UNIVERSIDAD DE GRANADA

FACULTAD DE TRADUCCIÓN E INTERPRETACIÓN

Trabajo Fin de Máster

# CORPUS-BASED IDENTIFICATION OF HYPONYMY SUBTYPES AND KNOWLEDGE PATTERNS IN THE ENVIRONMENTAL DOMAIN

Juan Carlos Gil Berrozpe

Directora: Pamela Faber

Máster Universitario en Traducción Profesional

Junio de 2017

**Acknowledgements**

First and foremost, my deepest gratitude goes to Professor Pamela Faber, director and supervisor of this research, for guiding me in the right direction at all times. Not only is she capable of always finding the best words and providing the best advice, but she is also an outstanding source of inspiration and encouragement.

Likewise, this research would have been impossible without the valuable contributions of Professor Pilar León. I appreciate every brilliant idea provided by her, as well as her constant endeavouring to make me give the best of myself.

Finally, I would like to extend my words of gratitude to all the remaining people who have supported me during the development of this work and who have trusted in my capabilities in every moment. To all of you, thank you.

## Table of contents

**List of tables**



**List of figures**

# CORPUS-BASED IDENTIFICATION OF HYPONYMY SUBTYPES AND KNOWLEDGE PATTERNS IN THE ENVIRONMENTAL DOMAIN

## 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 terminological 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 (KRCs). 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 causality, function, and meronymy (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 this proposal, a pilot study (Gil-Berrozpe et al., in press) was conducted to ascertain whether the generic-specific relation could be subdivided in EcoLexicon (Faber et al. 2014, 2016), a multilingual and multimodal TKB on environmental science. For this purpose, the EcoLexicon

English Corpus was processed with Sketch Engine (Kilgarriff et al. 2004), where the word sketch (WS) module was used. Sketch Engine (available at: https://www.sketchengine.co.uk/) is a corpus query system characterized by its ability to identify the grammatical relations in which a term participates, apart from more common corpus management and analysis functionalities. Accordingly, WSs are automatic corpus-derived summaries of a word's grammatical and collocational behavior (Kilgarriff et al. 2004). In this pilot study, the taxonomies of ROCK (an entity) and EROSION (a process) were reconstructed. 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 work presents the preliminary 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 main objective of this work was to identify through corpus analysis the behavior of hyponymic relations in the environmental domain with regard to concept nature. In order to achieve this objective, the following operational objectives were determined:

- To examine the theoretical features of hyponymy and its refinement.
- To review the utility of corpus-based analysis for extracting semantic information.

- To analyze the formulation of hyponymic KPs in order to obtain accurate KRCs.
- To carry out a semi-automatic extraction of hierarchies from a specialized corpus by using customized WSs based on hyponymic KPs.
- To explore the correlation between hyponymic KPs and different semantic categories.
- To study the correlation between hyponymy subtypes and different semantic categories.
- To establish a general inventory of hyponymy subtypes applicable to a TKB.

The rest of this work is organized as follows. Section 2 presents the theoretical framework, which focuses on four broad topics: (i) the theoretical premises of traditional and modern theories of terminology; (ii) the relevance of conceptual relations in TKBs; (iii) the theoretical features of hyponymic relations and their refinement; and (iv) the use of corpus-based analysis in terminology work. 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 preliminary results of this research are presented and discussed. Section 5 highlights the conclusions that can be derived from this study and outlines plans for future research. Finally, the bibliography cited is preceded by three appendixes in which semantic categories, hyponymic knowledge patterns, and hyponymy subtypes are described and exemplified.

## 2. Theoretical framework

The theoretical framework of this research is based on general and contemporary theories of terminology (Section 2.1), semantic or conceptual relations and their relevance in TKBs (Section 2.2), the theoretical features of hyponymic or generic-specific relations and their refinement (Section 2.3), and the use of corpus-based analysis in terminology work (Section 2.4).

## 2.1. From traditional to contemporary theories of terminology

The General Terminology Theory (GTT), as proposed by Wüster (1968, 1979), was the first theoretical proposal in the field of terminology and, in comparison to subsequent theories, its focus was mainly on prescription. According to the GTT, concepts are independent of context and the terms that designate them. In this way, GTT does not account for pragmatic or communicative elements that are inherent in specialized language texts, which are regarded as irrelevant. Furthermore, the GTT associates specialized language with the univocity principle, which states that each concept can only be designated by one term, and each term can only refer to one concept.

In the GTT, concepts are considered to be the abstract units of thought that refer to elements in the real world, whereas terms are viewed as the linguistic labels that designate these concepts. In addition, one of the essential premises of the GTT is that there is a clear division between specialized language and general language, since specialized terms are totally different from general language words because of the monosemic aspect of univocity. This theory has an onomasiological approach, which means that concepts are taken as the starting point for term specification.

The main objectives of GTT were the following (Cabré 2003: 173): (i) the elimination of ambiguity in technical language through the standardization of its terminology to facilitate technical communication; (ii) the dissemination of the benefits of standardized terminology among specialized language users; and (iii) the transformation of terminology into a science.

While it is true that the GTT was a major breakthrough in the consolidation of terminology as a discipline, it did not address the social, pragmatic and cognitive elements of language, aspects that would be further explored in the subsequent theories. According to Edo Marzá (2012: 105), the GTT has serious limitations. Not only does it simplify terminology as a discipline and its various applications, but it also places excessive emphasis on the standardization of terminology work. Moreover, it does not provide an account of of reality from a communicative perspective and thus fails to reflect the behavior of terms in specialized texts. There is thus a significant gap between the premises of the GTT and linguistic reality. Nonetheless, the importance of the GTT was (and still is) widely acknowledged because, at the time, it offered the only set of principles and premises for compiling terminological data (Faber 2009: 111).

It was not until the early 1990s when more descriptive approaches arose as a reaction to the GTT. These new approaches have been labeled ‘social and communicative terminology theories’. Unlike the GTT, these theories describe terminological units in discourse, analyzing the sociological and discourse conditions that give rise to different types of texts (Faber 2009: 113). Since they emphasize how terms are used in real communicative contexts, they provide a more realistic view of terminology. These theories are socioterminology (Gaudin 1993, 2003) and the Communicative Terminology Theory (CTT) (Cabré 1999, 2000, 2003).

Socioterminology, proposed by Gaudin (1993, 2003), applies the principles of sociolinguistics to terminology. This theory opts for a more descriptive and diachronic study of language, and studies terms in real contexts. In contrast to the GTT, this approach acknowledges the existence of terminology variation and, subsequently, of synonymy and polysemy in specialized language. Furthermore, this theory highlights the importance of taking into consideration social and ethnic aspects that characterize the communication between experts and specialists. According to socioterminology, language is in constant change, which means that concept systems and definitions are not static. Therefore, the standardization pursued by Wüster is regarded as an extremely difficult goal to achieve.

The Communicative Terminology Theory (CTT), proposed by Cabré (1999, 2000, 2003), expands the scope of socioterminology to reflect the complexity of terms or specialized language units in real situational contexts (Cabré 2003: 164) and from a social, linguistic and cognitive perspective (Faber 2009: 114). The CTT introduces the notion of multidimensionality in terminology and identifies terms as polyhedrons with three dimensions: (i) the linguistic dimension, which describes the features of a specialized language unit as it is used in language; (ii) the communicative dimension, which describes how a specialized language unit is used in different contexts or communicative situations; and (iii) the cognitive dimension, which focuses on a description of concepts and conceptual relations. Interestingly, the CTT highlights the importance of describing hierarchical and non-hierarchical conceptual relations, as part of the cognitive facet of terminology. Similarly to socioterminology but unlike the GTT, the CTT states that the boundaries between specialized language and general language (i.e. terms and words) are not clear. Therefore, it is the context that determines whether a particular unit is a lexical unit or a terminological unit (Cabré 2003: 190). Moreover, and in order to better examine the communicative situation in which a specialized language unit is activated, terms are analyzed *in vivo*, based on their real use in context (Edo Marzá 2012: 105).

In summary, the CTT significantly improved the study of terminology. Nonetheless, despite its valuable contributions, it has the following drawbacks (Faber 2009: 115): (i) it does not opt for any specific linguistic model; (ii) it does not explain conceptual relations and their potential constraints; (iii) it does not offer a clear explanation of specialized meaning and its components.

In the early 2000s, the theories that are commonly known as 'cognitive-based terminology theories' appeared on the horizon as a result of the 'cognitive shift' (Evans & Green 2006) affecting linguistic theory. These approaches, which focused on the conceptual network underlying language, implemented premises from cognitive linguistics and psychology in regard to concept description and category structure (Faber 2009: 116). These theories are the Sociocognitive Terminology Theory (STT) (Temmerman 2000, 2001) and Frame-based Terminology (FBT) (Faber 2009, 2012, 2015).

The Sociocognitive Terminology Theory (STT) was developed by Temmerman (2000, 2001), who applied sociocognitive premises to terminology as a reaction to the constraints of traditional theories. The STT is mainly characterized by its emphasis on conceptual organization, and its focus on category structure from the perspective of cognitive linguistic approaches (Faber 2009: 117). This theory also adopts Rosch's Prototype Theory (Rosch 1978) as a model of categorization, which states that humans tend to cognitively organize entities into sets of categories with prototypes that contain the most typical features of the category member. Accordingly, the STT proposes the use of units of understanding, which can have a prototypical structure (categories) or a non-prototypical structure (concepts). Furthermore, the STT emphasizes the study of terms and concepts from a diachronic perspective, and proposes ontologies as the best method to implement conceptual representations through what is known as 'termontography' (Temmerman et al. 2005). Termontography links ontologies to multilingual terminological information, and implements them in terminological resources (Faber 2009: 118), what is evidently related to both conceptual relations and terminological knowledge bases (see Section 2.2).

Frame-based Terminology (FBT), proposed by Faber (2009, 2012, 2015), is a recent approach that shares many aspects of the CTT and the STT. It combines a descriptive approach with elements of corpus linguistics, cognitive semantics (Geeraerts 2010) and frame semantics (Fillmore 1976, 1982, 1985) to create structured specialized domains and non-language-specific representations. FBT is a cognitively-oriented terminology theory that operates on the premise that, in scientific and technical communication, specialized knowledge units activate domain-specific semantic frames that are in consonance with the users' background knowledge (Faber et al. 2016: 73). These frames are cognitive structuring devices based on experience that provide the background knowledge for the words in a language, as well as the way that those words are used in discourse (Buendía Castro 2013: 64). In addition to this, FBT sets its focus on the following aspects (Faber 2009: 123): (i) conceptual organization, reflected through frames or events; (ii) the multidimensionality of terminological units, expressed through both hierarchical and non-hierarchical relations; and (iii) the extraction of semantic and syntactic information through the use of multilingual corpora.

As reflected in this brief overview, terminology theories have evolved from a static perspective of terms and specialized language, to a more dynamic and multidimensional view of how concepts interact with each other in different communicative situations or contexts. Interestingly, one of the key aspects of contemporary terminology theories (e.g. the Sociocognitive Terminology Theory, Frame-based Terminology) is the description of the conceptual relations in semantic networks.

### 2.2. Conceptual relations and terminological knowledge bases

As stated in the previous section, the study of terminology and specialized language has been undergoing a 'cognitive shift' (Faber 2009: 111) over the last decade. This has placed greater attention on conceptual representation and knowledge organization. Modern terminology theories, which adopt more descriptive approaches, reflect dynamic phenomena (such as variation or multidimensionality) and emphasize the importance of accounting for hierarchical and non-hierarchical relations. These approaches are particularly relevant to terminology work, since an accurate representation of conceptual or semantic relations is required to build a comprehensive knowledge resource, such as a terminological knowledge base (TKB).

Conceptual relations (also known as 'concept relations' or 'semantic relations') are links created by human thought processes to describe a type of interaction between concepts. In the end, these relations enable the creation of conceptual systems, which are sets of concepts structured according to a domain knowledge, and which can be consulted by different user groups. Since a conceptual system is subject to time-induced changes of reality and cognition, it may show variable degrees of formality as it evolves during the steps of terminology work.

ISO Standard 704 (2009) states that there are two main types of conceptual relations: (i) hierarchical relations, which include both generic relations and partitive relations; and (ii) non-hierarchical relations, which include associative relations. Generic relations refer to generic-specific or hyponymic relations, i.e. a relation between a hypernym or superordinate concept and its hyponym or subordinate concept. According to ISO Standard 704 (2009: 9):

> A generic relation exists between two concepts when the intension of the subordinate concept includes the intension of the superordinate concept plus at least one additional delimiting characteristic. […] The superordinate concept in a generic relation is called the generic concept [hypernym] and the subordinate concept is called the specific concept [hyponym].

Partitive relations refer to part-whole or meronymic relations, i.e. a relation between a holonym or whole concept and a meronym or partial concept. According to ISO Standard 704 (2000: 9):

> A partitive relation is said to exist when the superordinate concept represents a whole, while the subordinate concepts represent parts of that whole. The parts come together to form the whole. The superordinate concept in a partitive relation is called the comprehensive concept [holonym] and the subordinate concept is called the partitive concept [meronym]. Subordinate concepts at the same level and sharing the same dimension are also called coordinate concepts.

Associative relations refer to any kind of non-hierarchical relation, i.e. a relation due to functionality, causality, location, time, etc. According to ISO Standard 704 (2009: 17–18):

> Associative relations are non-hierarchical. An associative relation exists when a thematic connection can be established between concepts by virtue of experience. Some associative relations exist when dependence is established between concepts with respect to their proximity in space or time. […] Some relations involve events in time such as a process dependent on time or sequence; others relate cause and effect.

However, conceptual relations can be more than the three general categories previously mentioned. For instance, Nuopponen (2005: 127) states that the distinction between generic, partitive and associative relations might be enough for traditional terminology work and term banks, but more advanced resources for terminology management or concept modelling could benefit from a wider range of conceptual relations. Therefore, Nuopponen (2005) proposed a comprehensive inventory of conceptual relations, divided into six main categories with many different subcategories: (i) logical concept relations; (ii) ontological concept relations; (iii) contiguity concept relations; (iv) influence concept relations; (v) functional concept relations; and (vi) interactional concept relations.

Accordingly, Meyer et al. (1992: 159) recognize the importance of representing conceptual relations in terminological resources, since they state that term banks or similar databases would be more useful both for people and for machines if they were organized in a way that resembled the organization of concepts in the mind. The combination of the linguistic approach of term banks and the cognitive approach of knowledge bases results in what are known as terminological knowledge bases (TKBs). Therefore, TKBs are cognitive terminological resources that represent the specialized knowledge of a certain field through related concepts and the terms that designate them in one or various languages (Gil-Berrozpe & Faber, in press). According to León-Araúz et al. (2013), TKBs should account for the representation of natural and contextual knowledge dynamism. Various issues must thus be considered when designing and creating a TKB. On the one hand, the organization of the knowledge field should accurately represent the concepts and the semantic relations linking them. On the other hand, access to information and its retrieval should facilitate knowledge acquisition.

To enhance the representation of conceptual information in terminological resources like TKBs, different conceptual relations have been examined from multiple perspectives. For instance, causal relations have different subtypes and show many ways of expressing causality, since causation can be formally expressed by passive, active, subject-object, nominal or verbal propositions (Marshman 2002, Marshman et al. 2002). Another example is found in meronymic relations, which were decomposed by Winston et al. (1987) into six subtypes according to the way parts contribute to the structure of the wholes: component-integral object (WHEEL-CAR), member-collection (SOLDIER-ARMY), portion-mass (METER-KILOMETER), stuff-object (ALCOHOL-WINE), feature-activity

(PAYING-SHOPPING), and place-area (OASIS-DESERT). However, and in spite of their importance in taxonomies, the decomposition of hyponymic relations has been little researched in comparison to other semantic relations.

### 2.3. Hyponymy and its refinement

Hyponymy (the generic-specific relation) is generally defined as a relation of inclusion whose converse is hyperonymy (Murphy 2006: 446). Of all conceptual or semantic relations, hyponymy is considered to be the backbone of ontology-based terminological resources because it is the origin of all concept hierarchies (Barrière 2004a: 233–234). Furthermore, hyponymy is also central to many models of the lexicon for three main reasons (Murphy 2003: 217): (i) its inference-invoking nature; (ii) its importance in definition; and (iii) its relevance to selectional restrictions in grammar.

In addition to this, hyponymy is the semantic relation that plays the most important role in our conscious thinking about what a word means. Accordingly, classical or Aristotelian definitions use hyponymy to describe a concept, since they consist of *genus* and *differentiae*, i.e. a hyperonym and the qualities that distinguish the defined hyponym from the larger class (Murphy 2003: 217). For example, in the Aristotelian definition of the concept TABLE, “piece of furniture” would be the *genus*, and “supported by one or more legs and having a flat top surface on which objects can be placed” would be the *differentiae*.

However, hyponymic relations are complex, and thus hypernym-hyponym pairs can be studied from multiple perspectives. As in causality or meronymy, hyponymy can also be refined to provide an enhanced representation of generic-specific relations. In this line, two main proposals have been made as a means to improve the description of hyponymic relations: (i) the specification of hyponymy subtypes (Miller 1998, Murphy 2003), and (ii) the establishment ‘facets’ and/or ‘microsenses’ (Cruse 1995, 2002).

Regarding hyponymy subtypes, Murphy (2003: 219) states that hyponymy can indeed be decomposed in the same way as other semantic relations, but it is unclear the number of subtypes and whether they can provide a valid and comprehensive taxonomy of hyponymic relations. The most commonly accepted distinction is between taxonomic hyponymy (‘is-a-kind-of’ relation) and functional hyponymy (‘is-used-as-a-kind’ relation) (Miller 1998). For example, COW is in a taxonomic relation to ANIMAL (a COW is an ANIMAL), but in a functional relation to LIVESTOCK (a COW functions as LIVESTOCK). Moreover, whilst taxonomic relations are always analytic, functional relations are vaguer since they are not logically necessary relations (not every COW is LIVESTOCK) (Murphy 2003: 220).

On the other hand, Cruse (1995, 2002) proposes ‘facets’ as a means to distinguish between different types of hyponymy. ‘Facets’ are dimensions or aspects of a concept that show a high degree of autonomy and distinctness (Cruse 2002: 4), making it possible to describe that concept from any of those multiple perspectives independently. For instance, Cruse (2002: 4) highlights two ‘facets’ or dimensions in the hyponyms of BOOK, and divides them into two sets: ‘physical object’ (such as HARDBACK or PAPERBACK) and ‘abstract text’ (such as NOVEL or BIOGRAPHY). In these cases, the cohyponyms of the same hyperonym display within-set incompatibility, but between-set compatibility (a certain BOOK can be simultaneously a NOVEL and a HARDBACK, but a HARDBACK cannot be a PAPERBACK at the same time).

Furthermore, another important phenomenon in the specification of hyponymic relations is the existence of ‘microsenses’ (Cruse 2002: 5). A ‘microsense’ is a specific meaning of a concept (e.g. regarding its properties, attributes or functions) which is only activated in a certain context, and which makes it differ from the meaning of the same concept in a different context. For example, although KNIFE generally has a single sense, it can be classified in different domains under a variety of hyperonyms (WEAPON, TOOL, SURGICAL INSTRUMENT, etc.).

At this stage, it is clear that hyponymy itself is a broad conceptual relation that contains many specific nuances that could be exploited as a means to decompose it and obtain a more fine-grained vision of generic-specific relations.

In particular, TKBs displaying conceptual networks with different semantic relations would benefit greatly from this enhancement.

### *2.3.1. 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.

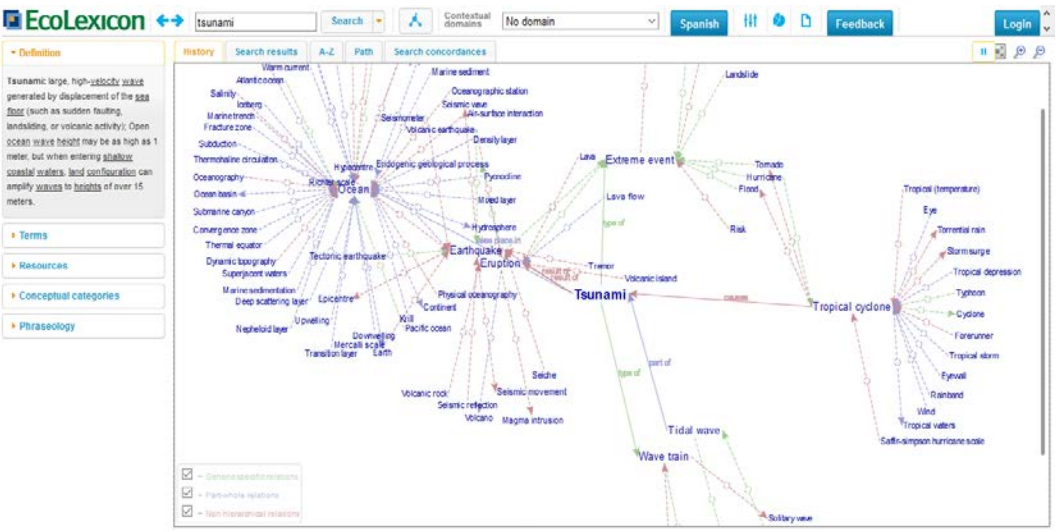



Figure 1. Visual interface of EcoLexicon (conceptual network of TSUNAMI)

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 type, 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).

