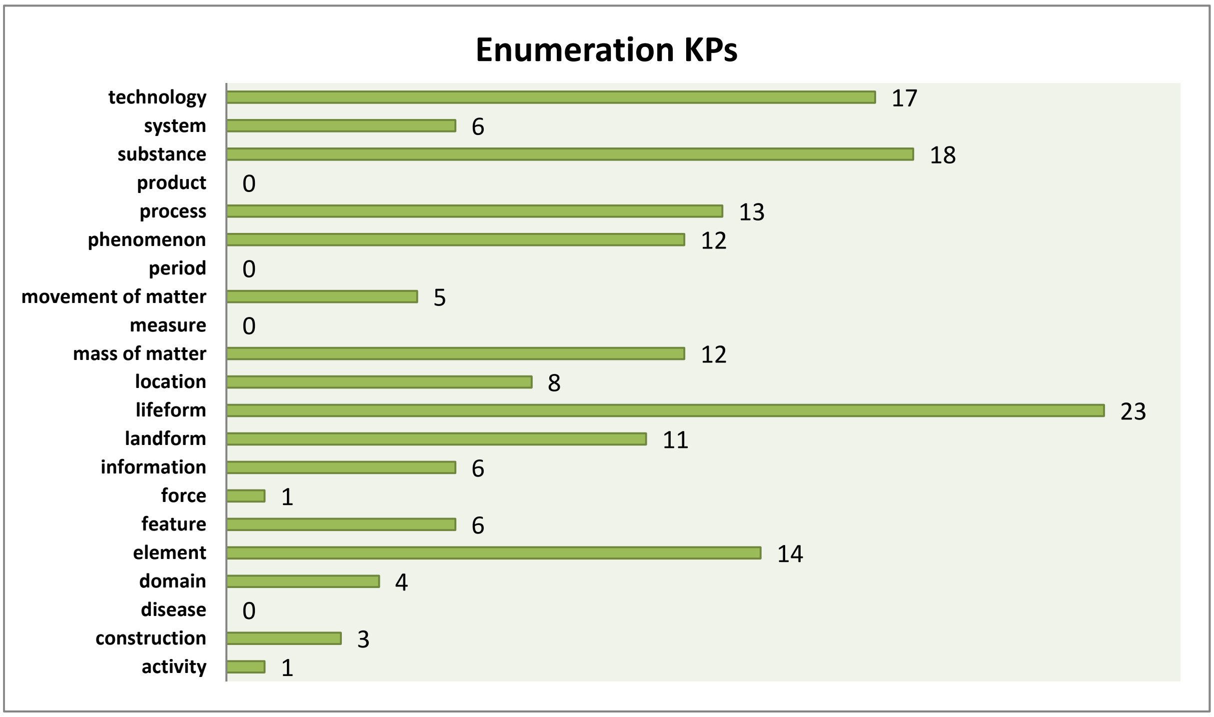

Figure 16: Enumeration KPs per semantic category

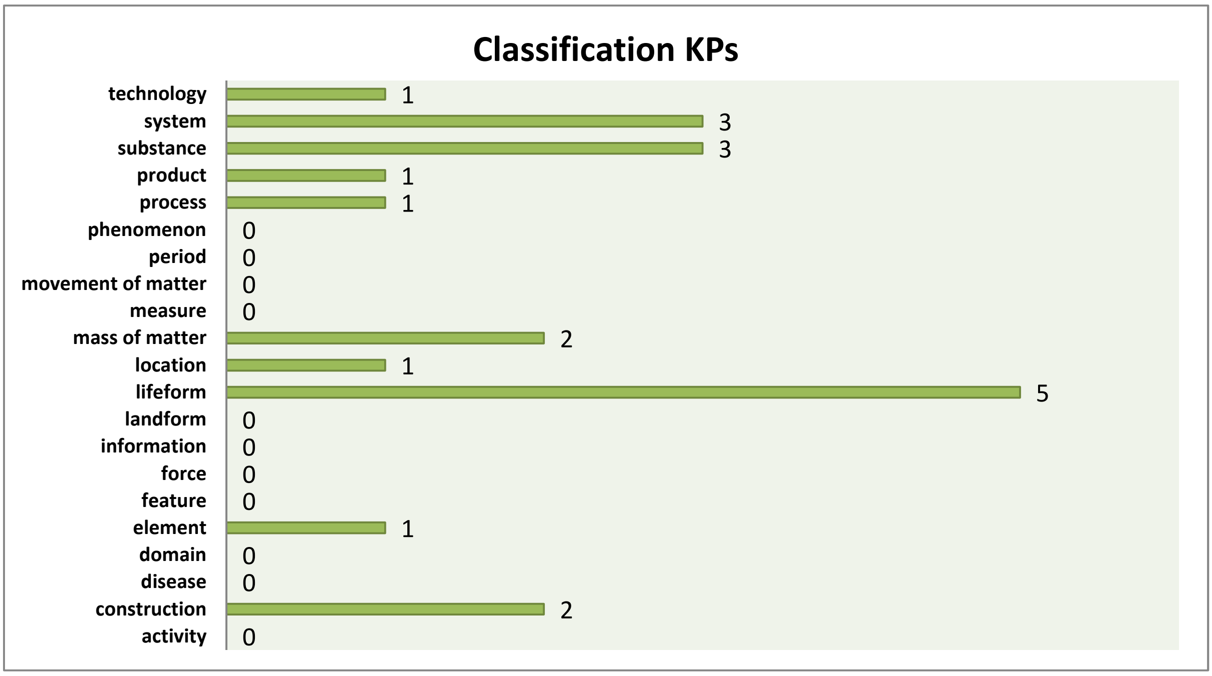

Figure 17: Classification KPs per semantic category

4.2 Hyponymy subtypes analysis: general results

Figure 18 shows the distribution of the 1,912 hyponyms in 13 semantic categories.