A recent pilot study on hyponymy refinement (Gil-Berrozpe et al., in press) was based 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. After correcting property inheritance and enriching the hierarchies with new concepts, a fine-grained set of subtypes was specified. Interestingly, the results of this pilot study indicated that hyponymy subtypes were based on whether the concept was an entity (ROCK) or a process (EROSION).

On the one hand, five different entity-related hyponymy subtypes were established according to the dimensions triggered by each entity: state-based hyponymy (dependent on the state of matter of the hyponyms, as in SOLID ROCK), formation-based hyponymy (dependent on the formation process of the hyponyms, as in SEDIMENTARY ROCK), composition-based hyponymy (dependent on the components or the constituents of the hyponyms, as in SILTSTONE), location-based hyponymy (dependent on the physical situation or location of the hyponyms, as in PLUTONIC ROCK), and attribute-based hyponymy (dependent on the traits or features of the hyponyms, as in PERMEABLE ROCK).

On the other hand, four process-related hyponymy subtypes were established in relation to the characteristics of each process: agent-based hyponymy (dependent on the agent or the promoter that causes the hyponyms, as in SEA EROSION); patient-based hyponymy (dependent on the entity or location affected by the hyponyms, as in CHANNEL SCOUR); result-based hyponymy (dependent on the results and effects of the hyponyms, as in GULLY EROSION); and, again, attribute-based hyponymy (dependent on the traits or features of the hyponyms, as in POTENTIAL EROSION).

Therefore, this pilot study (Gil-Berrozpe et al., in press) provided some interesting insights in relation to hyponymy refinement. Nonetheless, the scope of the corpus analysis was only limited to expanding the hierarchies of an entity and a process. For this reason, it was necessary to pay more attention to the linguistic markers expressing hyponymy and to a wider series of semantic categories.

### 2.4. Corpus-based analysis in terminology work

In a broad sense, corpora are large compilations of texts where experts express their knowledge and make it accessible to the public (Bourigault and Slodzian 1999). However, a more precise definition of what is meant by 'corpus' in terminology work is provided by Pearson (1998: 43):

> (…) a corpus is an artefact; it is selected, chosen or assembled according to explicit criteria. It is stored in electronic form. It consists of pieces of naturally occurring language. In this context, we understand naturally occurring to mean that the pieces of language have not been tampered with or edited. The corpus may, however, be annotated during or after the compilation process; grammatical tags or SGML markups (e.g. indicating text origin, authorship) may be added to facilitate information retrieval. A corpus may be used as a "sample of the language" (Sinclair) or because it is "representative of a given language" (Francis). A corpus may be a collection of transcribed spoken and/or written pieces of language, contrary to what the use of the word text might suggest.

In relation to this, Sager (1990: 55) adds that texts are the basic material of terminologists, and that the terminological resources developed by them aim at reflecting the usage of terms in natural language contexts. Corpus analysis is thus commonly used by terminologists in first instance to find terms in large corpora and extract their syntactic and semantic information. Traditionally, corpora have been analyzed and processed by manually reading concordance lines related to a particular term. León-Araúz et al. (2016: 73) state that this time-consuming task has led to the development of new corpus-based methods and applications to analyze and extract linguistic information.

Accordingly, a more efficient method of extracting information from corpora is to search for knowledge-rich contexts (KRCs), which are contexts "indicating at least one item of domain knowledge that could be useful for conceptual analysis" (Meyer 2001: 281). Such contexts are highly informative since they provide conceptual information and domain knowledge (Meyer 2001), and usually codify these semantic relations in the form of knowledge patterns (KPs), which can be used to find KRCs in corpora. KPs are the lexico-syntactic patterns between terms encoded in a proposition in real texts (Meyer 2001).

Bielinskiene et al. (2012: 18) state that important conceptual characteristics are expressed in KRCs in the form of semantic relations (such as hyponymy or meronymy) and that they can be identified through KPs. In this line, these elements make it possible to extract the relevant terminography-oriented knowledge about the concept from a corpus and then use that information to provide a starting point for any terminological purpose (Bielinskiene et al. 2012: 18). However, this task is complicated by the fact there are no user-friendly publicly available applications that allow terminologists to find KRCs in their own corpora with ready-made KPs. This means that terminologists still tend to rely on manual work to extract all the semantic information that they need for the description of specialized concepts (León-Araúz et al. 2016: 73).

Because of their interest for terminology work, corpus-based analysis and KPs have become a major research topic over the years as a method of automatically or semi-automatically extracting linguistic information concerning different conceptual or semantic relations. Table 1 shows a typology of KPs (herein referred to as 'lexical patterns') proposed by Bowker and Pearson (2002: 219), along with the type of conceptual knowledge expressed and some examples of possible KRCs:

| Lexical patterns | Conceptual knowledge | Example |
|---|---|---|
| 'is a', 'kind of', 'type of', 'includes' | generic-specific relations | The tabor is a type of drum. |
| 'has a', 'contains', 'consists of', 'includes' | part-whole relations | A snare drum has a batter head and a snare head. |
| 'used for', 'used to', 'employed to' | function relations | A wooden stick is used to strike the drum head. |
| 'causes', 'produces', 'produced by', 'results from' | cause-effect relations | Striking the drum head causes the snares to vibrate. |
| 'also called', 'also known as', 'sometimes referred to as' | possible synonymy | The tambourine, also known as the tambourin provençale, is the largest of all the tabors. |

Table 1. Lexical patterns and possible KRCs (Bowker & Pearson 2002: 219)

Nonetheless, these conceptual relations have not received an equal amount of attention. For example, meronymic or part-whole relations have been widely researched (Berland & Chamiak 1999, Girju et al. 2003). They can be codified by prepositional phrases, possessives, and partitive verbs. One of their most interesting features is the fact that many meronymic KPs can be polysemic (León-Araúz et al. 2016: 74). For example, 'including' expresses hyponymy and meronymy at the same time; and 'formed by' expresses meronymy and causality (León-Araúz 2014). Moreover, non-hierarchical relations have also been studied and implemented as KPs. For instance, causality (Marshman 2002, Marshman et al. 2002) can be expressed by passive, active, subject-object, nominal or verbal propositions, involving all kinds of causative nouns and verbs. However, the most commonly studied patterns are hyponymic KPs, because of their importance in relation to categorization and property inheritance (Barrière 2004a).

### *2.4.1. Corpus-based extraction of hyponymic information*

As previously stated, much attention has been paid to the development of semi-automatized procedures for extracting relevant KRCs in recent years. As with the rest of semantic relations, the strategies used to formulate hyponymic KPs to find useful hyponymic KRCs are language-specific. This is the reason why many authors have decided to discover the linguistic markers of different languages.

For example, Hearst (1998) proposed the automatic extraction of hyponymic patterns from texts by looking at sentences that contain hypernym-hyponym pairs in WordNet. In this way, six patterns in English (including simplified examples like 'X such as Y’ or ‘X and other Y’) were identified by mere observation of text and by analyzing the context between a hyponym and its superordinate found in the corpus. However, it was noted that these hyponymic patterns generated a large number of mistakes, either because the extracted relation was too far away, because there were subjective opinions with no interest, or because of parsing errors. Hearst (1998) only evaluated the pattern containing the expression ‘or other’, and 63% of all the contexts found contained actual hypernym-hyponym pairs.

Another example of automatic acquisition of hyponymic relations from corpora is Liu et al. (2006). Although they state that the two main approaches to automatic or semi-automatic hyponymy extraction are pattern-based extraction and statistics-based extraction, they performed a pattern-based extraction of hypernym-hyponym pairs from publicly available Chinese text. They focused on ‘is-a’ patterns, as they were the most frequent in their target corpora. Because of the effectiveness of their method (around 90%), this approach looked promising for Chinese, but more from a computational perspective.

Bielinskiene et al. (2012) proposed a series of 18 definitional patterns as a means of obtaining hypernym-hyponym pairs in Lithuanian. These patterns included prototypical examples such as ‘X is Y’ or ‘X is considered as Y’. However, their proposal also included polysemic KPs such as ‘X constitutes Y’ or ‘X includes Y’, which can also be used to find meronymic KRCs. Unfortunately, their results showed that 55% of all cases (127 contexts from 227) were irrelevant because grammar, the absence of hyponymy or other inaccuracies related to the expressions (Bielinskiene et al. 2012: 24).

More recently, Baisa & Suchomel (2015) tested hyponymy extraction in a specialized Czech corpus on the domain of land surveying by using different sketch grammars in Sketch Engine. Corpus querying in Sketch Engine is based on corpus query language (CQL), which makes it possible to formalize grammar patterns in the form of regular expressions combined with POS-tags. CQL expressions in Sketch Engine can be stored in a sketch grammar, which produces word sketches (WSs).

Finally, León-Araúz et al. (2016) went one step further by creating a total of 56 sketch grammars with the objective of automatically extracting semantic information from large corpora: 18 generic-specific grammars, 17 part-whole grammars, 10 cause grammars, 7 function grammars, and 4 location grammars. These sketch grammars are dynamic and have different permutations or variations so as to encompass all the different aspects that can involve every relation independently, and avoid possible problems such as noise or loops. As explained by León-Araúz et al. (2016: 76):

> In the development of our sketch grammars (a total of 56), we […] 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.

Since the hyponymic KP analysis (Sections 3.1 and 4.1) in this work involves a practical application of the 18 hyponymic sketch grammars developed by León-Araúz et al., their formulation will be explained in the upcoming sections.

## 3. Materials and methods

This 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. The default word sketches (WSs) provided by Sketch Engine represent different linguistic relations, of which only the default 'modifier' WS was used for the extraction of hyponyms for the hyponymy subtypes analysis, through the analysis of multiword terms (see Section 3.2). As previously mentioned in Section 2.4.1, apart from the default options, the system also permits the creation of customized WSs by storing corpus query language (CQL) expressions in new sketch grammars.

The corpus was thus compiled by implementing the 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 2 shows a summarized version of the KPs, whereas Table 3 shows an example of one of the patterns converted into a CQL sketch grammar.

| |
|---|
| **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 2. Hyponymic KPs (León-Araúz et al. 2016)

```
1:"N.*"",|\("?[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"]
```

Table 3. CQL representation of a hyponymic KP

The grammar in Table 2 can be interpreted as follows: 1 fills the role of the hyperonym, which must be a noun. Then it can optionally be followed by a comma or a bracket, by ‘that’ or ‘which’, or any modal verb. After that, we find the KP itself, acting as an anchor point, which is ‘classified’, ‘categorized’ or ‘categorised’ preceded by the lemma ‘be’, or a comma or a bracket, optionally followed by the preposition ‘by’ and any word that is not a verb, plus the preposition ‘in’ or ‘into’. Then there may be any number of words (included zero) that are not verbs optionally followed by lemmas such as ‘type’, ‘kind’, ‘example’, ‘group’, etc. Then again there may be any number of words that are not verbs followed by 2 others, namely, the hyponym, which must also be a noun, and none of the previous lemmas.

Examples (1) and (2) are two of the concordances that can be matched with this grammar:

(1) Mild climates can be classified into three subtypes: humid subtropical climates, marine west-coast climates, and Mediterranean climates.

(2) Water is commonly categorized into surface water and groundwater.

### 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 following CQL query: [ws(".*-n","\"%w\" is the generic of...",".*-n")]. The resulting subcorpus contained a total of 938,386 potential hyponymic concordances (Figure 2).

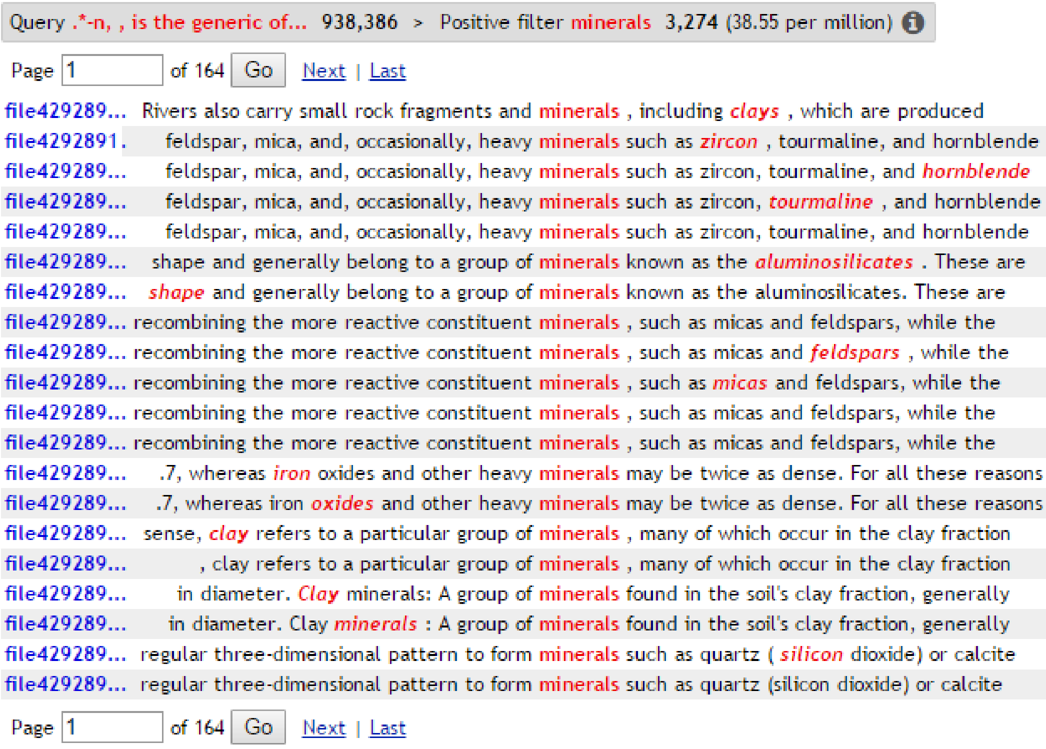



Figure 2. Concordances retrieved from the hyponymic subcorpus

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.

Therefore, a randomized portion of the hyponymic subcorpus was examined, from which 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 |

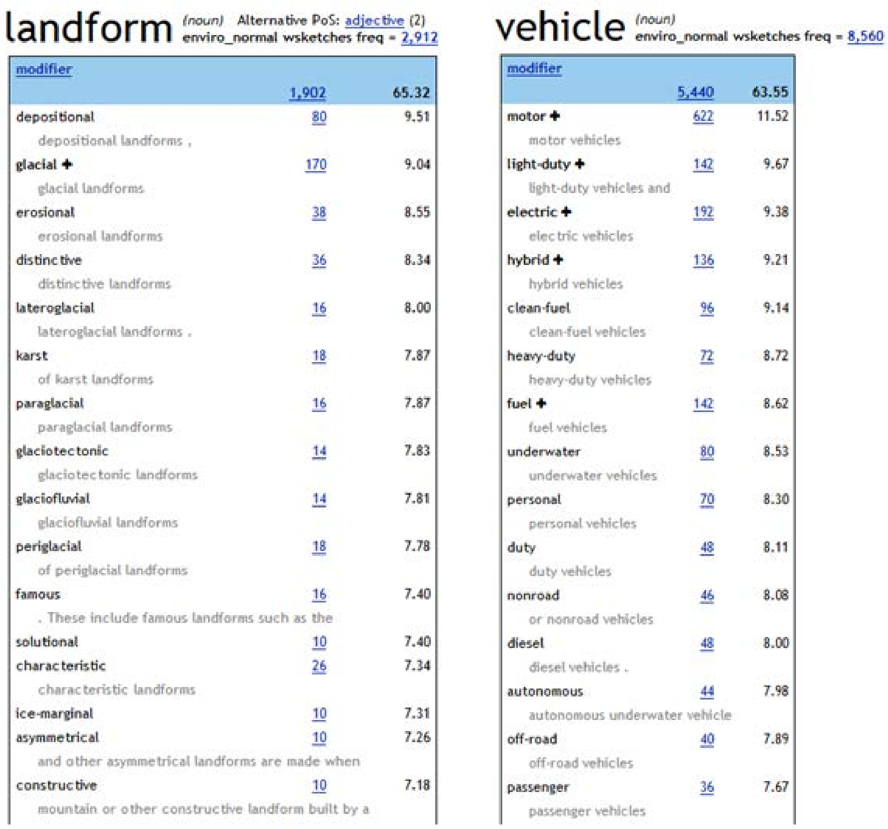
Figure 3. Extract of the hyponymic KP table

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, the 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 hyponymic 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 the previous results seemed to indicate that hyponymy subtypes depended on the nature of the concept (Gil-Berrozpe & Faber 2016), this hypothesis had to be confirmed through the use of more fine-grained semantic categories (e.g. activity, landform, chemical element, etc.).

Therefore, a second compilation of hypernym-hyponym pairs had to be performed, though this time with a greater focus on semantic categories. For this reason, 109 hypernyms of concepts belonging to a wide range of semantic categories were extracted: 32 natural entities, 32 artificial entities, 21 natural processes, 17 artificial processes, and 7 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 | 80 | 9.51 |
| depositional landforms , | | |
| glacial + | 170 | 9.04 |
| glacial landforms | | |
| erosional | 38 | 8.55 |
| erosional landforms | | |
| distinctive | 36 | 8.34 |
| distinctive landforms | | |
| lateroglacial | 16 | 8.00 |
| lateroglacial landforms . | | |
| karst | 18 | 7.87 |
| of karst landforms | | |
| paraglacial | 16 | 7.87 |
| paraglacial landforms | | |
| glaciotectonic | 14 | 7.83 |
| glaciotectonic landforms | | |
| glaciofluvial | 14 | 7.81 |
| glaciofluvial landforms | | |
| periglacial | 18 | 7.78 |
| of periglacial landforms | | |
| famous | 16 | 7.40 |
| . These include famous landforms such as the | | |
| solutional | 10 | 7.40 |
| characteristic | 26 | 7.34 |
| characteristic landforms | | |
| ice-marginal | 10 | 7.31 |
| asymmetrical | 10 | 7.26 |
| and other asymmetrical landforms are made when | | |
| constructive | 10 | 7.18 |
| mountain or other constructive landform built by a | | |

vehicle (noun)
enviro_normal wsketches freq = 8,560

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

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, the 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. Preliminary 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 the hyponymic KP analysis. As can be observed, 21 semantic categories were found. (See Appendix A for the description and typical examples of each category.)

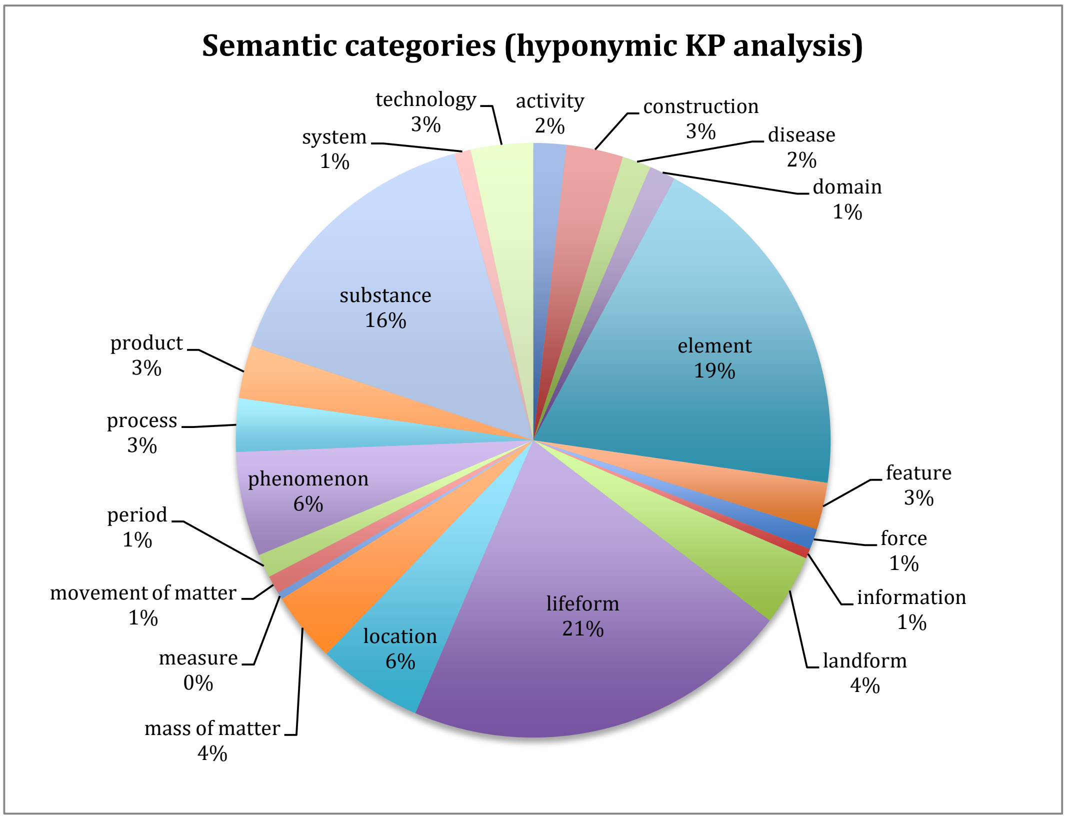



Figure 6. Semantic categories of the concepts in the hyponymic KP analysis

The results of this 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 a comprehensive list of all the patterns identified.)

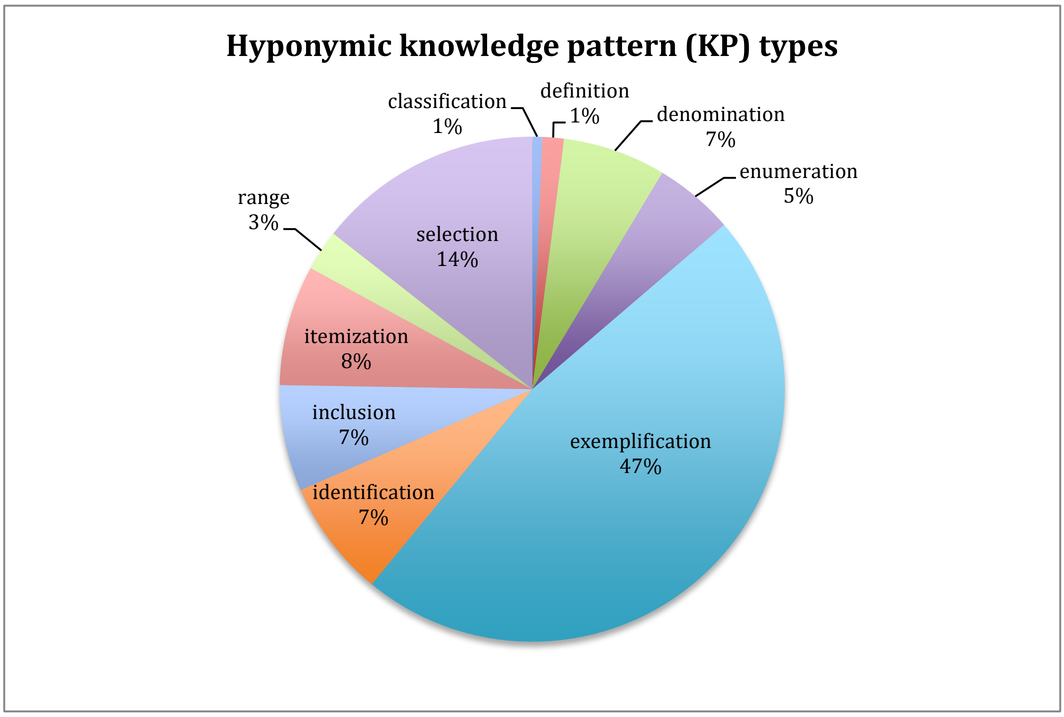



Figure 7. Hyponymic knowledge patterns

As reflected in the 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 type, comprised almost half of the sample analyzed. Because of the quantity of information in these patterns (e.g. HYPER such as HYPO, HYPER like HYPO), they were typical of the most common semantic categories, namely, chemical element, lifeform, and substance. The second most significant group of categories were location, phenomenon, landform, and construction. The other semantic categories were found in significantly fewer concordances.

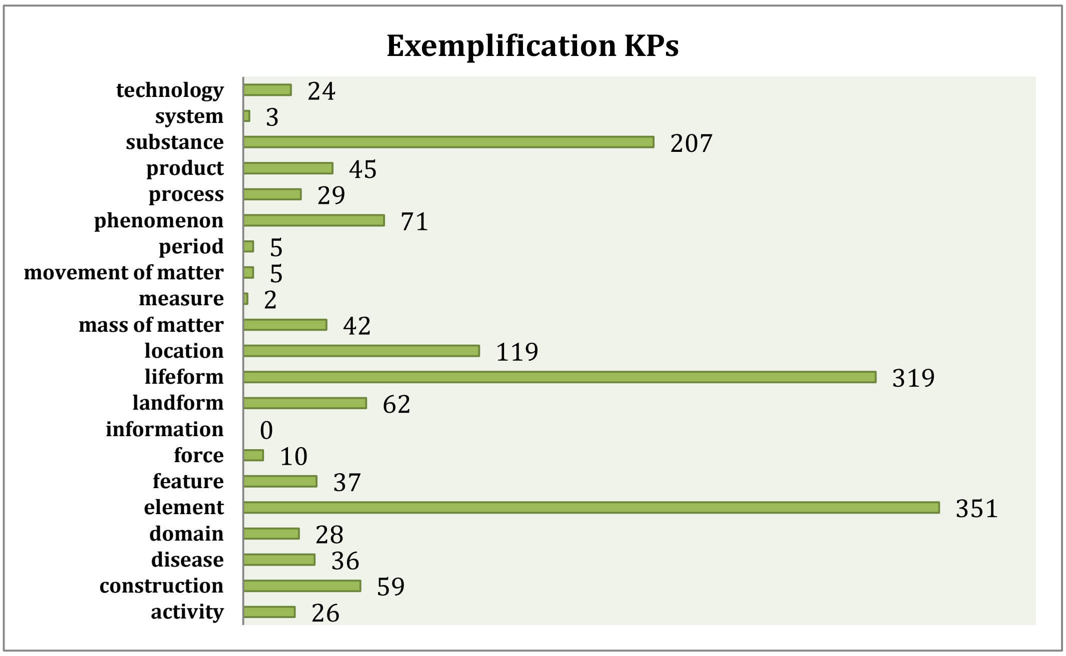



Figure 8: Exemplification KPs per semantic category

Since exemplification KPs were the most common, the only conclusion that can be derived is that the occurrences of exemplification KPs per semantic category were proportional to the ratios of semantic categories shown in Figure 7. In other words, they did not show any preference for a specific semantic category and, accordingly, presented a universal value.

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. These three pattern types also presented polyvalent structures (e.g. HYPER, especially HYPO; HYPER, mainly HYPO // HYPO and other HYPER, types of HYPER: HYPO // HYPER including HYPO, HYPER types including HYPO) that, in a similar way to exemplification KPs, can link concepts belonging to any semantic category on a non-preferential basis.

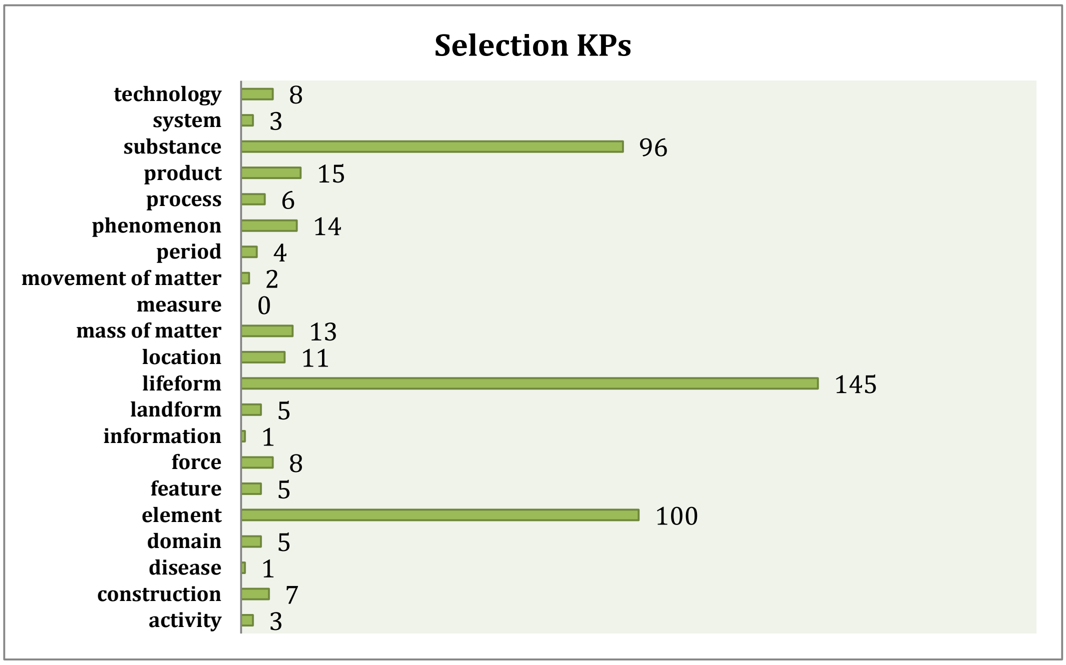



Figure 9. Selection KPs per semantic category

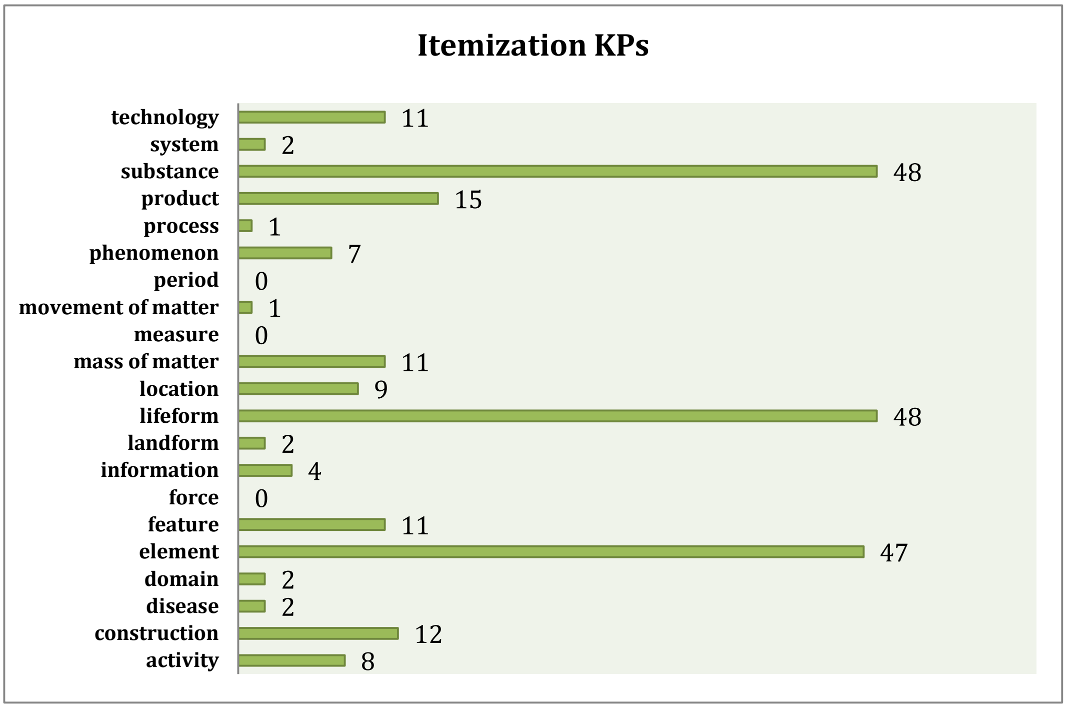



Figure 10. Itemization KPs per semantic category

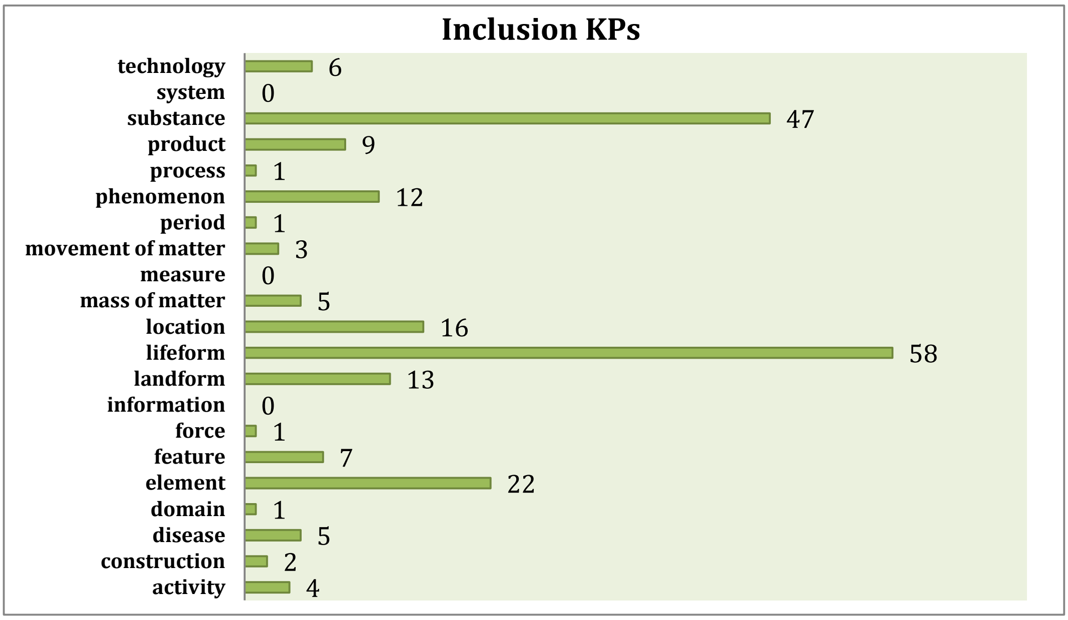



Figure 11. Inclusion KPs per semantic category

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, encyclopedias, whose function is 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 (e.g. HYPO is a HYPER, a type of HYPER is a HYPO) and denomination KPs (e.g. a type of HYPER called HYPO, a type of HYPER known as HYPO) are also activated by semantic categories in relation to the ratios shown in Figure 7. 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.

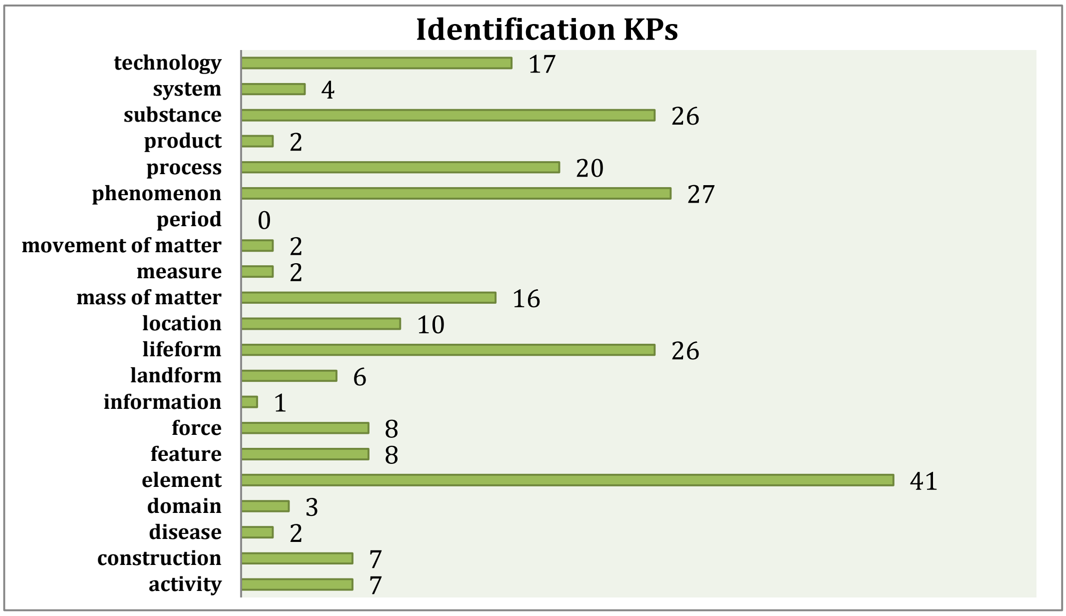



Figure 12. Identification KPs per semantic category

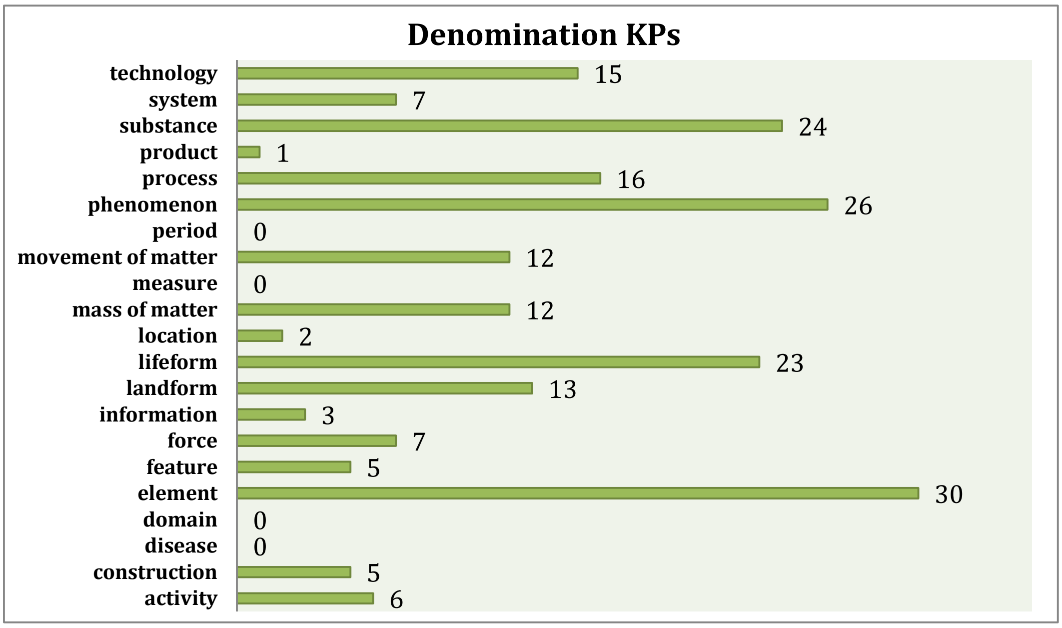



Figure 13. Denomination KPs per semantic category

The same could also be true of definition KPs (Figure 14), where the categories of technology and phenomenon shared the second position, after substance. Once again, the KP expressions in this category (e.g. HYPO: a HYPER, HYPO: a type of HYPER) specifically define a concept in terms of its superordinate, so they were commonly shown linking concepts of complex understanding.