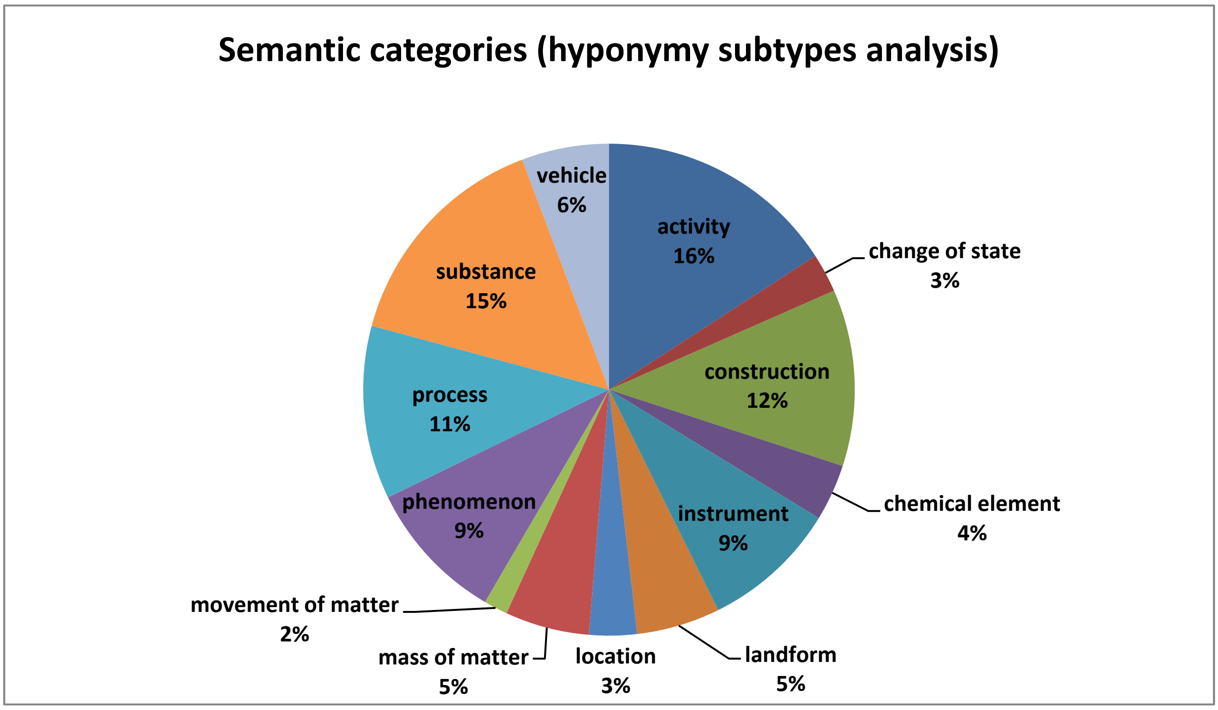

Figure 18: Semantic categories of the concepts of the hyponymy subtypes analysis

Although most of the semantic categories identified during this analysis coincide with those of the hyponymic KP analysis, the categories of *disease, domain, feature, force, information, lifeform, measure, period, product, system* and *technology* do not appear. This was due to the manual selection process. On the other hand, because of the higher frequency of other concept types, it was possible to identify three more semantic categories that are exclusive to the hyponymy subtype analysis: *instrument*, *vehicle*, and *change of state* (Appendix A).

The decomposition of the generic-specific relation was based on common features in the cases analyzed. This led to the identification of 32 different subtypes in the 1,912 hypernym-hyponym pairs (Figure 19). Appendix C describes and exemplifies the full inventory of hyponymy subtypes. In this inventory, a distinction can be made between relational hyponymy subtypes (those specifying a relation between the components of hyponym-hypernym pairs) and attributional hyponymy subtypes (those specifying an intrinsic feature of the hyponym).

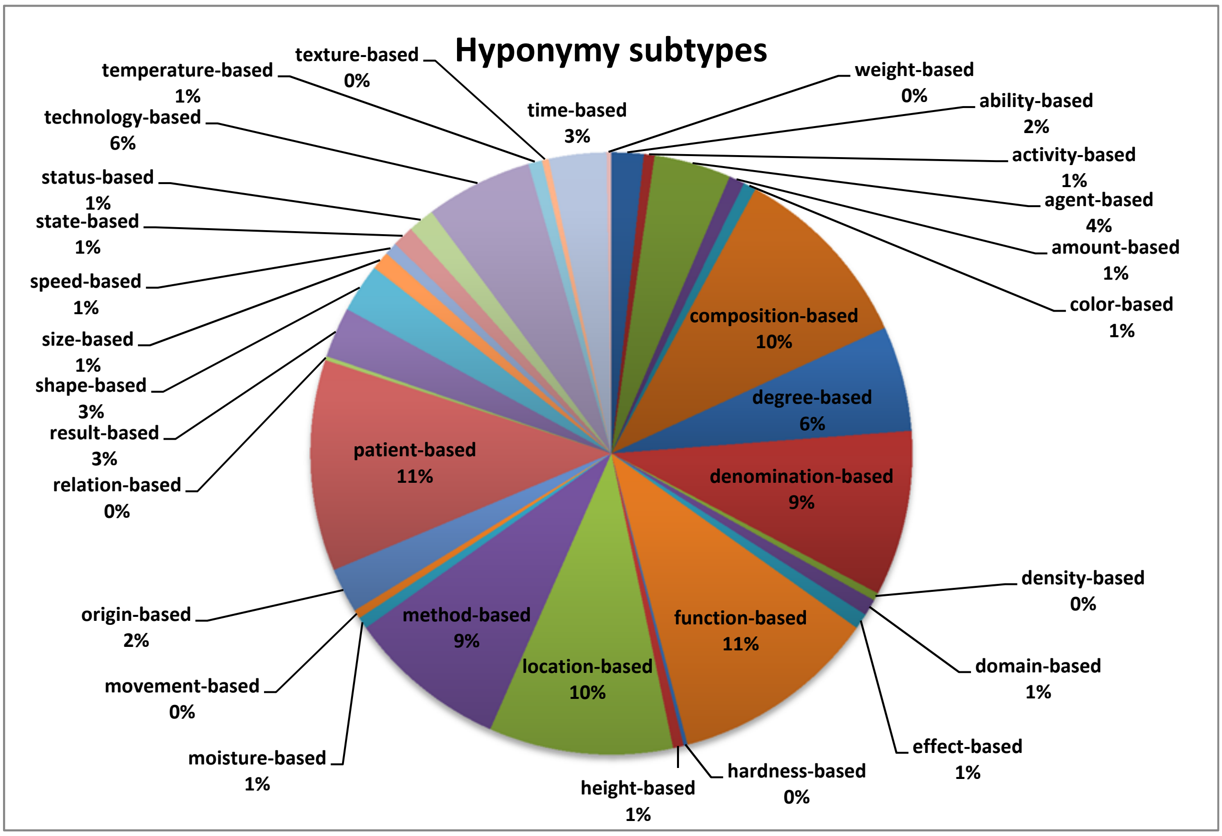

Figure 19: Hyponymy subtypes

As can be observed in Figure 19, the most frequently activated hyponymy subtypes were relational, particularly *patient-based*, *function-based*, *composition-based* and *location-based* hyponymy. On the contrary, attributional hyponymy subtypes (such as *degree-based*, *shape-based*, *ability-based* or *size-based*) were found to be less representative. This seems to indicate that when environmental knowledge is categorized into subtypes, there is a greater emphasis on how concepts interact with each other, rather than on the intrinsic characteristics of individual concepts.

4.2.1 Correlations between hyponymy subtypes and semantic categories

For the sake of conciseness, this section focuses on the 12 most recurrent hyponymy subtypes, derived from 1,582 hypernym-hyponym pairs (83% of the sample). These are *patient-based*, *function-based*, *composition-based*, *location-based*, *denomination-based*, *method-based*, *technology-based*, *degree-based*, *agent-based*, *time-based*, *result-based*, and *shape-based* hyponymy.

In both *patient-based* hyponymy (Figure 20) and *method-based* hyponymy (Figure 21), there was a predominance of the categories of activity, process, phenomenon, and change of state. There were no entity-related semantic categories because these two subtypes of hyponymy are exclusive to process-related semantic categories.