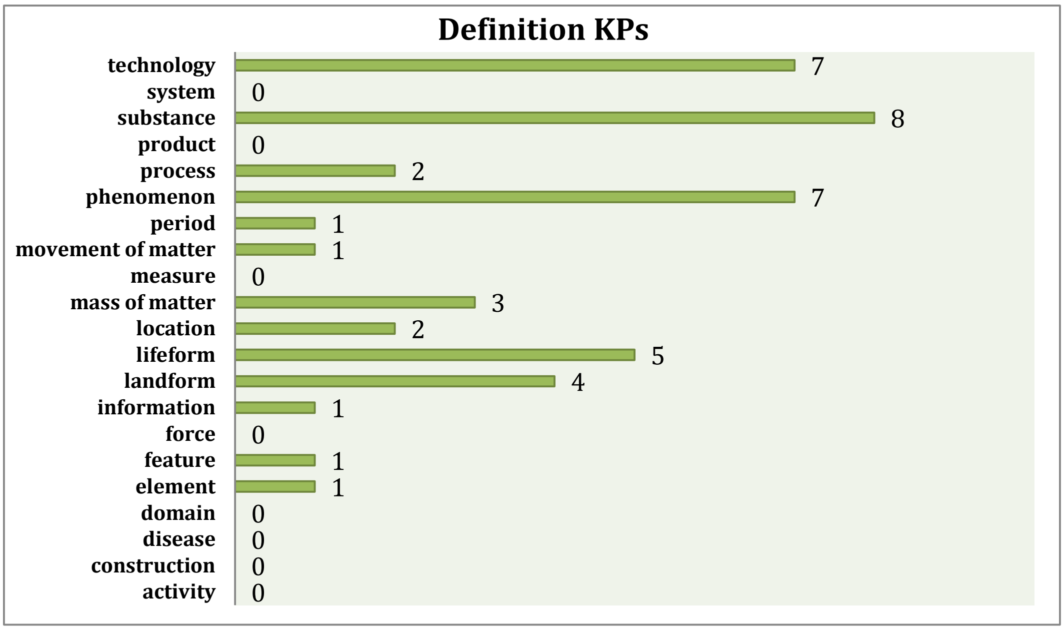



Figure 14. Definition KPs per semantic category

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

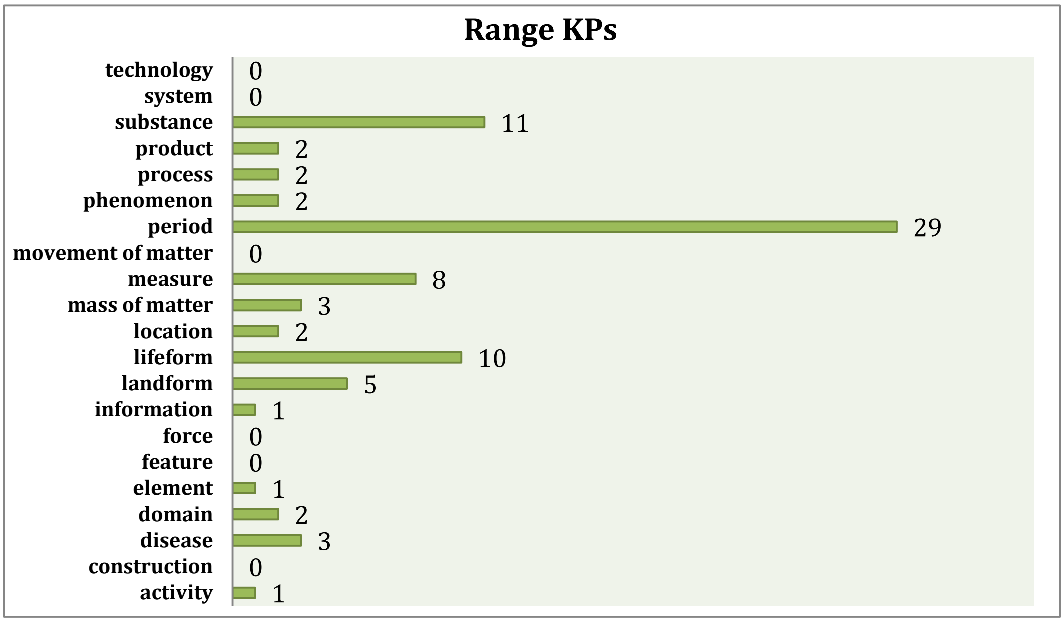



Figure 15. Range KPs per semantic category

Finally, in the case of enumeration KPs (Figure 16) and classification KPs (Figure 17), it was not possible to extract any specific correlation pattern. The results showed that enumeration KPs (e.g. # types of HYPER: HYPO, # types of HYPER are: HYPO), in the same way as exemplification KPs, were applicable to any concept type without specific preferences.

Moreover, the data obtained in these preliminary results for classification KPs (e.g. HYPER is classified into HYPO, HYPER is divided into HYPO) was insufficient to draw any definitive conclusions. However, the most frequent semantic categories linked by classification KPs should refer to those concepts that are commonly studied through scientific classifications (i.e. lifeform, chemical element and substance). Even though further research is needed to confirm this hypothesis, the results in Figure 17 showed lifeform and substance in first and second position, respectively.

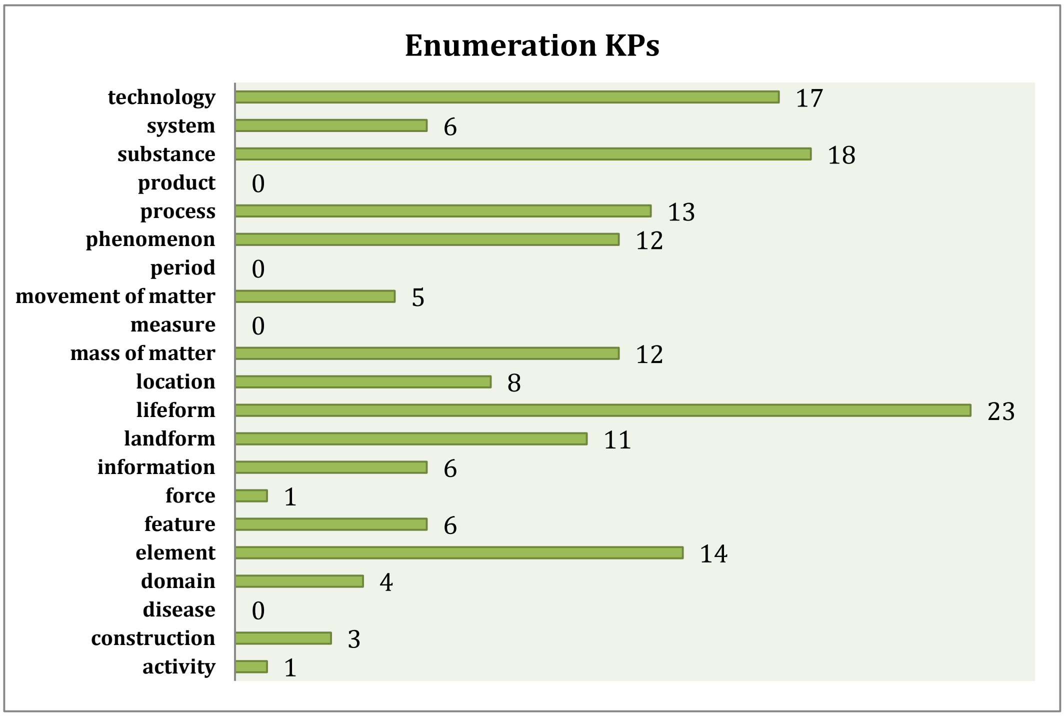



Figure 16. Enumeration KPs per semantic category

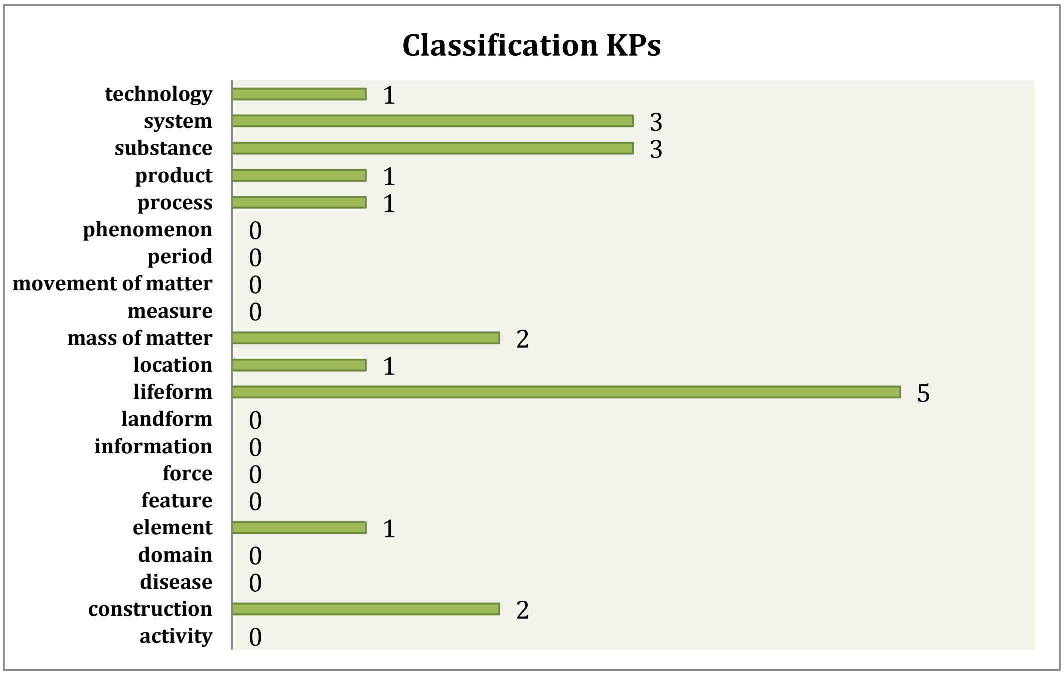



Figure 17. Classification KPs per semantic category

## 4.2. Hyponymy subtypes analysis: general results

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

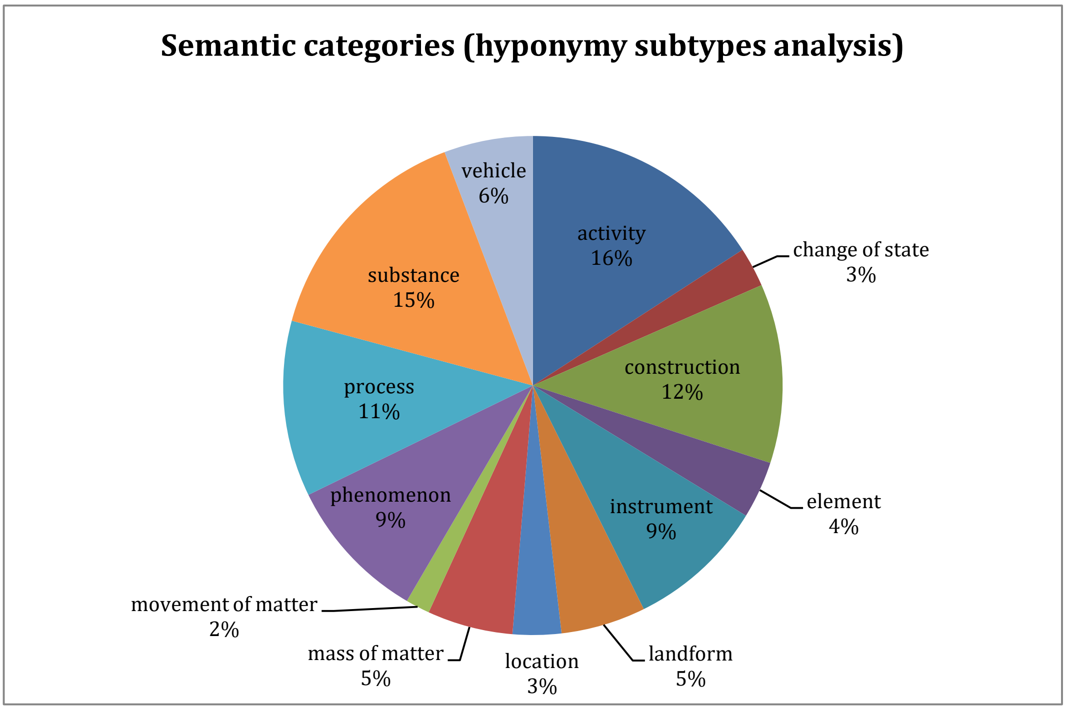



Figure 18. Semantic categories of the concepts in the hyponymy subtypes analysis

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 subtypes analysis: instrument, vehicle, and change of state (Appendix A).

The decomposition of the generic-specific relation was based on shared 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).

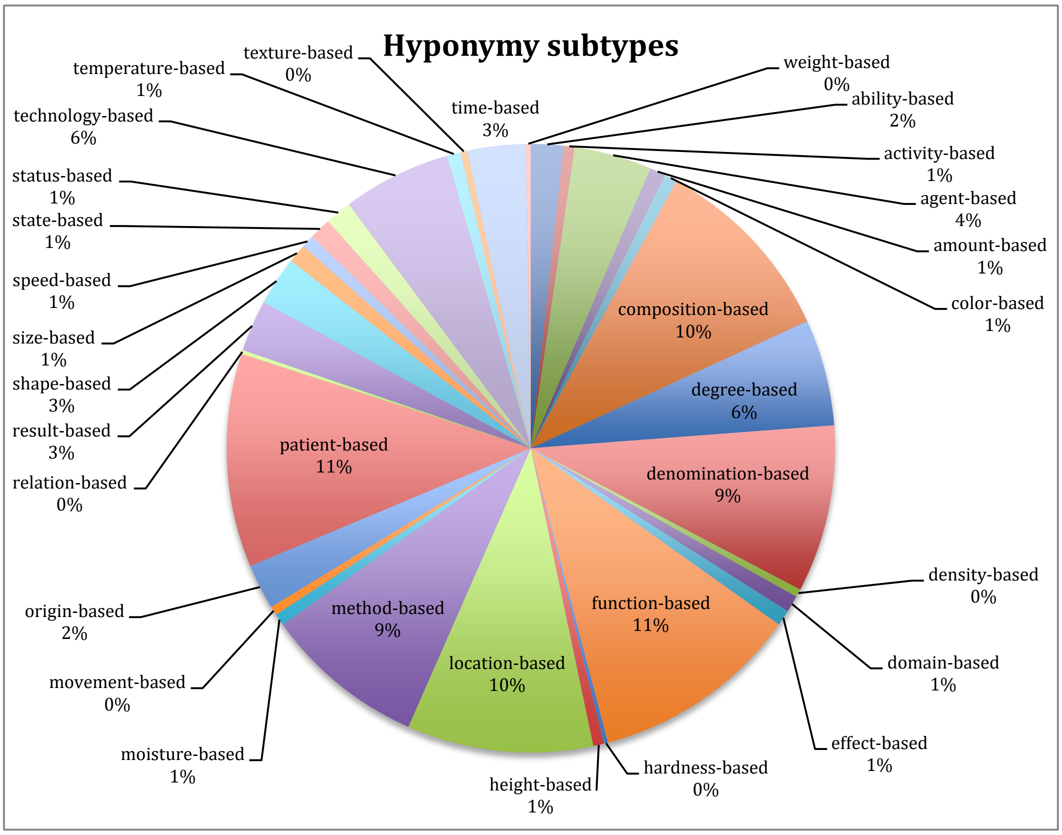



Figure 19. Hyponymy subtypes

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*

The most interesting results obtained are shown in 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 (e.g. FISH FARMING // SUSTAINABLE MANUFACTURING), process (e.g. SOIL EROSION // OXYGENIC REACTION), phenomenon (e.g. STEAM ERUPTION // SUPERCELL TORNADO), and change of state (e.g. ICE CAP MELTING // DIRECT SUBLIMATION). There were no entity-related semantic categories because these two subtypes of hyponymy appear to be exclusive to process-related semantic categories.

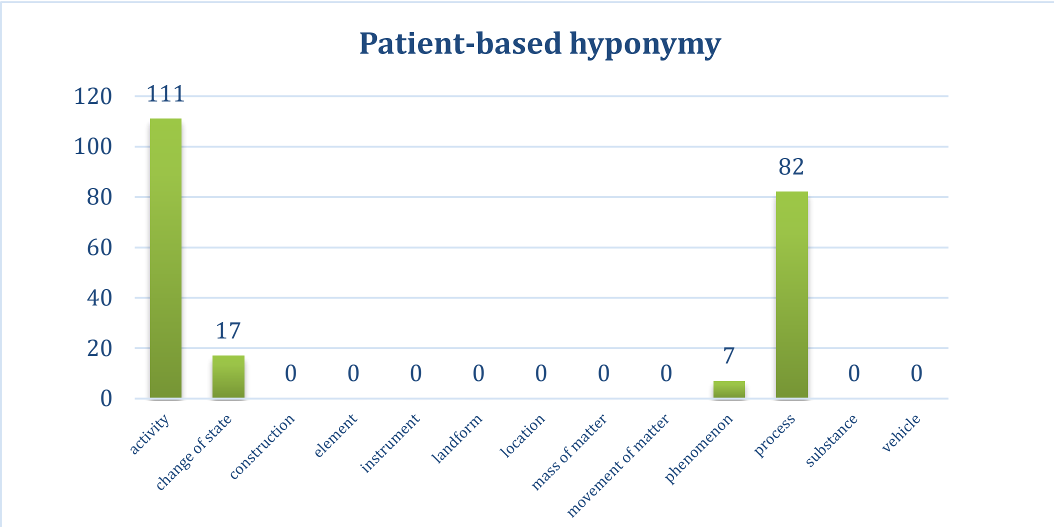



Figure 20. Patient-based hyponymy subtypes per semantic category

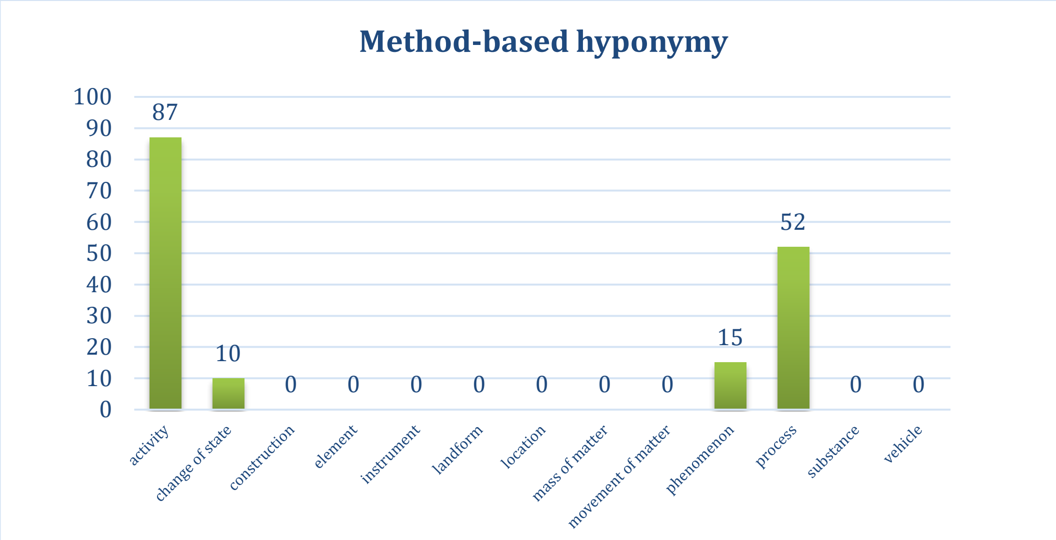



Figure 21. Method-based hyponymy subtypes per semantic category

As can be observed, in both cases 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 referred to process-related semantic categories, namely activity (e.g. PASTORAL FARMING // WELL DRILLING), process (e.g. WIND EROSION // NET PHOTOSYNTHESIS), and phenomenon (e.g. VOLCANIC ERUPTION // TSUNAMIGENIC EARTHQUAKE).