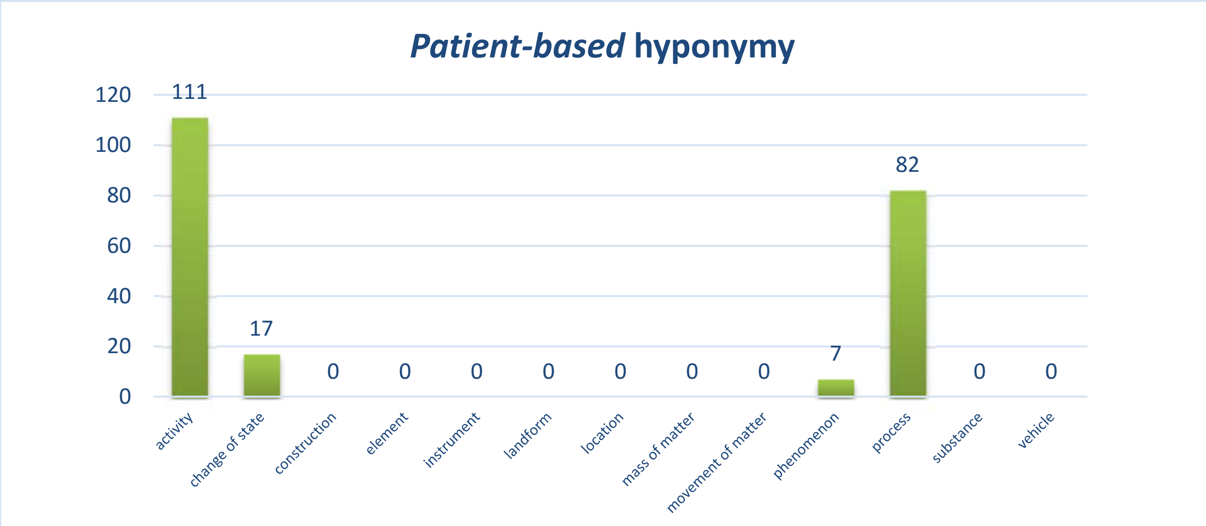

Figure 20: Patient-based hyponymy subtypes per semantic category

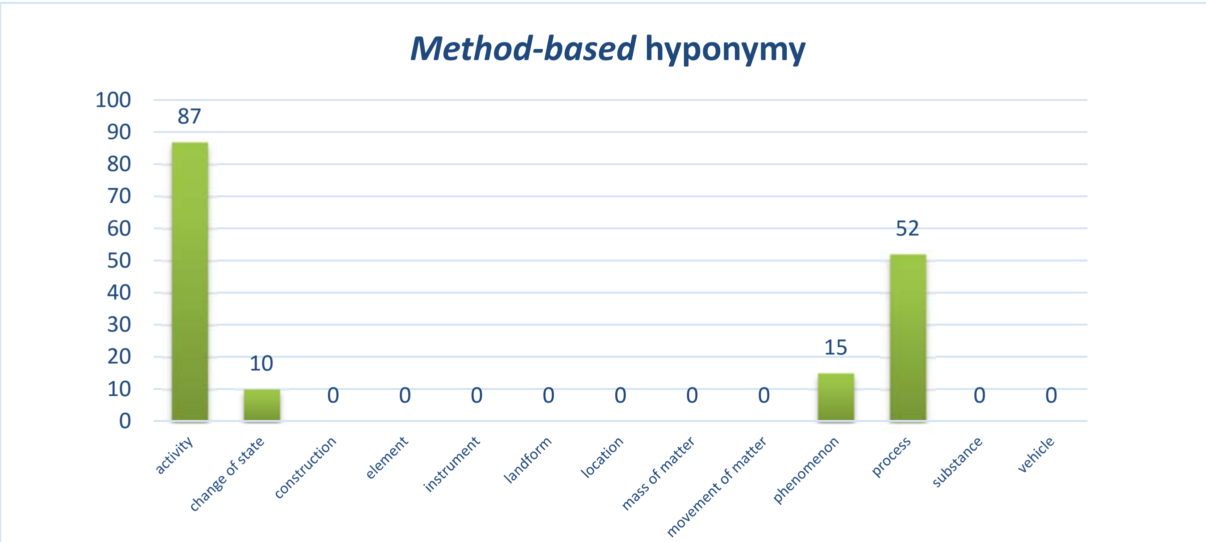

Figure 21: Method-based hyponymy subtypes per semantic category

As can be observed, the most frequent semantic categories were found to be activity and process, which are mostly composed of artificial or deliberate actions and processes. This sharply contrasted with the categories of phenomenon and change of state, which were mostly composed of natural processes. This could indicate that patient and method are what distinguish artificial processes from natural processes, since a natural change is not purposeful or deliberate.

As for *agent-based* hyponymy (Figure 22) and *result-based* hyponymy (Figure 23), once again most of the examples refer to process-related semantic categories, namely activity, process, and phenomenon.