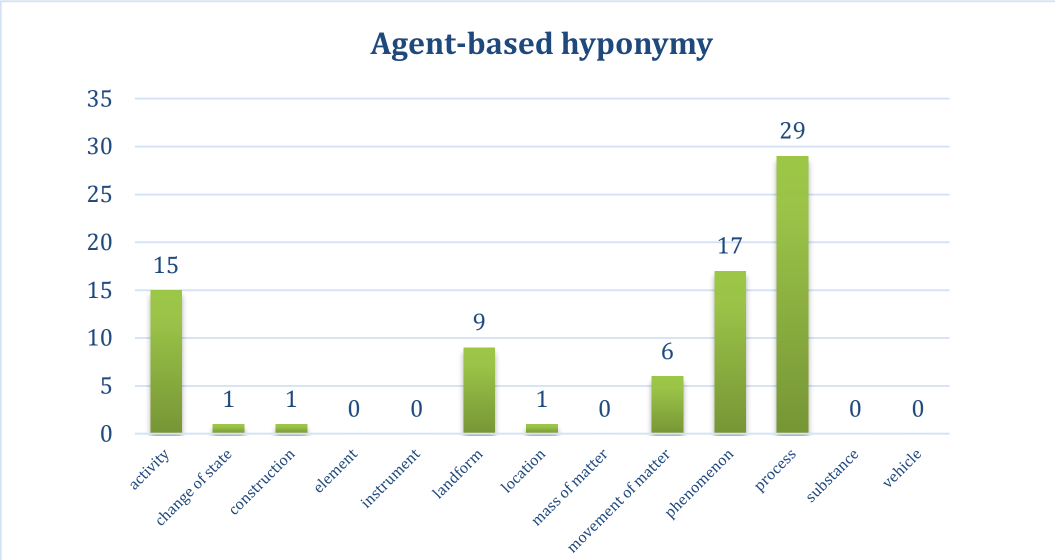



Figure 22. Agent-based hyponymy subtypes per semantic category

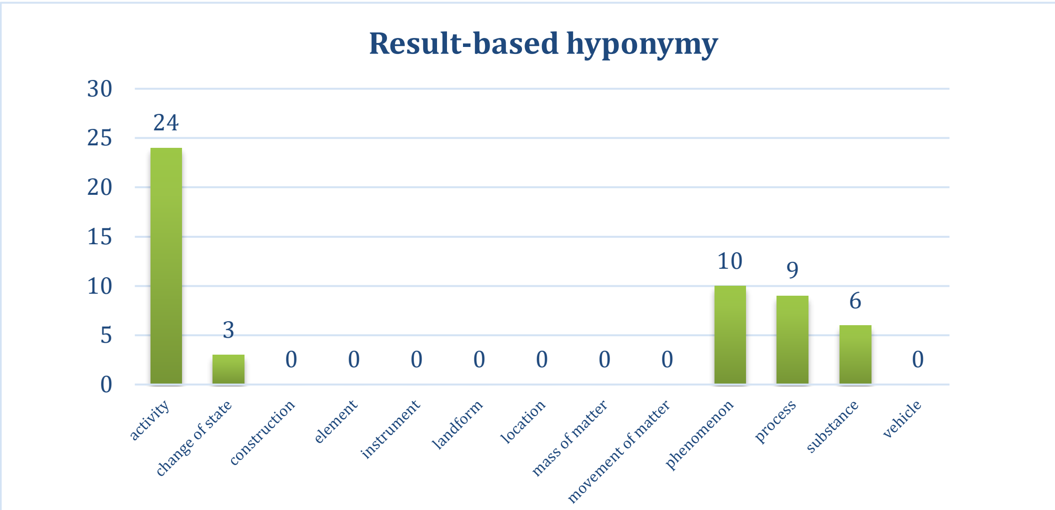



Figure 23. Result-based hyponymy subtypes per semantic category

Interestingly, these hyponymy subtypes also included 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, etc.); (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, etc.).

Similarly, degree-based hyponymy (Figure 24) was also mostly exclusive to process-related semantic categories, such as phenomenon (e.g. MAJOR STORM), activity (e.g. LARGE-SCALE EXPLOITATION), process (e.g. CHRONIC EROSION), and change of state (e.g. PARTIAL MELTING). Furthermore, and in contrast to the results of the previous hyponymy subtypes, the category of phenomenon was mostly characterized by degree (e.g. CATACLYSMIC ERUPTION, LOW-MAGNITUDE EARTHQUAKE, KILLER TORNADO, etc.).

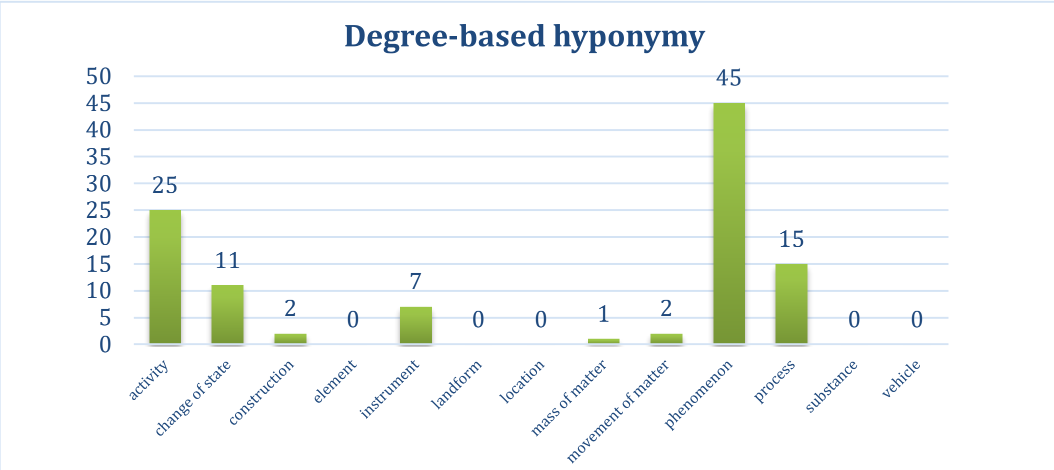



Figure 24. Degree-based hyponymy subtypes per semantic category

Composition-based hyponymy (Figure 25) showed that the most recurrent semantic categories were those involving natural entities, namely substance (e.g. INORGANIC SOLID) and chemical element (e.g. SULPHUR COMPOUND). These were 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.).

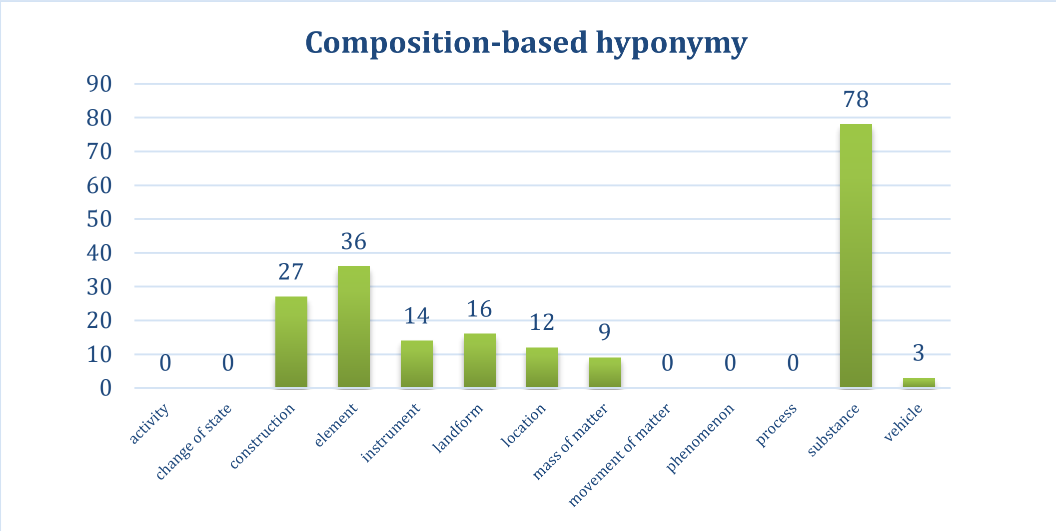



Figure 25. Composition-based hyponymy subtypes per semantic category

Location-based hyponymy (Figure 26) typically occurred with entity-related categories such as substance (e.g. OCEAN WATER), construction (e.g. SPACE STATION), mass of matter (e.g. TROPICAL OCEAN), and landform (e.g. MOUNTAIN CANYON). 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.).

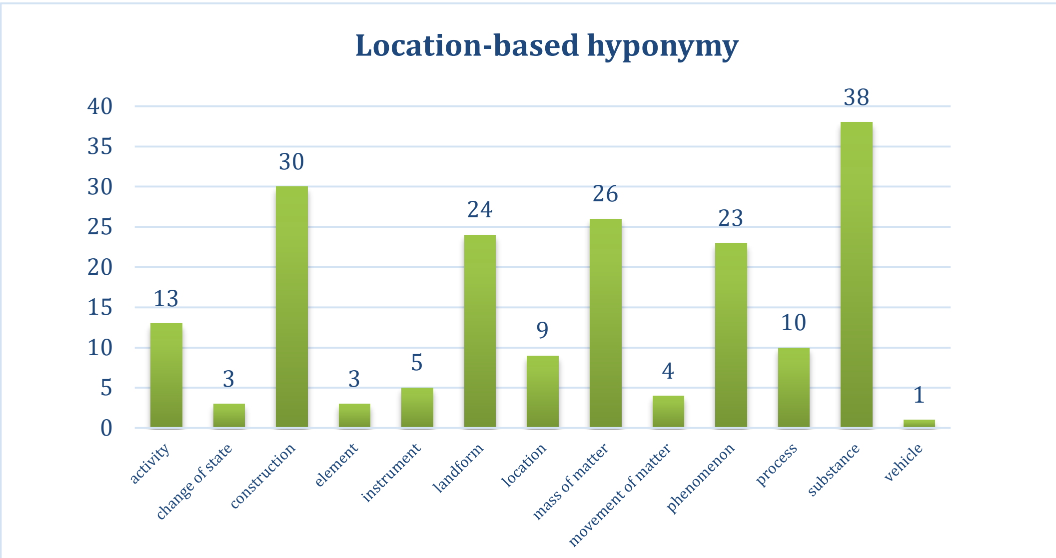



Figure 26. Location-based hyponymy subtypes per semantic category

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 (e.g. RECORDING BAROMETER // DIGITAL FILTER), vehicle (e.g. RECREATIONAL BOAT // ELECTRIC VEHICLE), and construction (e.g. PRODUCTION FACILITY // HYDROELECTRIC DAM). However, and rather surprisingly, construction, which was 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). On the other hand, it is interesting to see how function-based hyponymy also revealed a small amount of process-related semantic categories, namely artificial processes expressed in the form of activity (e.g. EXPLORATION DRILLING, MANUFACTURING INDUSTRY, COMMERCIAL EXPLOITATION), since they are characterized by the purpose or the final aim to be achieved in that event.

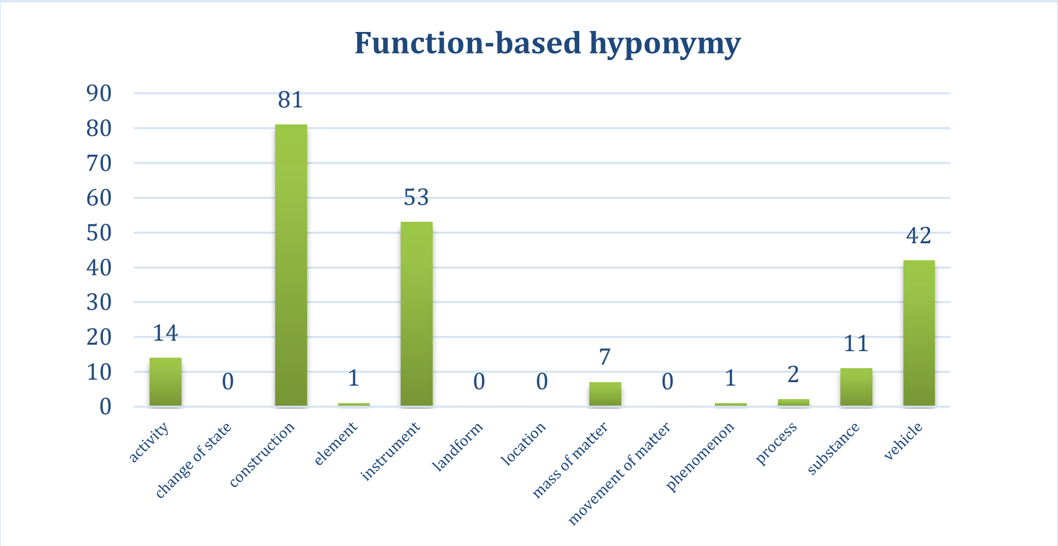



Figure 27. Function-based hyponymy subtypes per semantic category

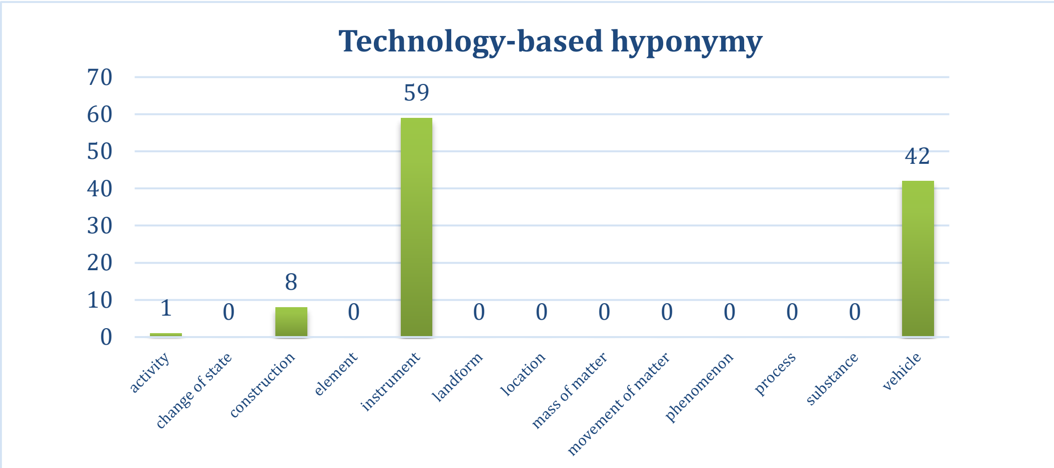



Figure 28. Technology-based hyponymy subtypes per semantic category

Regarding denomination-based hyponymy (Figure 29), almost all of the semantic categories activated were entities: landform (e.g. BIG THOMPSON CANYON), location (e.g. NEW YORK CITY), mass of matter (e.g. ATLANTIC OCEAN), construction (e.g. VOSTOK STATION), and instrument (e.g. DOPPLER RADAR). 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).

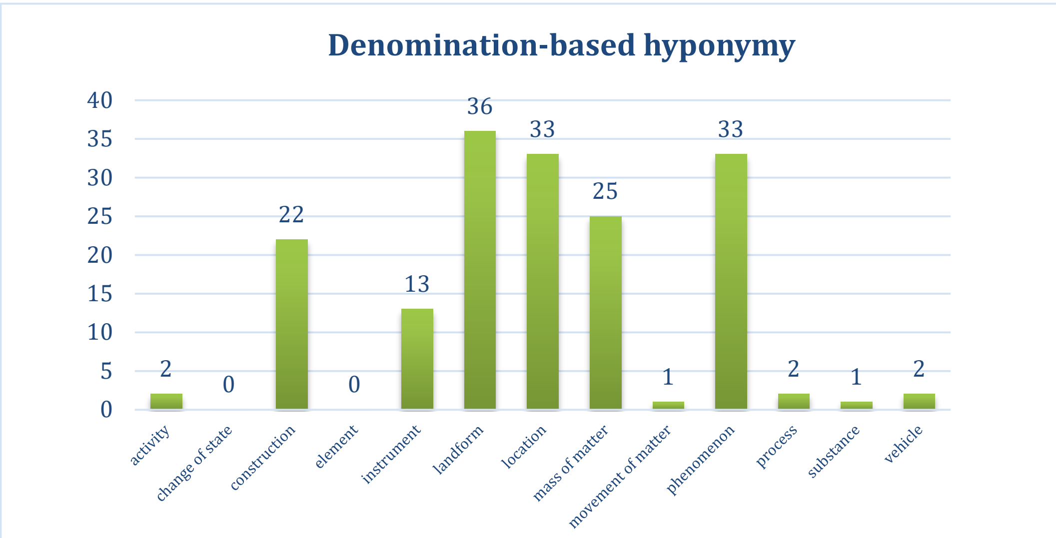



Figure 29. Denomination-based hyponymy subtypes per semantic category

Time-based hyponymy (Figure 30) was related to natural semantic categories, which were both processes such as phenomenon (e.g. WINTER STORM) or movement of matter (e.g. EQUINOCTIAL TIDE), and entities such as substance (e.g. YOUNG ROCK) or mass of matter (e.g. PRIMITIVE ASTEROID). In fact, time is also a natural factor that affects the environmental domain and, overall, phenomena (e.g. SUMMER PRECIPITATION, LATE-SEASON HURRICANE, PERIODIC DROUGHT). However, it rarely occurred with artificial concepts, but there were still some interesting results in the semantic category of construction (e.g. PREHISTORIC SETTLEMENT).

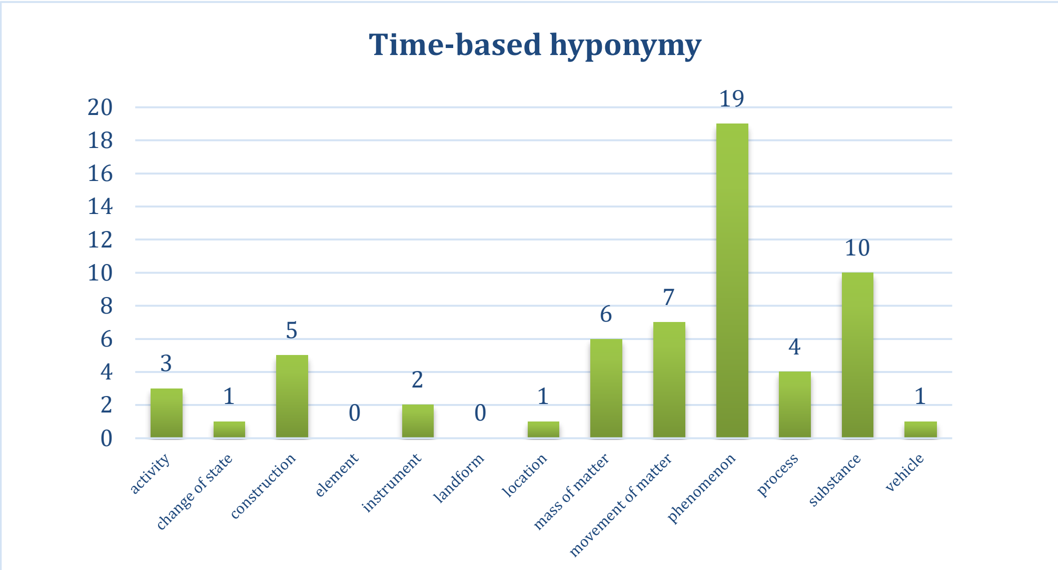



Figure 30. Time-based hyponymy subtypes per semantic category

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

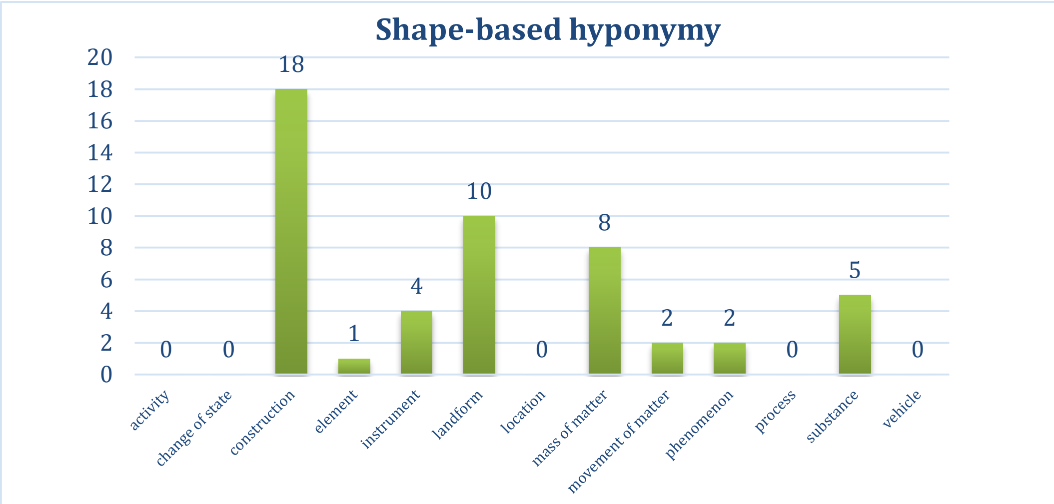



Figure 31. Shape-based hyponymy subtypes per semantic category

The following figures and comments correspond to the 20 remaining hyponymy subtypes, derived from 330 hypernym-hyponym pairs (17% of the sample). Since the amount of information is significantly lower in comparison to the hyponymy subtypes that have been previously analyzed, no definitive conclusions can be drawn from the preliminary results and they will be explored in future research. Nonetheless, and despite the lack of conclusive results, it is interesting to see how certain correlations tended to occur between these hyponymy subtypes and certain semantic categories. These subtypes are origin-based, ability-based, status-based, state-based, size-based, domain-based, amount-based, effect-based, temperature-based, color-based, speed-based, moisture-based, activity-based, height-based, movement-based, density-based, texture-based, weight-based, hardness-based, and relation-based.