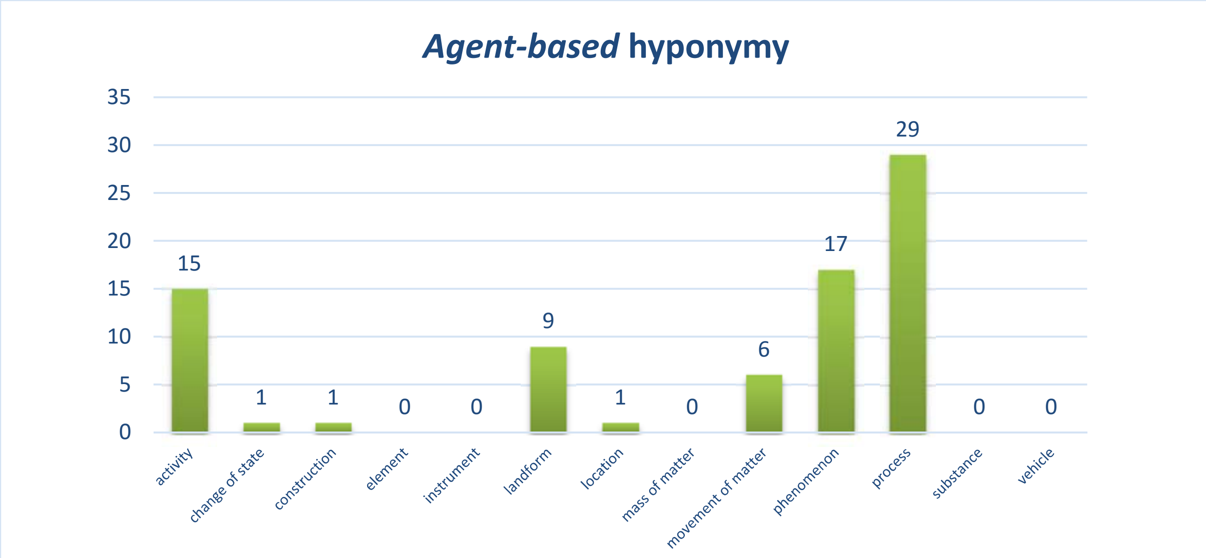

Figure 22: Agent-based hyponymy subtypes per semantic category

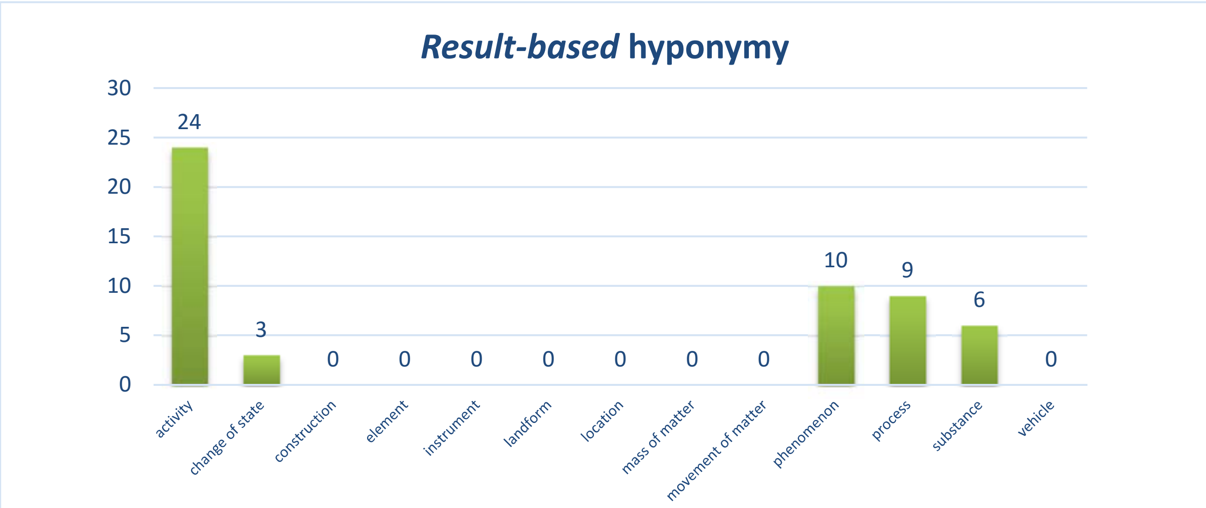

Figure 23: Result-based hyponymy subtypes per semantic category

Interestingly, these hyponymy subtypes also include two entity-related categories: (i) landform in the case of *agent-based* hyponymy, since there are some landforms characterized by the agent that has created them (e.g. GLACIAL LANDFORM, FLUVIAL LANDFORM, VOLCANIC ISLAND); (ii) substance in the case of *result-based* hyponymy, since substances can sometimes be characterized as the result of a process (e.g. DEGRADATION PRODUCT, OXIDATION PRODUCT, FISSION PRODUCT).

Similarly, *degree-based* hyponymy (Figure 24) is also mostly exclusive to process-related semantic categories, such as phenomenon, activity, process, and change of state. Furthermore, and in contrast to the previous results, the category of phenomenon is mostly characterized by degree (e.g. CATACLYSMIC ERUPTION, LOW-MAGNITUDE EARTHQUAKE, KILLER TORNADO, etc.).

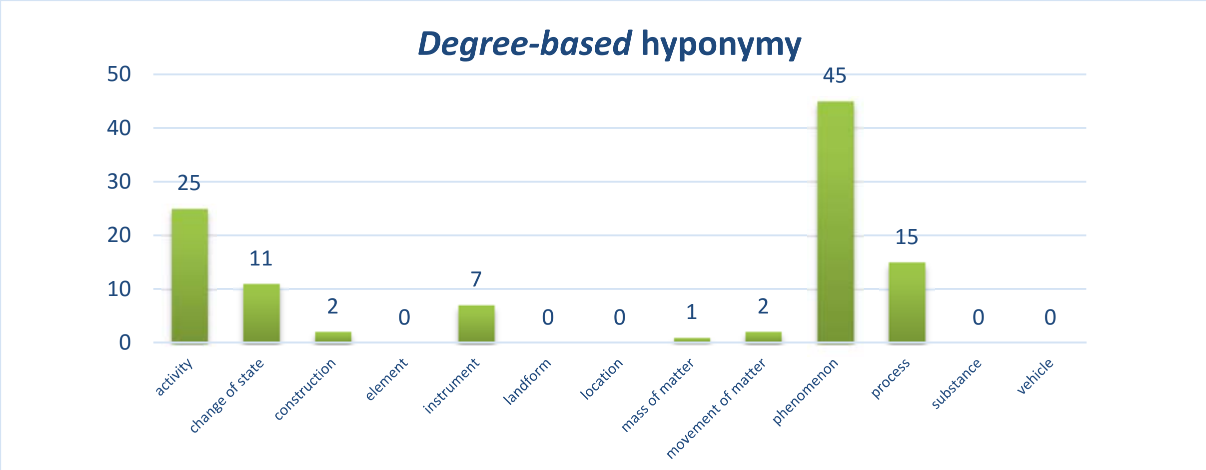

Figure 24: Degree-based hyponymy subtypes per semantic category

Composition-based hyponymy (Figure 25) shows that the most recurrent semantic categories are those involving natural entities, namely substance and chemical element. These are followed by the category of construction, which is composed of artificial entities that can be characterized by their components or their material (e.g. WOODEN BUILDING, RUBBLE MOUND BREAKWATER, CONCRETE DAM, etc.).

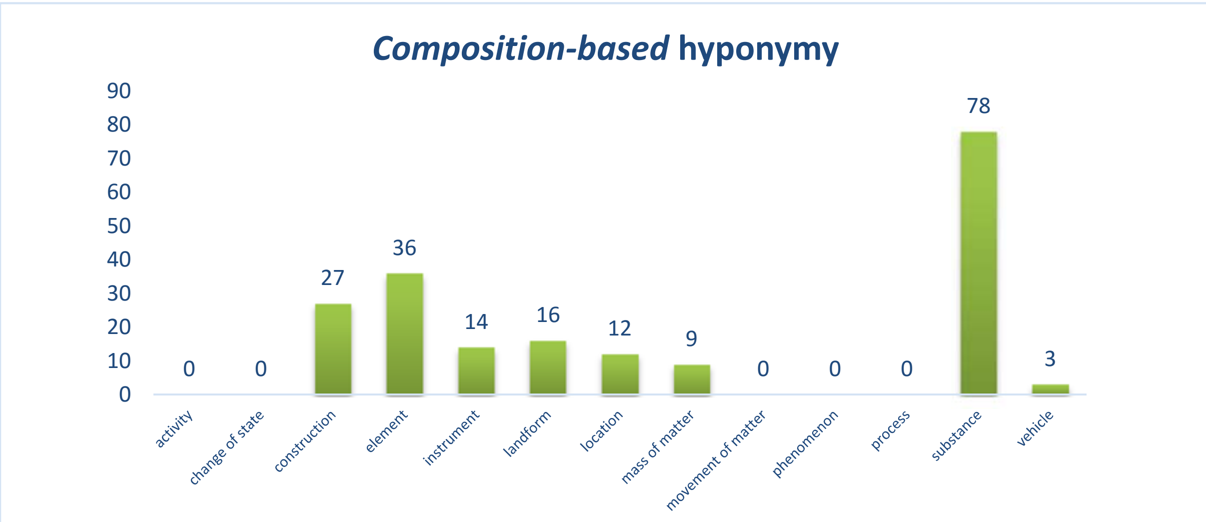

Figure 25: Composition-based hyponymy subtypes per semantic category

Location-based hyponymy (Figure 26) typically occurs with entity-related categories such as substance, construction, mass of matter, and landform. However, the category of phenomenon is also significant because natural processes are also characterized by the location where they occur (e.g. SUBMARINE EARTHQUAKE, MOUNTAIN CYCLOGENESIS, FOREST FIRE, etc.).

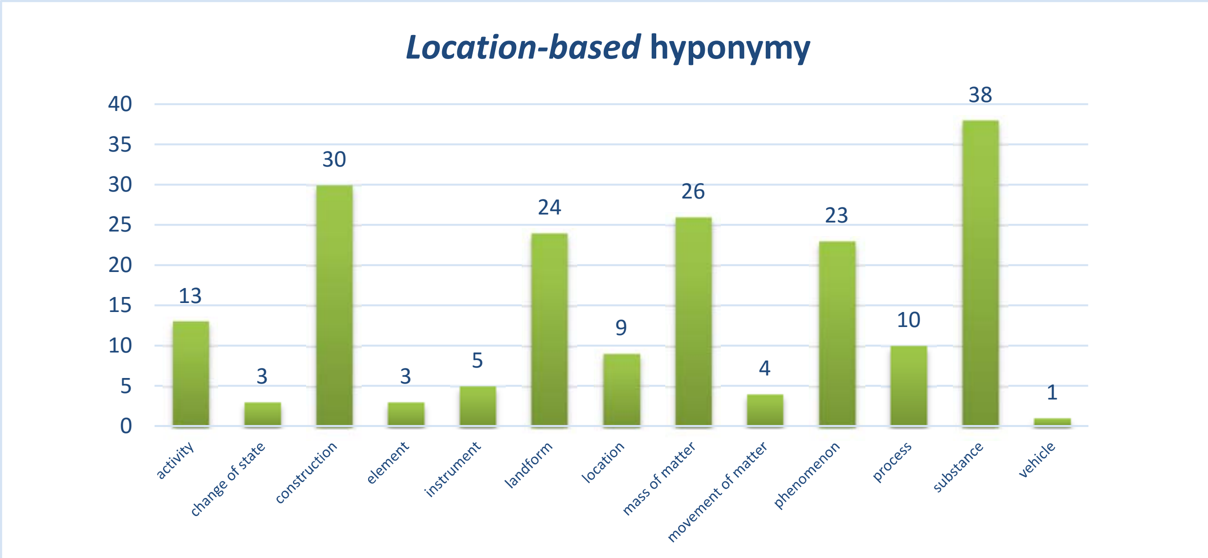

Figure 26: Location-based hyponymy subtypes per semantic category

In the case of *function-based* hyponymy (Figure 27) and *technology-based* hyponymy (Figure 28), the most frequently-activated semantic categories were those pertaining to artificial entities: instrument, vehicle, and construction. However, rather surprisingly, construction, which is the most recurrent category in *function-based* hyponymy, appeared less frequently in relation to *technology-based* hyponymy. This seems to indicate that the identifying feature of a construction is its purpose (e.g. PROCESSING FACILITY, PROTECTION STRUCTURE, LANDING DOCK), rather than its technology (e.g. NUCLEAR FACILITY, COAL-FIRED STATION, ORGANIC FARM).

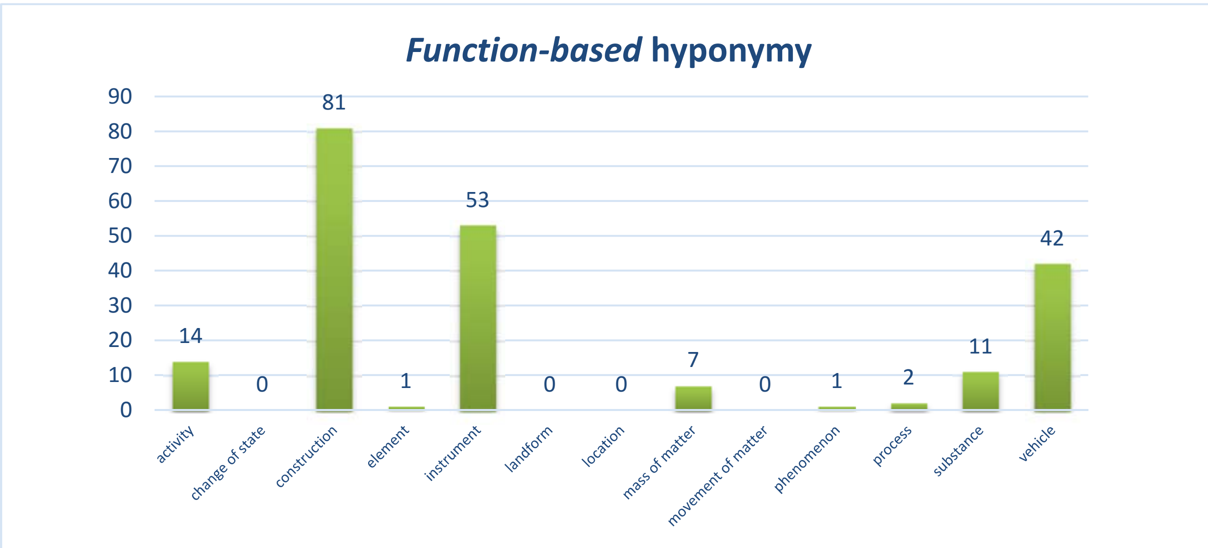

Figure 27: Function-based hyponymy subtypes per semantic category

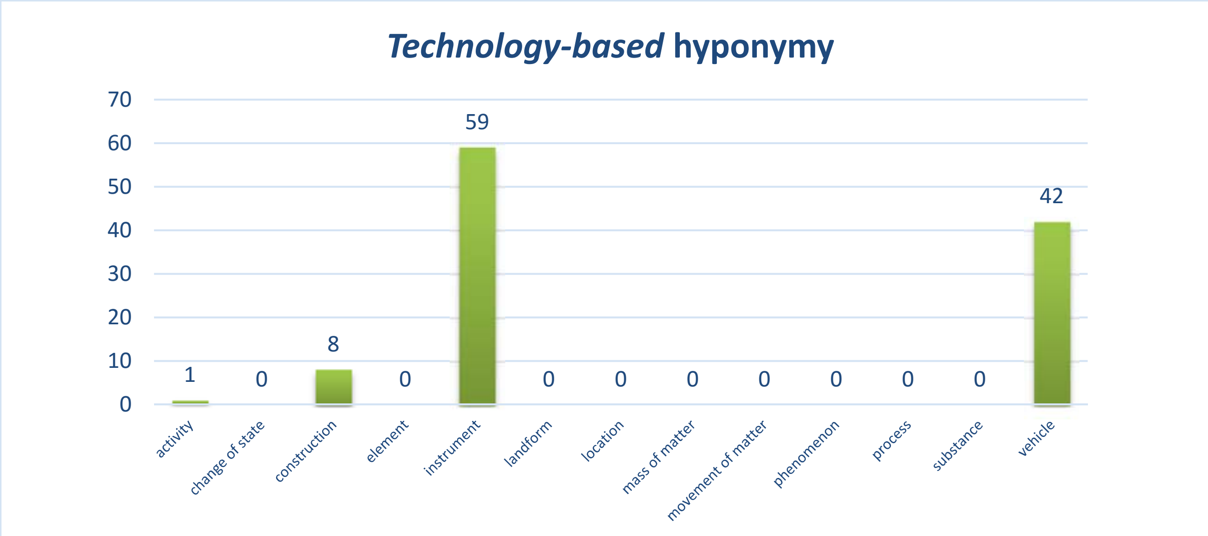

Figure 28: Technology-based hyponymy subtypes per semantic category

Regarding *denomination-based* hyponymy (Figure 29), almost all of the semantic categories activated were entities: landform, location, mass of matter, construction, and instrument. However, the category of phenomenon was in second position along with location, since certain meteorological events tend to receive denominations specifying the location where they occur (e.g. SUMATRA EARTHQUAKE, OKLAHOMA TORNADO, SAHEL DROUGHT).