Origin-based hyponymy (Figure 32) was shown to be almost exclusive to natural entities, involving the semantic category of substance (e.g. PINE WOOD), but also landform (e.g. DEPOSITIONAL LANDFORM), chemical element (e.g. NATIVE METAL) and mass of matter (e.g. ARTIFICIAL DUNE). Surprisingly, there was only one artificial entity, a construction (e.g. ARTIFICIAL DOCK), and two natural processes (e.g. NATURAL CYCLE, ALLERGIC REACTION), so this subtype appeared to be bound to natural entities.

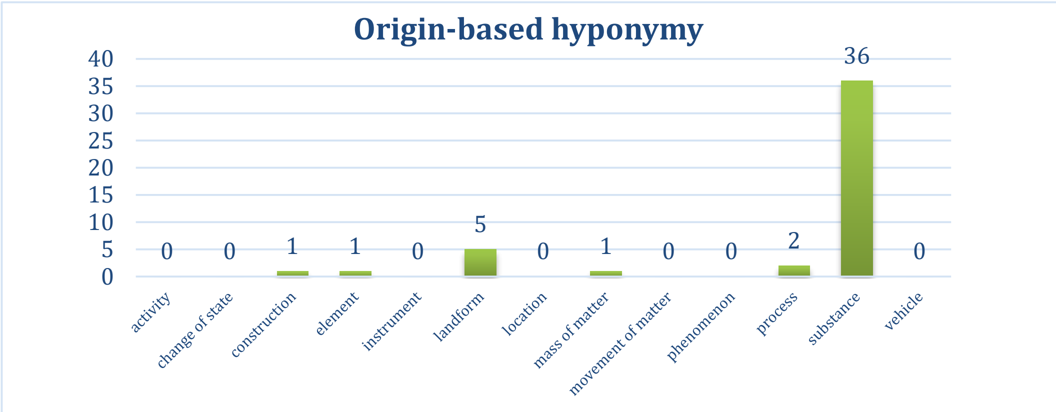



Figure 32. Origin-based hyponymy subtypes per semantic category

In a similar way, amount-based hyponymy (Figure 33), color-based hyponymy (Figure 34), and effect-based hyponymy (Figure 35) were also shown to be almost exclusive to natural entities, frequently being involved to the semantic categories of substance (e.g. TRACE GAS // AMBER LIQUID // POISONOUS SUBSTANCE) and chemical element (e.g. RARE METAL // SILVERY METAL // TOXIC METAL). However, different categories of concepts were also found in the results of each subtype, as artificial entity (e.g. COLLECTIVE FARM) and natural process (e.g. SOLITARY WAVE, SINGLE STORM) in amount-based hyponymy, artificial entity (e.g. GREEN BUILDING) and natural process (e.g. RED TIDE) in color-based hyponymy, and artificial entity (e.g. EFFECTIVE WEAPON) in effect-based hyponymy.

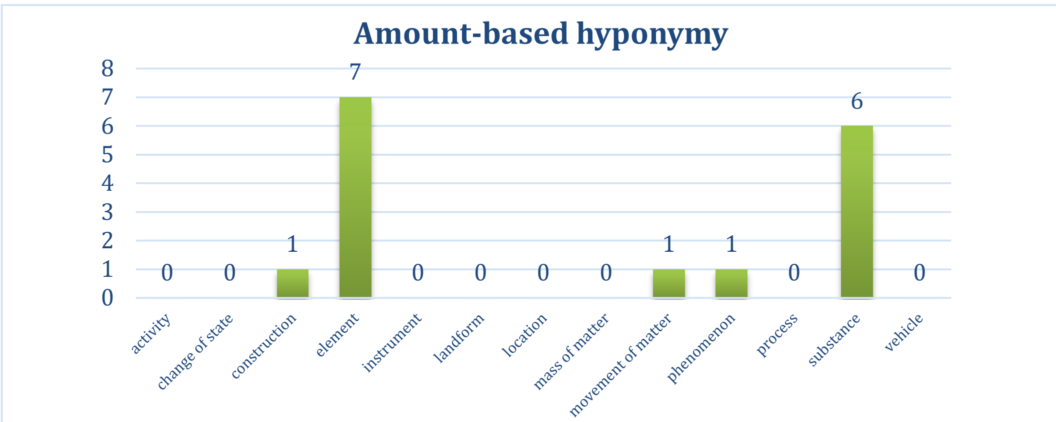



Figure 33. Amount-based hyponymy subtypes per semantic category

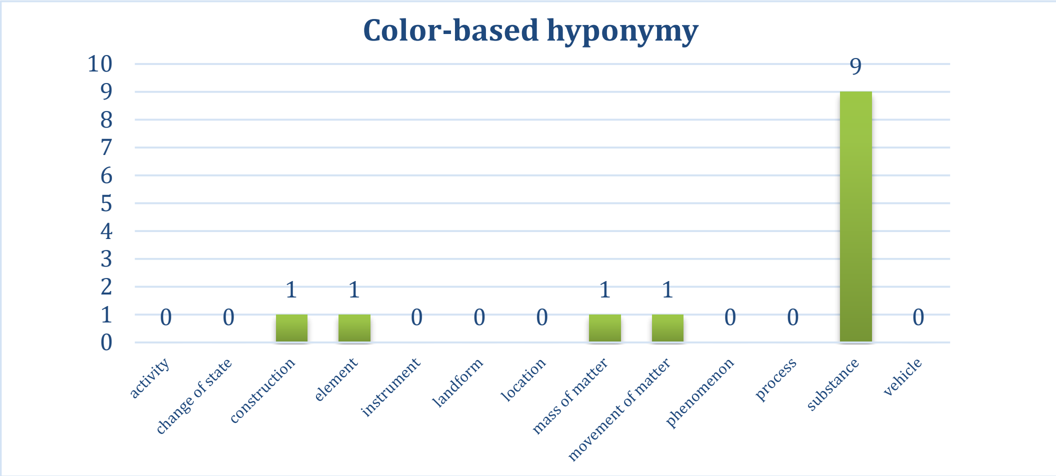



Figure 34. Color-based hyponymy subtypes per semantic category

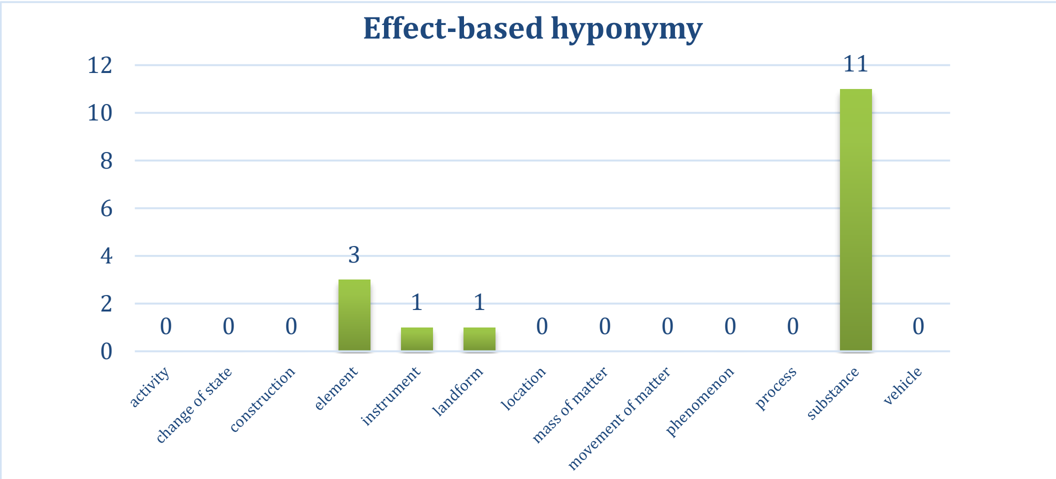



Figure 35. Effect-based hyponymy subtypes per semantic category

The seven following hyponymy subtypes were completely exclusive to natural entities: activity-based hyponymy (Figure 36), density-based hyponymy (Figure 37), moisture-based hyponymy (Figure 38), state-based hyponymy (Figure 39), relation-based hyponymy (Figure 40), temperature-based hyponymy (Figure 41), and texture-based hyponymy (Figure 42). Activity-based hyponymy was characterized by substance (e.g. RADIOACTIVE SUBSTANCE), chemical element (e.g. ALKALI METAL), and mass of matter (e.g. ACTIVE DUNE); density-based hyponymy was represented by substance (e.g. COMPRESSED AIR), element (e.g. HEAVY METAL), and location (e.g. DENSE FOREST); moisture-based hyponymy was manifested in substance (e.g. SATURATED SOIL), mass of matter (e.g. DRY LAND),

and location (e.g. ARID DESERT); state-based hyponymy was linked to substance (e.g. MOLTEN ROCK), chemical element (e.g. VOLATILE COMPOUND), and mass of matter (e.g. GASEOUS PLANET); relation-based hyponymy was found in element (e.g. PARENT COMPOUND) and substance (e.g. FOREIGN SUBSTANCE); temperature-based hyponymy was related to substance (e.g. HOT GAS), mass of matter (e.g. WARM OCEAN), and location (e.g. COLD DESERT); and texture-based hyponymy was only bound to substance (e.g. COARSE SAND).

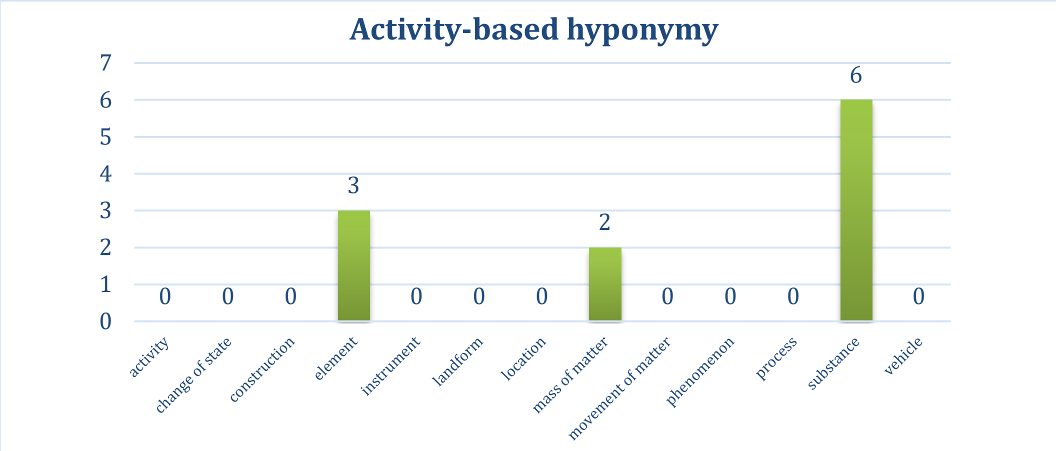



Figure 36. Activity-based hyponymy subtypes per semantic category

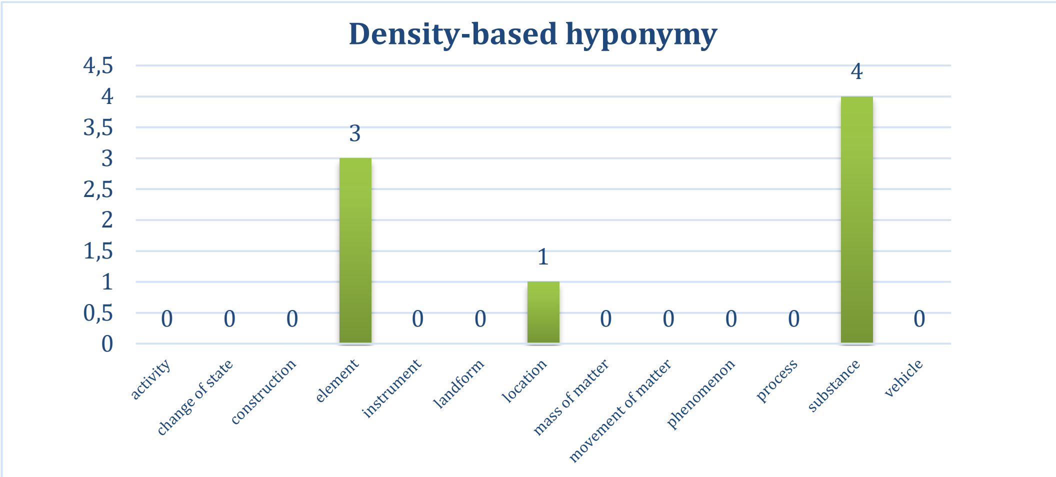



Figure 37. Density-based hyponymy subtypes per semantic category

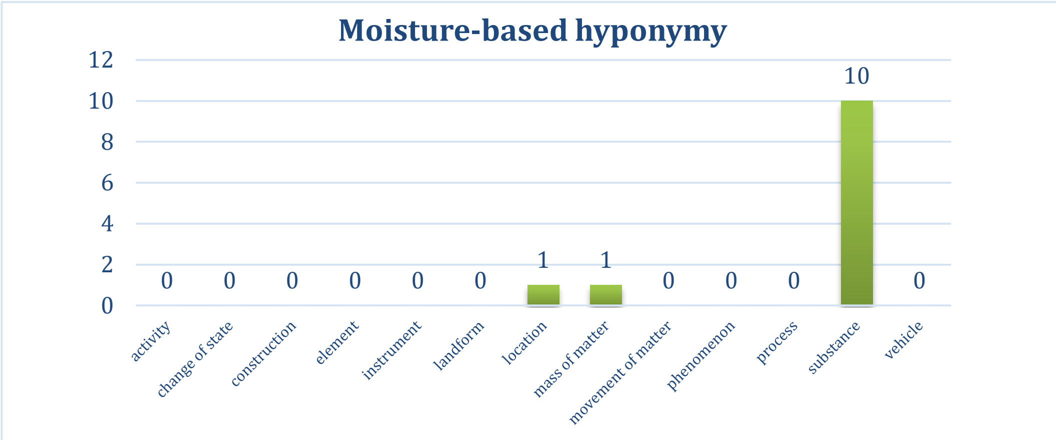



Figure 38. Moisture-based hyponymy subtypes per semantic category

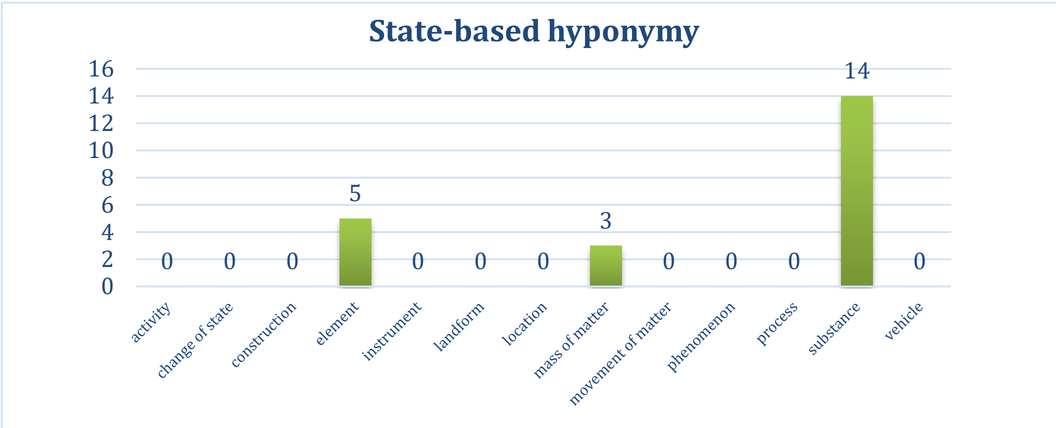



Figure 39. State-based hyponymy subtypes per semantic category

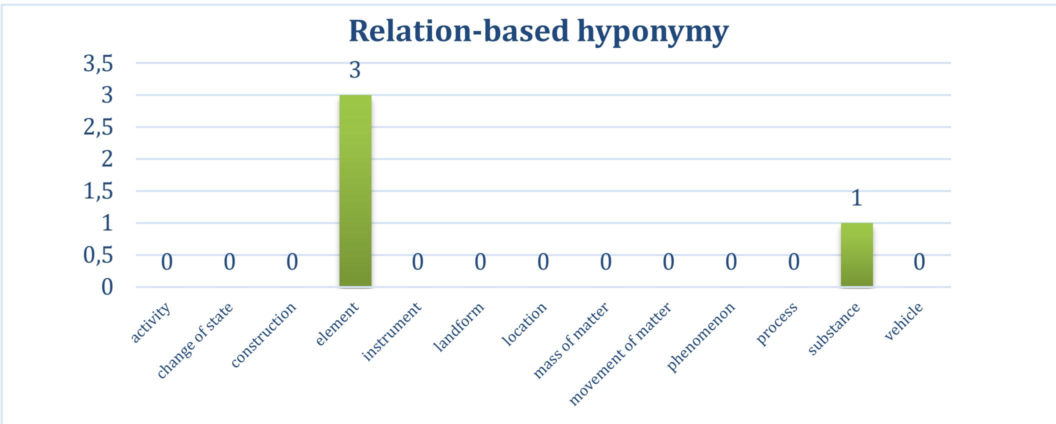



Figure 40. Relation-based hyponymy subtypes per semantic category

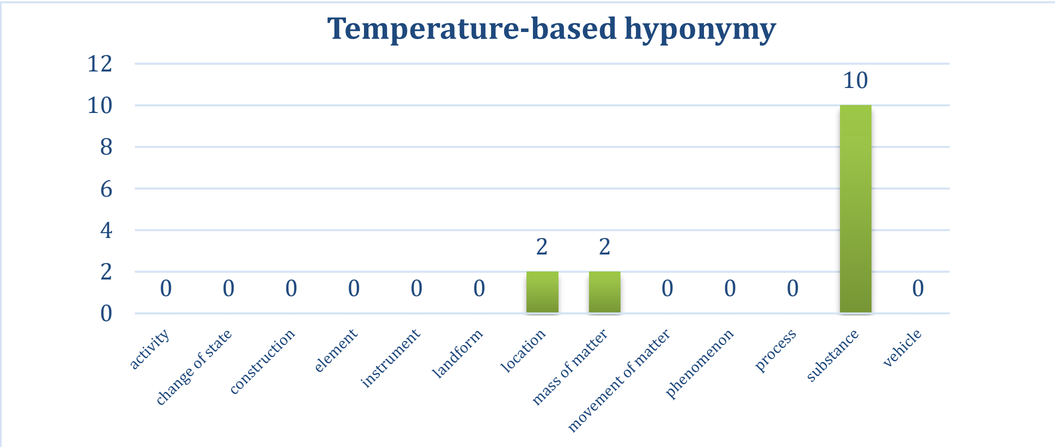



Figure 41. Temperature-based hyponymy subtypes per semantic category

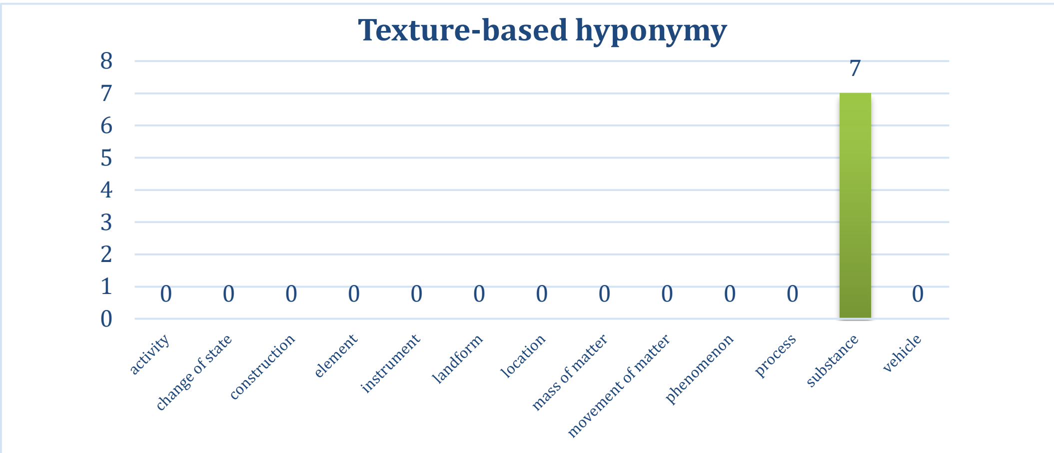



Figure 42. Texture-based hyponymy subtypes per semantic category

As for size-based hyponymy (Figure 43), this subtype was exclusive to entity-related semantic categories. The results showed both artificial entities, namely vehicle (e.g. COMPACT CAR) and construction (e.g. MAJOR CITY); and natural entities, especially mass of matter (e.g. DWARF PLANET), substance (e.g. TINY CRYSTAL), and landform (e.g. SMALL-SCALE LANDFORM). Interestingly, size-based hyponymy appears to be mostly common of large entities, rather than small entities or elements.