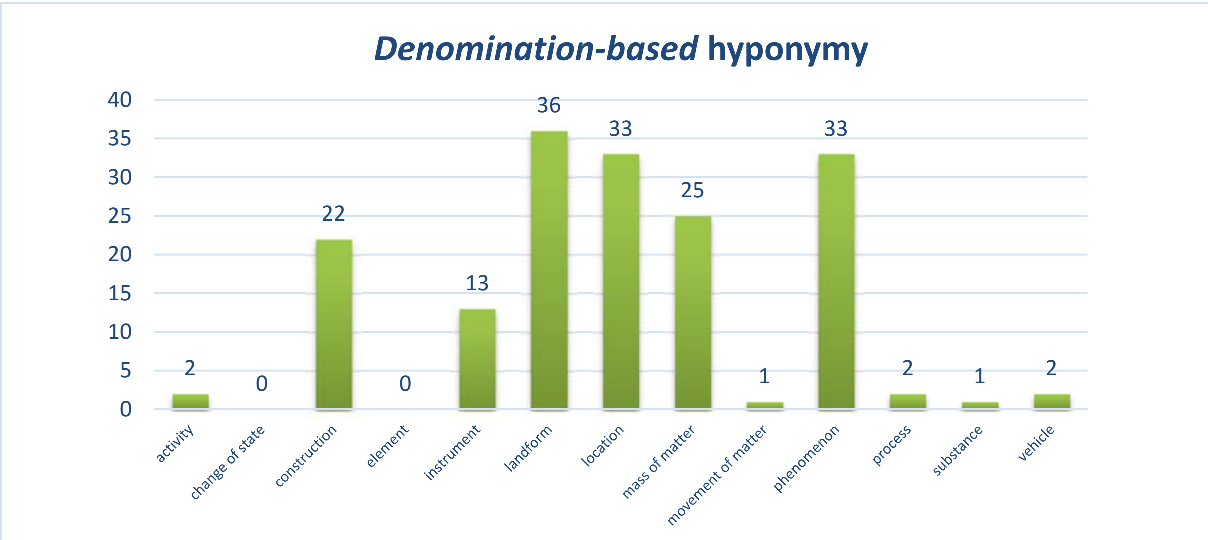

Figure 29: Denomination-based hyponymy subtypes per semantic category

Time-based hyponymy (Figure 30) was related to natural semantic categories, which were both processes (phenomenon and movement of matter) and entities (substance and mass of matter). In fact, time is also a natural factor that affects the environmental domain and phenomena (e.g. SUMMER PRECIPITATION, LATE-SEASON HURRICANE, PERIODIC DROUGHT). However, it rarely occurs with artificial concepts.

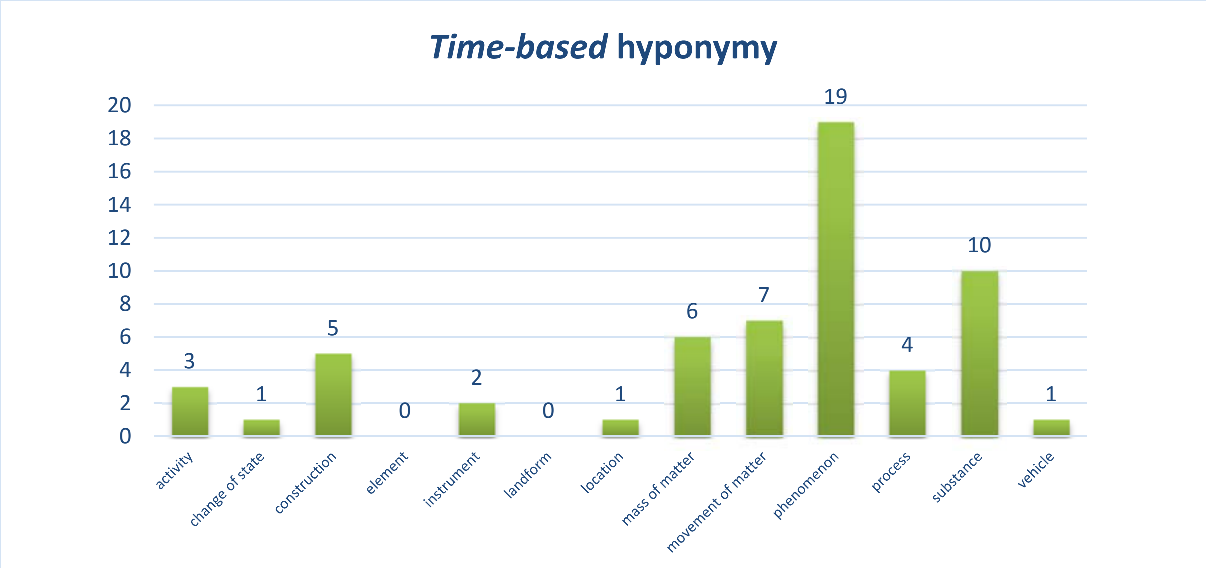

Figure 30: Time-based hyponymy subtypes per semantic category

Finally, with regard to *shape-based* hyponymy (Figure 31), the most recurrent semantic categories were the following artificial and natural entities: construction, landform, and mass of matter. Interestingly, shape occurred most frequently in the case of large formations (e.g. STAR DUNE, RING DIKE, VERTICAL BREAKWATER) than in the case of smaller formations or entities. Furthermore, two process-related semantic categories, movement of matter and phenomenon, are also registered in the table. They include concepts such as WEDGE TORNADO or CROWN FIRE, also characterized by the physical shape acquired by those processes.

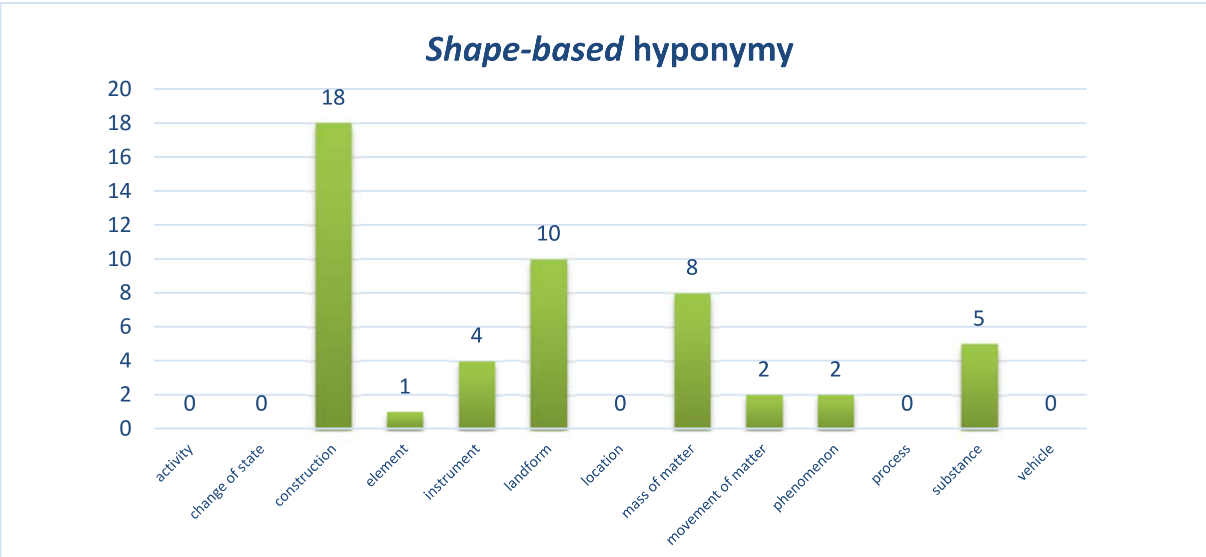

Figure 31: Shape-based hyponymy subtypes per semantic category

5. Conclusion

Hyponymy is a complex semantic relation that can be studied by analyzing concept hierarchies. The results obtained showed that the semantic category of concepts constrained their occurrence in different hyponymy subtypes. By analyzing and classifying hyponymic knowledge patterns and hyponymy subtypes, this study highlights the importance of accounting for semantic categories in the study of the generic-specific relation.