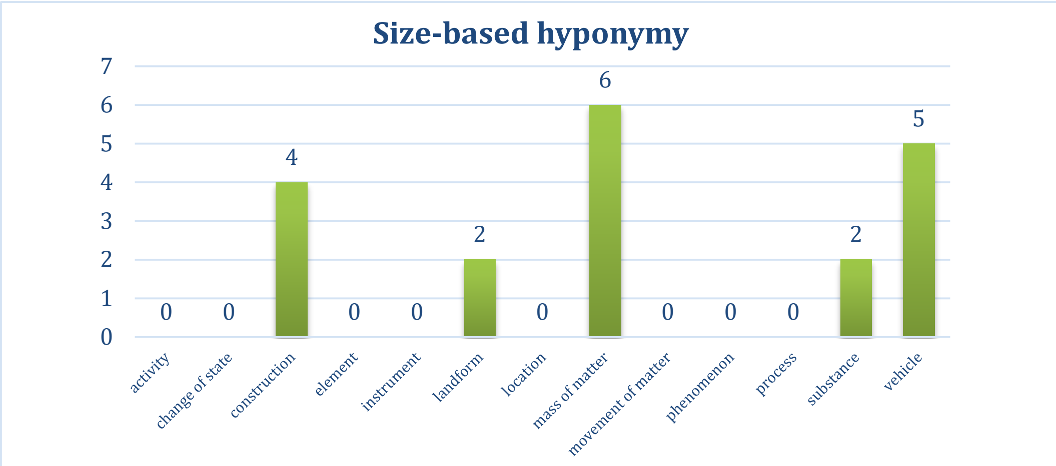



Figure 43. Size-based hyponymy subtypes per semantic category

Status-based hyponymy (Figure 44) was mostly exclusive of entity-related semantic categories. These categories involved both artificial entities, primarily construction (e.g. PUBLIC BUILDING), vehicle (e.g. CONCEPT CAR), and instrument (e.g. UNPROTECTED THERMOMETER); and natural entities, characteristically substance (e.g. UNTREATED WOOD), mass of matter (e.g. UNDEVELOPED LAND), and chemical element (e.g. INCOMPATIBLE ELEMENT). However, unlike the previous size-based hyponymy, this subtype also included phenomenon (e.g. CONTROLLED FIRE), which is a process-related semantic category.

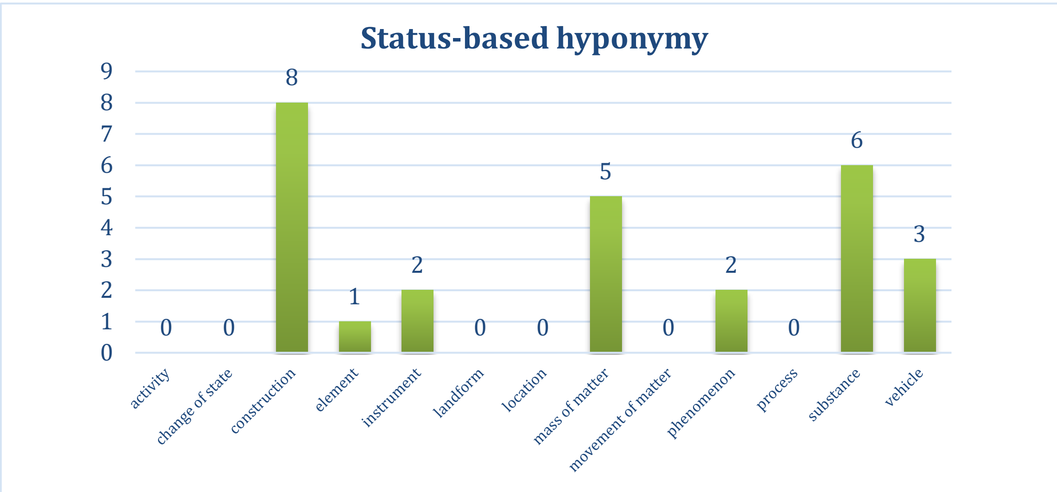



Figure 44. Status-based hyponymy subtypes per semantic category

Apart from hyponymy subtypes mainly related to natural entities, movement-based hyponymy (Figure 45) was mostly associated with natural processes, namely movement of matter (e.g. EBB TIDE) and process (e.g. OUTGOING RADIATION). However, the results found were insufficient to conclude that this subtype was exclusive to natural processes, since there were also examples of natural entities (e.g. INCOMING ASTEROID) and artificial entities (e.g. OCEAN-GOING DREDGE).

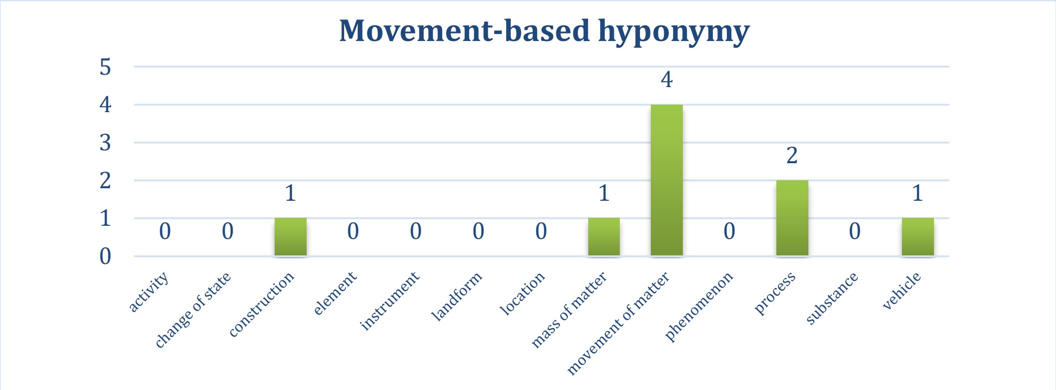



Figure 45. Movement-based hyponymy subtypes per semantic category

Similarly, speed-based hyponymy (Figure 46) was almost exclusive to process-related semantic categories, both natural processes such as process (e.g. SPONTANEOUS DECOMPOSITION) or change of state (e.g. FLASH EVAPORATION), and artificial processes such as activity (e.g. RAPID RECYCLING). Nevertheless, once again it was impossible to state that this subtype is exclusive to processes, since more results would have been needed, and since also found were artificial entities such as instrument (e.g. SLOW FILTER).

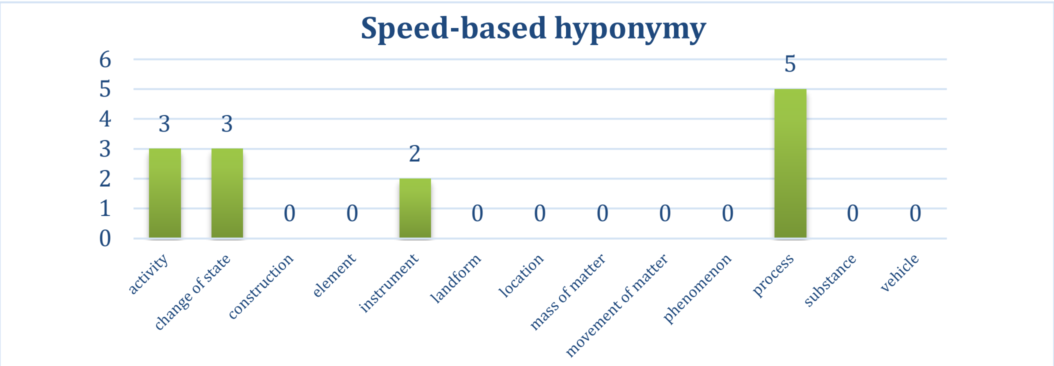



Figure 46. Speed-based hyponymy subtypes per semantic category

Ability-based hyponymy (Figure 47), domain-based hyponymy (Figure 48), and height-based hyponymy (Figure 49) were identified as polyvalent subtypes that appeared in relation to different kinds of semantic categories without special preferences. Ability-based hyponymy was linked to natural entities like substance (e.g. INFLAMMABLE LIQUID), artificial entities like construction (e.g. ADJUSTABLE GROIN), and natural processes like phenomenon (e.g. INVISIBLE DROUGHT).

Domain-based hyponymy was associated with artificial entities like instrument (e.g. SCIENTIFIC INSTRUMENT), natural processes like phenomenon (e.g. AGRICULTURAL DROUGHT), and artificial processes like activity (e.g. TOURISM INDUSTRY). Height-based hyponymy tended to occur in the case of natural entities like substance (e.g. DEEP WATER), natural processes like movement of matter (e.g. HIGH TIDE), and artificial entities like construction (e.g. HIGH-RISE BUILDING).

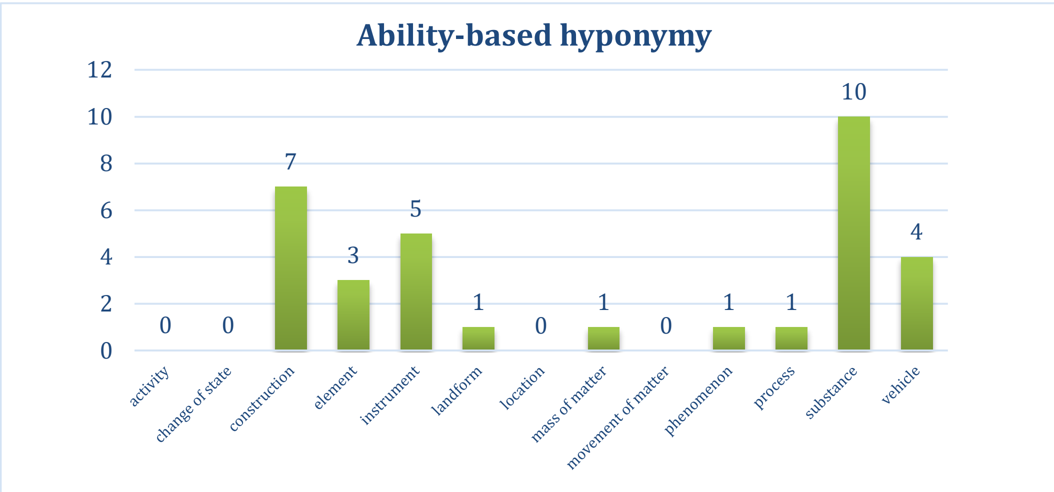



Figure 47. Ability-based hyponymy subtypes per semantic category

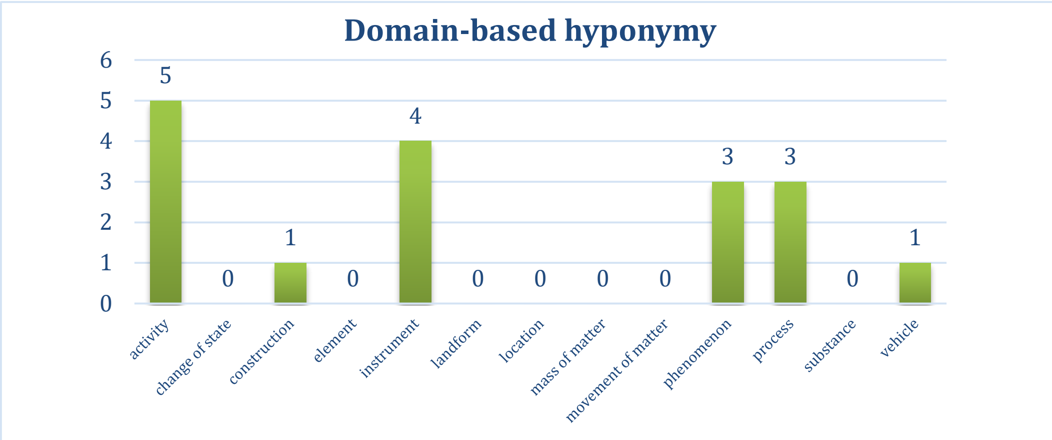



Figure 48. Domain-based hyponymy subtypes per semantic category

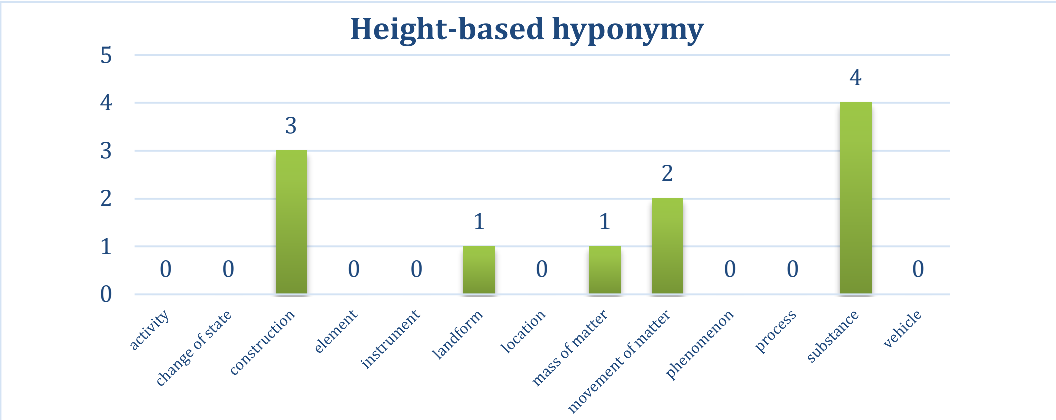



Figure 49. Height-based hyponymy subtypes per semantic category

Finally, the information obtained in these preliminary results for weight-based hyponymy (Figure 50) and hardness-based hyponymy (Figure 51) was not sufficient to draw any definitive conclusions because of the low number of concordances. On the one hand, weight-based hyponymy was only described in relation to vehicle (e.g. LIGHT-DUTY TRUCK), an artificial entity. On the other hand, hardness-based hyponymy was referred to substance (e.g. SOFT WOOD), a natural entity, and to construction (e.g. HARD STRUCTURE), an artificial entity. Nevertheless, these hyponymy subtypes were attributional and, similarly to previous attributional subtypes (e.g. density-based, moisture-based, temperature-based, etc.), exclusive to entity-related semantic categories.

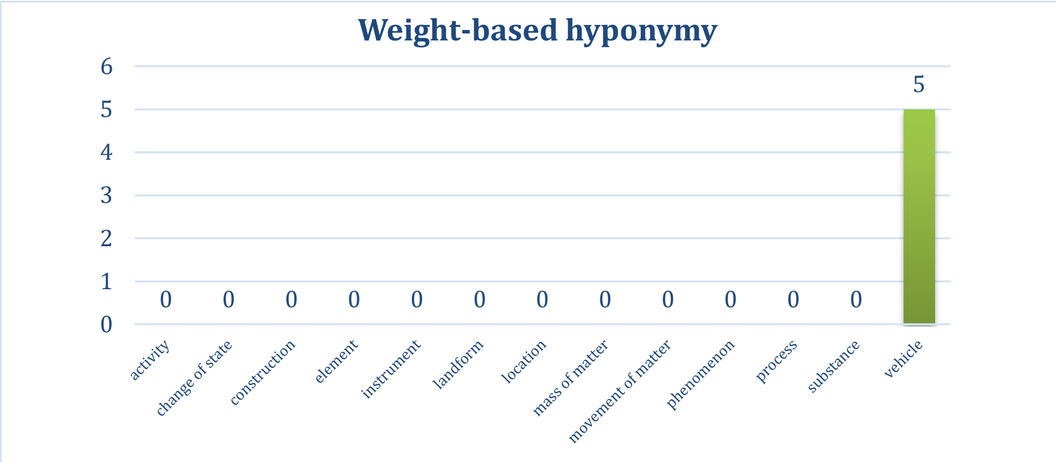



Figure 50. Weight-based hyponymy subtypes per semantic category

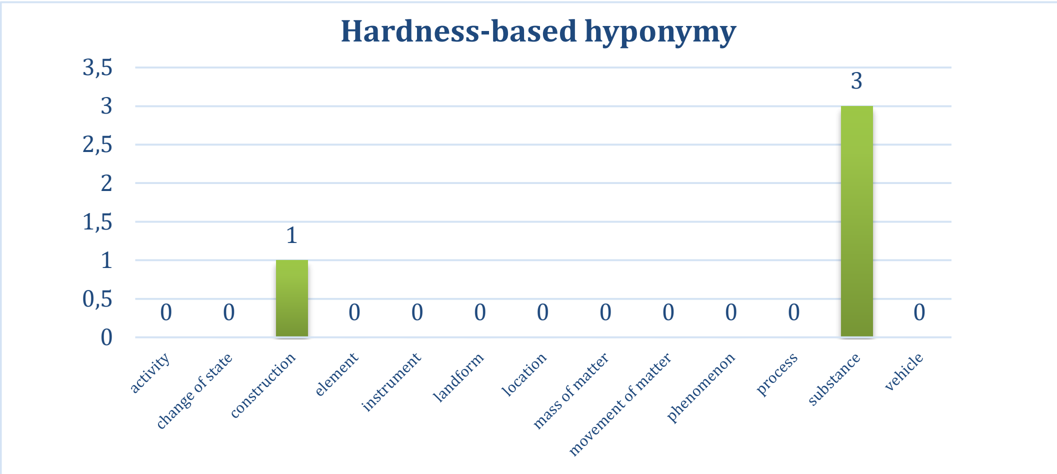



Figure 51. Hardness-based hyponymy subtypes per semantic category

## 5. Conclusion

This work has presented a preliminary corpus-based analysis of the behavior of hyponymic relations in the environmental domain with regard to concept nature. The research involved an analysis of the correlations between hyponymic knowledge patterns (KPs) and semantic categories, and the correlations between hyponymy subtypes and semantic categories.

In contrast to the General Terminology Theory (GTT), which contemplates static conceptual structures, modern theories of terminology emphasize dynamic conceptual representations and knowledge organization. Accordingly, the integration of specialized knowledge representation, cognitive linguistics and cognitive semantics into the study of terminology and specialized communication has enriched this discipline with a more cognitive perspective and given it an interdisciplinary approach. Indeed, in recent years, terminology has demonstrated to be a discipline in constant evolution, thanks to the adoption of innovative proposals that facilitate the organization and representation of information in knowledge structures.

The practical application of contemporary terminology theories has led to the development of comprehensive knowledge resources, such as terminological knowledge bases (TKBs), which facilitate both information display and its retrieval. One of the key elements of these terminological resources are conceptual or semantic relations, cognitive links that describe any kind of interaction between concepts and which result in the creation of conceptual systems. Even though conceptual relations are generally classified as hierarchical or non-hierarchical, they can be further specified.

Of all these relations, this work has focused on generic-specific or hyponymic relations because of their importance in terminology work. Hyponymy (the type_of relation) is the backbone of ontology-based terminological resources because it is the origin of all concept hierarchies. However, generic-specific relations are complex, and thus hypernym-hyponym pairs can be studied from multiple perspectives. This work has shown how hyponymy can be decomposed into subtypes or analyzed based on different 'facets' or 'microsenses'. This makes it possible to obtain a more fine-grained vision of hyponymic relations.

On the other hand, one the most recent approaches to the study of hyponymy and other conceptual relations is corpus analysis. This method has been traditionally used by terminologists to find terms in large corpora and extract their syntactic and semantic information. However, corpus analysis has recently begun to be used as a means to identify knowledge patterns (KPs) that codify semantic relations between terms in knowledge-rich contexts (KRCs). Consequently, much attention is currently focused on the development of semi-automatized procedures for extracting relevant KRCs and, in particular, for extracting hyponymic information.

As shown in the practical part of this work, hyponymy is indeed a complex relation that can be studied by analyzing concept hierarchies. The results obtained in this research study 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.

As can be observed in the preliminary results of this work, 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. Range KPs were mostly associated with time period and measure since these categories are generally composed of values that are characterized by the distance between them in terms of time, space, intensity, etc.