Our results showed that certain KPs (i.e. *exemplification*, *selection*, *itemization*, and *inclusion*) were linked to semantic categories that are the basis of scientific classifications (lifeform and chemical element). Furthermore, other KPs (*identification*, *denomination*, and *definition*) were found to have a more explanatory structure, and were thus most frequently linked to more complex semantic categories involving various participants (phenomenon, process, and technology). They thus invited a more detailed description and/or explanation to facilitate reader understanding. *Range* KPs were mostly associated with time period and measure since these categories are generally composed of values that are characterized by the space/distance between them in terms of time, space, intensity, etc.

The analysis of hyponymy showed that certain subtypes (*agent-based*, *patient-based*, *result-based*, *method-based*, and *degree-based*) closely correlated with process-related semantic categories (activity, phenomenon, process, and change of state). On the other hand, other hyponymy subtypes (*composition-based*, *technology-based*, and *function-based*) were directly linked to entity-related semantic categories (substance, landform, construction, and instruments). In addition, a distinction was made between natural and artificial concepts.

These results open the door to further studies on hyponymy not only in the environmental domain, but also in regard to specialized knowledge in general. In future research, we plan to analyze the whole English EcoLexicon corpus after a previous revision of the customized hyponymic word sketch grammars in order to reduce repetitions and false positives. Regarding hyponymy subtypes, another interesting feature to be explored in future work is the relation between certain subtypes identified (such as *composition-based*, *function-based*, or *origin-based*) and Pustejovsky's (1995) *qualia* structure (with formal, constitutive, telic, and agentive roles).

It would also be necessary to study the distinction between relational and attributional hyponymy subtypes. Another phenomenon to be explored is the correlation between hyponymic KPs and hyponymy subtypes. All of this information related to hyponymy refinement will make it possible to specify a more accurate set of hyponymic relations in the environmental domain.

6. Acknowledgements

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

7. References

Agbago, A. & Barrière, C. (2005). Corpus Construction for Terminology. *Proceedings of the Corpus Linguistics 2005 Conference*, pp. 1–14. Birmingham, United Kingdom.

Barrière, C. (2004a). Knowledge-rich Contexts Discovery. *Proceedings of the 17th Canadian Conference on Artificial Intelligence (AI'2004)*, pp. 187–201. London (Ontario), Canada.

Barrière, C. (2004b). Building a Concept Hierarchy from Corpus Analysis. *Terminology*, 10(2), pp. 241–263.

Bielinskiene, A., Boizou, L., Kovalevskaite, J., & Utka, A. (2012). Towards the Automatic Extraction of Term-defining Contexts in Lithuanian. In A. Tavast, K. Muischnek & M. Koit (Eds.) *Human Language Technologies: The Baltic Perspective*, pp. 18–26. Amsterdam/Berlin/Tokyo/Washington DC: IOS Press.

Cabré, M.T. (1999). *La terminología: representación y comunicación.* Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

Condamines, A. (2002). Corpus Analysis and Conceptual Relation Patterns. *Terminology*, 8(1), pp. 141–162.

Cruse, D.A. (2002). Hyponymy and its Varieties. In R. Green, C.A. Bean, & S.H. Myaeng, (eds.) *The Semantics of Relationships: An Interdisciplinary Perspective*, pp. 3–22. Dordrecht/Boston/London: Kluwer Academic Publishers.

Faber, P. (2009). The Cognitive Shift in Terminology and Specialized Translation. *Monografías de Traducción e Interpretación (MonTI)*, 1, pp. 107–134. Valencia: Universitat de València.

Faber, P. (2015). Frames as a Framework for Terminology. In H.J. Kockaert & F. Steurs (eds.) *Handbook of Terminology*, 1, pp. 14–33. Amsterdam/Philadelphia: John Benjamins.

Faber, P. (ed.) (2012). *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin/Boston: De Gruyter Mouton.

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

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*, pp. 73–80. Portorož, Slovenia.

Gil-Berrozpe, J.C. & Faber, P. (2016). Refining Hyponymy in a Terminological Knowledge Base. *Proceedings of the 2nd Joint Workshop on Language and Ontology (LangOnto2) & Terminology and Knowledge Structures (TermiKS) at the 10th edition of the Language Resources and Evaluation Conference (LREC 2016)*, pp. 8–15. Portorož, Slovenia.

Gil-Berrozpe, J.C., León-Araúz, P., & Faber, P. (in press). Subtypes of Hyponymy in the Environmental Domain: Entities and Processes. *Proceedings of the 10th International Conference on Terminology & Ontology: Theories and Applications (TOTh 2016).* Chambéry, France.

Girju, R., Badulescu, A., & Moldovan, D. (2003). Learning Semantic Constraints for the Automatic Discovery of Part-Whole Relations. *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 1–8.

Jacquemin, C. & Bourigault, D. (2005). Term Extraction and Automatic Indexing. In R. Mitkov (ed.) *The Oxford Handbook of Computational Linguistics.* Oxford: Oxford University Press.

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

León-Araúz, P. (2014). Semantic Relations and Local Grammars for the Environment. In S. Joeva, S. Mesfar & M. Silberztein (eds.), *Formalising Natural Languages with NooJ 2013*, pp. 87–102. Newcastle-upon-Tyne: Cambridge Scholars Publishing.

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

Marshman, E. (2002). The Cause Relation in Biopharmaceutical Texts: Some English Knowledge Patterns. *Proceedings of Terminology and Knowledge Engineering (TKE 2002)*, pp. 89–94. Nancy, France.

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*, pp. 279–302. Amsterdam/Philadelphia: John Benjamins.

Murphy, M.L. (2003). *Semantic Relations and the Lexicon: Antonymy, Synonymy and Other Paradigms.* Cambridge: Cambridge University Press.

Murphy, M.L. (2006). Hyponymy and Hyperonymy. In K. Brown (ed.) *Encyclopedia of Language and Linguistics*, 1, pp. 446–448. New York: Elsevier.

Pustejovsky, J. (1995). *The Generative Lexicon.* Cambridge, MA: MIT Press.

Schumann, A.K. (2012). Knowledge-Rich Context Candidate Extraction and Ranking with KnowPipe. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*, pp. 3626–3630.

Temmerman, R. (2000). *Towards New Ways of Terminology Description: The Sociocognitive Approach.* Amsterdam/Philadelphia: John Benjamins.

Appendix A: Semantic categories: description and examples

SEMANTIC CATEGORY	DESCRIPTION	EXAMPLES
activity	activities, techniques and behaviors	AGRICULTURE REPRODUCTION LAND USE PLANNING
change of state	natural processes involving the change of state of a certain matter	ICE MELTING FLASH EVAPORATION SNOW SUBLIMATION
chemical element	chemical elements and compounds	CHLOROFLUOROCARBON MERCURY NICOTINAMIDE
construction	man-made buildings and structures	TOWER MILL BREAKWATER PIPELINE
disease	illnesses and conditions	BLACK LUNG DISEASE CANCER MALARIA
domain	scientific or knowledge fields	BIOLOGY METEOROLOGY COASTAL ENGINEERING
feature	properties, characteristics and variables	SOIL MOISTURE BODY SIZE DENSITY
force	types of energy	HEAT WAVE SOLAR ENERGY ELECTRICITY
information	documents and data	CLIMOGRAPH BIOLOGICAL CLASSIFICATION BATHYMETRIC CHART
instrument	man-made inventions or creations used as instruments	MONITORING INSTRUMENT DIGITAL BAROMETER SAND FILTER
landform	geographical and geological features	ISLAND KARST MOUNTAIN
lifeform	living beings or organisms	SEABIRD MANGROVE TREE PROTIST
location	spatial environments	MARINE BIOME TROPICAL RAIN FOREST EUROPE
mass of matter	massive entities composed of certain substances	PLANET OCEAN GLACIER
measure	measuring units	CELSIUS HORSEPOWER KILOMETER

movement of matter	types of mass movement	EBBING TIDE LANDSLIDE MUDFLOW
period	time periods or spans	MONTH SEASON HOUR
phenomenon	meteorological and geological phenomena	TSUNAMI RAIN VOLCANIC ERUPTION
process	natural and artificial processes with agents and patients	ABRASION WEATHERING GAS ADSORPTION
product	natural and artificial substances that are the result of a process	GLASSWARE DEODORANT COFFEE
substance	solid, liquid and gaseous substances or materials	GRANITE FOSSIL FUEL WOOD
system	scientific systems and models	THEORY OF RELATIVITY SCIENTIFIC LAW EMPIRICAL METHOD
technology	man-made creations and inventions	GENERATOR AIRCRAFT RADIOSONDE
vehicle	man-made inventions or creations used as vehicles	MOTOR VEHICLE ELECTRIC CAR DELIVERY TRUCK

Appendix B: Hyponymic knowledge patterns: description and examples

HYPONYMIC KP TYPE	DESCRIPTION	EXAMPLES
classification	they classify or divide the hypernym into hyponyms	HYPER is classified into HYPO HYPER is divided into HYPO types of HYPER are classified as HYPO
definition	they introduce the hyponym with a definition where the hypernym is the *genus*	HYPO: a HYPER HYPO: a type of HYPER HYPO, defined as HYPER
denomination	they introduce the hyponyms as particular denominations	a type of HYPER called HYPO a type of HYPER known as HYPO types of HYPER are called HYPO
enumeration	they show an exhaustive and numbered list of hyponyms for the hypernym	# types of HYPER: HYPO # HYPER: HYPO # types of HYPER occur: HYPO
exemplification	they present the hyponyms as examples, types or kinds	HYPER such as HYPO

	of the hypernym	HYPER types such as HYPO HYPER like HYPO
identification	they directly link the hyponym to the hypernym with a copulative verb	HYPO is a HYPER types of HYPER are HYPO a type of HYPER is a HYPO
inclusion	they present the hyponyms as concepts included in the notion of the hypernym	HYPER including HYPO HYPER types include HYPO among HYPER are HYPO
itemization	they introduce a non-exhaustive list of hyponyms for the hypernym	HYPO and other HYPER HYPO and other HYPER types types of HYPER: HYPO
range	they establish a span where several hyponyms can be found for the same hypernym	HYPER ranging from HYPO to HYPO HYPER types ranging from HYPO to HYPO
selection	they highlight main or preferred hyponyms for the hypernym	HYPER, especially HYPO HYPER, mainly HYPO HYPER, usually HYPO

Appendix C: Hyponymy subtypes

HYPONYMY SUBTYPE	DESCRIPTION	EXAMPLES
ability-based	hyponyms characterized by own abilities or characteristics	RENEWABLE RESOURCE HABITABLE PLANET AUTONOMOUS VEHICLE
activity-based	hyponyms characterized by the activity or stability of their composition	RADIOACTIVE SUBSTANCE ALKALI METAL ACTIVE DUNE
agent-based	hyponyms characterized by the agent that causes them	STORM TIDE AIR OXIDATION SPRINKLER IRRIGATION
amount-based	hyponyms characterized by their amount or quantity	TRACE ELEMENT RARE METAL SINGLE STORM
color-based	hyponyms characterized by their color	COLORLESS SOLID RED TIDE YELLOW LIQUID
composition-based	hyponyms characterized by their components or by their material	METALLIC ELEMENT CARBONATE SAND PINE FOREST
degree-based	hyponyms characterized by their degree of intensity, size or consequences	CATACLYSMIC ERUPTION LOW-MAGNITUDE EARTHQUAKE MEGA-SCALE EXTRACTION
denomination-based	hyponyms characterized by having a particular denomination with a proper noun	PACIFIC OCEAN SAHARA DESERT NEW YORK CITY
density-based	hyponyms characterized by their density or particle concentration	LIGHT ELEMENT DENSE WATER HEAVY METAL

domain-based	hyponyms characterized by the scientific or knowledge field to which they belong	AGRICULTURAL PRODUCT MUSICAL INSTRUMENT CHEMICAL INDUSTRY
effect-based	hyponyms characterized by the effects or consequences that they cause	TOXIC LIQUID HAZARDOUS SUBSTANCE GREENHOUSE GAS
function-based	hyponyms characterized by their function or purpose	DRINKING WATER SURVEILLANCE RADAR MANUFACTURING FACILITY
hardness-based	hyponyms characterized by their hardness level	SOFT WOOD HARD ROCK HARD STRUCTURE
height-based	hyponyms characterized by their height or depth level	SHALLOW WATER DEEP OCEAN HIGH TIDE
location-based	hyponyms characterized by their spatial location or position	OCEAN WATER SURROUNDING AIR TROPICAL STORM
method-based	hyponyms characterized by the method or the process that they involve	AEROBIC OXIDATION DIRECT SUBLIMATION INDUSTRIAL TREATMENT
moisture-based	hyponyms characterized by their moisture level	DRY SOLID SATURATED AIR ARID DESERT
movement-based	hyponyms characterized by their movement or direction	EBB TIDE OCEAN-GOING DREDGE OUTGOING RADIATION
origin-based	hyponyms characterized by their origin, i.e. the place where they come from or where they were created	NATURAL RESOURCE PINE WOOD COUNTRY ROCK
patient-based	hyponyms characterized by the patient that is affected by them	COAST EROSION ICE MELTING WATER TREATMENT
relation-based	hyponyms characterized by being related to other concepts	FOREIGN SUBSTANCE PARENT COMPOUND COVALENT SOLID
result-based	hyponyms characterized by the result that they cause, or by being the result of a process	TSUNAMIGENIC EARTHQUAKE PAPER INDUSTRY UNIMOLECULAR DECOMPOSITION
shape-based	hyponyms characterized by their shape	AMORPHOUS SOLID PARABOLIC DUNE L-SHAPED GROIN
size-based	hyponyms characterized by their size	TINY CRYSTAL GIANT PLANET COMPACT CAR
speed-based	hyponyms characterized by their speed	RAPID EROSION FLASH EVAPORATION SPONTANEOUS DECOMPOSITION
state-based	hyponyms characterized by the state of matter	SOLID SUBSTANCE FLUID ELEMENT

		MOLTEN ROCK
status-based	hyponyms characterized by a particular circumstance or situation	REGULATED SUBSTANCE UNTREATED WOOD CONTAMINATED SOIL
technology-based	hyponyms characterized by the technology that they use	MOTOR VEHICLE GREEN TECHNOLOGY DIGITAL BAROMETER
temperature-based	hyponyms characterized by their temperature	HOT GAS WARM OCEAN COLD AIR
texture-based	hyponyms characterized by their texture	VISCOUS LIQUID FINE SAND SOFT ROCK
time-based	hyponyms characterized by their duration, by their age, or by happening in a particular moment	WINTER ICE OLD ROCK ANNUAL PRECIPITATION
weight-based	hyponyms characterized by their weight	LIGHT-DUTY VEHICLE HEAVY-DUTY TRUCK LIGHT TRUCK