The analysis of hyponymy showed that certain subtypes (e.g. agent-based, patient-based, result-based, method-based, degree-based) closely correlated with process-related semantic categories (e.g. activity, phenomenon, process, change of state). On the other hand, other hyponymy subtypes (e.g. composition-based, technology-based, function-based) were directly linked to entity-related semantic categories (e.g. substance, landform, construction, instrument). Furthermore, the results also demonstrated that a distinction can be made between relational hyponymy subtypes (i.e. those involving another concept, like agent-based, result-based or location-based) and attributional hyponymy subtypes (i.e. those involving the intrinsic characteristics of the concept, like shape-based, texture-based or moisture-based). Relational hyponymy subtypes were generally associated with processes, whilst attributional hyponymy subtypes were mainly related to entities. In addition, further distinctions were made not only depending on whether concepts were entities or processes, but also between natural and artificial concepts. In this line, some hyponymy subtypes were shown to be mostly exclusive to natural concepts (e.g. origin-based, state-based, time-based), whereas other hyponymy subtypes were generally attributed to artificial concepts (e.g. function-based, technology-based, weight-based).

These preliminary results open the door to further studies on hyponymy not only in the environmental domain, but also in regard to specialized knowledge in general. As future research, we plan to analyze the entire English EcoLexicon corpus after a previous revision of the customized hyponymic word sketch grammars in order to reduce repetitions and false positives.

It would also be necessary to study in greater depth the distinction between relational and attributional hyponymy subtypes. Another interesting 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.

## Appendix A. Semantic categories: description and examples

| SEMANTIC CATEGORY | DESCRIPTION | EXAMPLES |
|---|---|---|
| **ACTIVITY** | activities, techniques and behaviors | AGRICULTURE<br>REPRODUCTION<br>LAND USE PLANNING |
| **CHANGE OF STATE** | natural processes involving the change of state of a certain matter | ICE MELTING<br>FLASH EVAPORATION<br>SNOW SUBLIMATION |
| **CHEMICAL ELEMENT** | chemical elements and compounds | CHLOROFLUOROCARBON<br>MERCURY<br>NICOTINAMIDE |
| **CONSTRUCTION** | man-made buildings and structures | TOWER MILL<br>BREAKWATER<br>PIPELINE |
| **DISEASE** | illnesses and conditions | BLACK LUNG DISEASE<br>CANCER<br>MALARIA |
| **DOMAIN** | scientific or knowledge fields | BIOLOGY<br>METEOROLOGY<br>COASTAL ENGINEERING |
| **FEATURE** | properties, characteristics and variables | SOIL MOISTURE<br>BODY SIZE<br>DENSITY |
| **FORCE** | types of energy | HEAT WAVE<br>SOLAR ENERGY<br>ELECTRICITY |
| **INFORMATION** | documents and data | CLIMOGRAPH<br>BIOLOGICAL CLASSIFICATION<br>BATHYMETRIC CHART |
| **INSTRUMENT** | man-made inventions or creations used as instruments | MONITORING INSTRUMENT<br>DIGITAL BAROMETER<br>SAND FILTER |
| **LANDFORM** | geographical and geological features | ISLAND<br>KARST<br>MOUNTAIN |
| **LIFEFORM** | living beings or organisms | SEABIRD<br>MANGROVE TREE<br>PROTIST |
| **LOCATION** | spatial environments | MARINE BIOME<br>TROPICAL RAIN FOREST<br>EUROPE |
| **MASS OF MATTER** | massive entities composed of certain substances | PLANET<br>OCEAN<br>GLACIER |
| **MEASURE** | measuring units | CELSIUS<br>HORSEPOWER<br>KILOMETER |
| **MOVEMENT OF MATTER** | types of mass movement | EBBING TIDE<br>LANDSLIDE<br>MUDFLOW |
| **PERIOD** | time periods or spans | MONTH<br>SEASON<br>HOUR |
| **PHENOMENON** | meteorological and geological phenomena | TSUNAMI<br>RAIN<br>VOLCANIC ERUPTION |

| | | |
|---|---|---|
| **PROCESS** | natural and artificial processes with agents and patients | ABRASION<br>WEATHERING<br>GAS ADSORPTION |
| **PRODUCT** | natural and artificial substances that are the result of a process | GLASSWARE<br>DEODORANT<br>COFFEE |
| **SUBSTANCE** | solid, liquid and gaseous substances or materials | GRANITE<br>FOSSIL FUEL<br>WOOD |
| **SYSTEM** | scientific systems and models | THEORY OF RELATIVITY<br>SCIENTIFIC LAW<br>EMPIRICAL METHOD |
| **TECHNOLOGY** | man-made creations and inventions | GENERATOR<br>AIRCRAFT<br>RADIOSONDE |
| **VEHICLE** | man-made inventions or creations used as vehicles | MOTOR VEHICLE<br>ELECTRIC CAR<br>DELIVERY TRUCK |

Table 4. Semantic categories description table

## Appendix B. Hyponymic knowledge patterns: description and patterns

| HYPONYMIC KP TYPE | DESCRIPTION | PATTERNS | # |
|---|---|---|---|
| **CLASSIFICATION** | they classify or divide the hypernym into hyponyms | **1.** HYPER *be* classified into HYPO | **13** |
| | | **2.** HYPER *be* divided into HYPO | **1** |
| | | **3.** HYPO *be* classified in HYPER | **2** |
| | | **4.** types of HYPER *be* classified as HYPO | **3** |
| | | **5.** HYPER *be* classified into types, namely HYPO | **1** |
| **DEFINITION** | they introduce the hyponym with a definition where the hypernym is the *genus* | **1.** HYPO *be* defined as HYPER | **1** |
| | | **2.** HYPO, defined as HYPER | **2** |
| | | **3.** HYPO: a HYPER | **31** |
| | | **4.** HYPO: a type of HYPER | **9** |
| **DENOMINATION** | they introduce the hyponyms as particular denominations | **1.** a HYPER called HYPO | **5** |
| | | **2.** a type of HYPER called HYPO | **60** |
| | | **3.** a type of HYPER known as HYPO | **40** |
| | | **4.** a type of HYPER referred to as HYPO | **2** |
| | | **5.** a type of HYPER termed HYPO | **3** |
| | | **6.** a type of HYPER, called HYPO | **19** |
| | | **7.** a type of HYPER, known as HYPO | **10** |
| | | **8.** a type of HYPER, named a HYPO | **1** |
| | | **9.** a type of HYPER, termed HYPO | **1** |
| | | **10.** HYPO refers to HYPER | **3** |
| | | **11.** types of HYPER *be* called HYPO | **45** |
| | | **12.** types of HYPER *be* known as HYPO | **10** |
| | | **13.** types of HYPER *be* referred to as HYPO | **8** |
| **ENUMERATION** | they show an exhaustive and numbered list of hyponyms for the hypernym | **1.** # HYPER: HYPO | **1** |
| | | **2.** # types of HYPER *be* considered: HYPO | **3** |
| | | **3.** # types of HYPER *be* distinguished: HYPO | **3** |
| | | **4.** # types of HYPER *be* HYPO | **10** |
| | | **5.** # types of HYPER *be* identified: HYPO | **2** |
| | | **6.** # types of HYPER *be* recognized: HYPO | **1** |
| | | **7.** # types of HYPER *be* required: HYPO | **1** |
| | | **8.** # types of HYPER *be*: HYPO | **3** |
| | | **9.** # types of HYPER occur: HYPO | **1** |
| | | **10.** # types of HYPER: HYPO | **48** |
| | | **11.** of HYPER, # types *be* HYPO | **1** |
| | | **12.** a type of HYPER, divided into # types: HYPO | **1** |
| | | **13.** HYPER *be* classified in # types, HYPO | **1** |
| | | **14.** HYPER *be* classified into # types, HYPO | **1** |
| | | **15.** HYPER *be* classified into # types: HYPO | **29** |
| | | **16.** HYPER *be* divided into # types, HYPO | **1** |
| | | **17.** HYPER *be* divided into # types: HYPO | **16** |
| | | **18.** # types of HYPER called HYPO | **2** |
| | | **19.** # examples of HYPER *be* HYPO | **3** |
| | | **20.** # types of HYPER *be* included: HYPO | **1** |
| | | **21.** # types of HYPER, ranging from HYPO to HYPO | **1** |
| | | **22.** # main types of HYPER *be* HYPO | **17** |
| | | **23.** # main types of HYPER, namely HYPO | **2** |
| | | **24.** # main types of HYPER: HYPO | **3** |
| | | **25.** # types of HYPER, namely HYPO | **7** |
| | | **26.** # types of HYPER, typically HYPO | **1** |

| | | | |
|---|---|---|---|
| **EXEMPLIFICATION** | they present the hyponyms as examples, types or kinds of the hypernym | **1.** examples of HYPER *be* HYPO | **9** |
| | | **2.** HYPER (e.g. HYPO) | **1** |
| | | **3.** HYPER (HYPO) | **1** |
| | | **4.** HYPER (such as HYPO) | **1** |
| | | **5.** HYPER like HYPO | **1** |
| | | **6.** HYPER species (such as HYPO) | **3** |
| | | **7.** HYPER species such as HYPO | **18** |
| | | **8.** HYPER species, such as HYPO | **27** |
| | | **9.** HYPER such as HYPO | **1025** |
| | | **10.** HYPER types (such as HYPO) | **6** |
| | | **11.** HYPER types such as HYPO | **29** |
| | | **12.** HYPER types, such as HYPO | **26** |
| | | **13.** HYPER, such as HYPO | **332** |
| | | **14.** HYPO *be* an example of a HYPER | **1** |
| **IDENTIFICATION** | they directly link the hyponym to the hypernym with a copulative verb | **1.** a HYPER *be* HYPO | **3** |
| | | **2.** a type of HYPER *be* a HYPO | **37** |
| | | **3.** a type of HYPER, a HYPO | **1** |
| | | **4.** HYPO *be* a HYPER | **170** |
| | | **5.** HYPO *be* a type of HYPER | **13** |
| | | **6.** HYPO, a type of HYPER | **4** |
| | | **7.** other HYPER *be* HYPO | **3** |
| | | **8.** types of HYPER *be* HYPO | **4** |
| **INCLUSION** | they present the hyponyms as concepts included in the notion of the hypernym | **1.** a type of HYPER that includes HYPO | **4** |
| | | **2.** among HYPER *be* HYPO | **1** |
| | | **3.** HYPER include HYPO | **4** |
| | | **4.** HYPER including HYPO | **109** |
| | | **5.** HYPER species including HYPO | **4** |
| | | **6.** HYPER species, including HYPO | **16** |
| | | **7.** HYPER type, which includes HYPO | **2** |
| | | **8.** HYPER types include HYPO | **39** |
| | | **9.** HYPER types including HYPO | **6** |
| | | **10.** HYPER types, including HYPO | **13** |
| | | **11.** HYPER, including HYPO | **14** |
| | | **12.** included in this type of HYPER *be* HYPO | **1** |
| **ITEMIZATION** | they introduce a non-exhaustive list of hyponyms for the hypernym | **1.** HYPO and other HYPER | **135** |
| | | **2.** HYPO and other HYPER species | **6** |
| | | **3.** HYPO and other HYPER types | **5** |
| | | **4.** HYPO, and other HYPER | **24** |
| | | **5.** HYPO, and other HYPER types | **1** |
| | | **6.** HYPO. These types of HYPER… | **1** |
| | | **7.** types of HYPER: HYPO | **3** |
| | | **8.** types of HYPER *be*: HYPO | **4** |
| | | **9.** HYPO and other HYPER classified as HYPO | **1** |
| | | **10.** HYPO and other HYPER (such as HYPO) | **1** |
| | | **11.** HYPO and other HYPER such as HYPO | **19** |
| | | **12.** HYPO and other HYPER, such as HYPO | **16** |
| | | **13.** HYPO, and other HYPER such as HYPO | **2** |
| | | **14.** HYPER include: HYPO | **3** |
| | | **15.** HYPER types include: HYPO | **5** |
| | | **16.** HYPER types including: HYPO | **1** |
| | | **17.** HYPO and other HYPER (including HYPO) | **2** |
| | | **18.** HYPO and other HYPER including HYPO | **2** |
| | | **19.** HYPO and other HYPER, including HYPO | **10** |

| | | | |
|---|---|---|---|
| **RANGE** | they establish a span where several hyponyms can be found for the same hypernym | **1.** HYPER ranging from HYPO to HYPO | **68** |
| | | **2.** HYPER types ranging from HYPO to HYPO | **4** |
| | | **3.** HYPER types, ranging from HYPO to HYPO | **3** |
| | | **4.** HYPER, ranging from HYPO to HYPO | **7** |
| **SELECTION** | they highlight main or preferred hyponyms for the hypernym | **1.** HYPER (especially HYPO) | **1** |
| | | **2.** HYPER primarily HYPO | **1** |
| | | **3.** HYPER species, especially HYPO | **1** |
| | | **4.** HYPER species, mainly HYPO | **1** |
| | | **5.** HYPER, characteristically HYPO | **1** |
| | | **6.** HYPER, especially HYPO | **226** |
| | | **7.** HYPER, generally HYPO | **4** |
| | | **8.** HYPER, mainly HYPO | **49** |
| | | **9.** HYPER, namely HYPO | **26** |
| | | **10.** HYPER, primarily HYPO | **31** |
| | | **11.** HYPER, typically HYPO | **24** |
| | | **12.** HYPER, usually HYPO | **56** |
| | | **13.** HYPO *be* the main types of HYPER | **2** |
| | | **14.** the main types of HYPER *be* HYPO | **17** |
| | | **15.** types of HYPER, mainly HYPO | **1** |
| | | **16.** types of HYPER, primarily HYPO | **2** |
| | | **17.** types of HYPER, typically HYPO | **4** |
| | | **18.** types of HYPER, usually HYPO | **1** |
| | | **19.** the main example of HYPER *be* HYPO | **2** |
| | | **20.** the main types of HYPER include HYPO | **2** |
| **TOTAL** | | | **3133** |

Table 5. Hyponymic knowledge patterns description table

## Appendix C. Hyponymy subtypes: description and examples

| HYPONYMY SUBTYPE | DESCRIPTION | EXAMPLES | # |
|---|---|---|---|
| **ABILITY-BASED** | hyponyms characterized by own abilities or characteristics | RENEWABLE RESOURCE<br>HABITABLE PLANET<br>AUTONOMOUS VEHICLE | **33** |
| **ACTIVITY-BASED** | hyponyms characterized by the activity or stability of their composition | RADIOACTIVE SUBSTANCE<br>ALKALI METAL<br>ACTIVE DUNE | **11** |
| **AGENT-BASED** | hyponyms characterized by the agent that causes them | STORM TIDE<br>AIR OXIDATION<br>SPRINKLER IRRIGATION | **79** |
| **AMOUNT-BASED** | hyponyms characterized by their amount or quantity | TRACE ELEMENT<br>RARE METAL<br>SINGLE STORM | **16** |
| **COLOR-BASED** | hyponyms characterized by their color | COLORLESS SOLID<br>RED TIDE<br>YELLOW LIQUID | **13** |
| **COMPOSITION-BASED** | hyponyms characterized by their components or by their material | METALLIC ELEMENT<br>CARBONATE SAND<br>PINE FOREST | **195** |
| **DEGREE-BASED** | hyponyms characterized by their degree of intensity, size or consequences | CATACLYSMIC ERUPTION<br>LOW-MAGNITUDE EARTHQUAKE<br>MEGA-SCALE EXTRACTION | **108** |
| **DENOMINATION-BASED** | hyponyms characterized by having a particular denomination with a proper noun | PACIFIC OCEAN<br>SAHARA DESERT<br>NEW YORK CITY | **170** |
| **DENSITY-BASED** | hyponyms characterized by their density or particle concentration | LIGHT ELEMENT<br>DENSE WATER<br>HEAVY METAL | **8** |
| **DOMAIN-BASED** | hyponyms characterized by the scientific or knowledge field to which they belong | AGRICULTURAL PRODUCT<br>MUSICAL INSTRUMENT<br>CHEMICAL INDUSTRY | **17** |
| **EFFECT-BASED** | hyponyms characterized by the effects or consequences that they cause | TOXIC LIQUID<br>HAZARDOUS SUBSTANCE<br>GREENHOUSE GAS | **16** |
| **FUNCTION-BASED** | hyponyms characterized by their function or purpose | DRINKING WATER<br>SURVEILLANCE RADAR<br>MANUFACTURING FACILITY | **212** |
| **HARDNESS-BASED** | hyponyms characterized by their hardness level | SOFT WOOD<br>HARD ROCK<br>HARD STRUCTURE | **4** |
| **HEIGHT-BASED** | hyponyms characterized by their height or depth level | SHALLOW WATER<br>DEEP OCEAN<br>HIGH TIDE | **11** |
| **LOCATION-BASED** | hyponyms characterized by their spatial location or position | OCEAN WATER<br>SURROUNDING AIR<br>TROPICAL STORM | **189** |
| **METHOD-BASED** | hyponyms characterized by the method or the process that they involve | AEROBIC OXIDATION<br>DIRECT SUBLIMATION<br>INDUSTRIAL TREATMENT | **164** |
| **MOISTURE-BASED** | hyponyms characterized by their moisture level | DRY SOLID<br>SATURATED AIR<br>ARID DESERT | **12** |
| **MOVEMENT-BASED** | hyponyms characterized by their movement or direction | EBB TIDE<br>OCEAN-GOING DREDGE<br>OUTGOING RADIATION | **9** |

| | | | |
|---|---|---|---|
| **ORIGIN-BASED** | hyponyms characterized by their origin, i.e. the place where they come from or where they were created | NATURAL RESOURCE<br>PINE WOOD<br>COUNTRY ROCK | **46** |
| **PATIENT-BASED** | hyponyms characterized by the patient that is affected by them | COAST EROSION<br>ICE MELTING<br>WATER TREATMENT | **217** |
| **RELATION-BASED** | hyponyms characterized by being related to other concepts | FOREIGN SUBSTANCE<br>PARENT COMPOUND<br>COVALENT SOLID | **4** |
| **RESULT-BASED** | hyponyms characterized by the result that they cause, or by being the result of a process | TSUNAMIGENIC EARTHQUAKE<br>PAPER INDUSTRY<br>UNIMOLECULAR DECOMPOSITION | **52** |
| **SHAPE-BASED** | hyponyms characterized by their shape | AMORPHOUS SOLID<br>PARABOLIC DUNE<br>L-SHAPED GROIN | **50** |
| **SIZE-BASED** | hyponyms characterized by their size | TINY CRYSTAL<br>GIANT PLANET<br>COMPACT CAR | **19** |
| **SPEED-BASED** | hyponyms characterized by their speed | RAPID EROSION<br>FLASH EVAPORATION<br>SPONTANEOUS DECOMPOSITION | **13** |
| **STATE-BASED** | hyponyms characterized by the state of matter | SOLID SUBSTANCE<br>FLUID ELEMENT<br>MOLTEN ROCK | **22** |
| **STATUS-BASED** | hyponyms characterized by a particular circumstance or situation | REGULATED SUBSTANCE<br>UNTREATED WOOD<br>CONTAMINATED SOIL | **27** |
| **TECHNOLOGY-BASED** | hyponyms characterized by the technology that they use | MOTOR VEHICLE<br>GREEN TECHNOLOGY<br>DIGITAL BAROMETER | **110** |
| **TEMPERATURE-BASED** | hyponyms characterized by their temperature | HOT GAS<br>WARM OCEAN<br>COLD AIR | **14** |
| **TEXTURE-BASED** | hyponyms characterized by their texture | VISCOUS LIQUID<br>FINE SAND<br>SOFT ROCK | **7** |
| **TIME-BASED** | hyponyms characterized by their duration, by their age, or by happening in a particular moment | WINTER ICE<br>OLD ROCK<br>ANNUAL PRECIPITATION | **59** |
| **WEIGHT-BASED** | hyponyms characterized by their weight | LIGHT-DUTY VEHICLE<br>HEAVY-DUTY TRUCK<br>LIGHT TRUCK | **5** |
| **TOTAL** | | | **1912** |

Table 6. Hyponymy subtypes description table

## References


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

Baisa, Vít & Suchomel, Vít (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 (2004a). Building a Concept Hierarchy from Corpus Analysis. *Terminology* 10(2), 241–263. Amsterdam/Philadelphia: John Benjamins.

Barrière, Caroline (2004b). Knowledge-rich Contexts Discovery. In *Proceedings of the 17th Canadian Conference on Artificial Intelligence (AI'2004)*, 187–201. Edmonton, Canada.

Berland, Matthew & Charniak, Eugene (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.

Bielinskiene, Agne; Boizou, Loïc; Kovalevskaite, Jolanta; & Utka, Andrius (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*, 18–26. Amsterdam/Berlin/Tokyo/Washington DC: IOS Press.

Bourigault, Didier & Slodzian, Monique (1999). Pour une terminologie textuelle. *Terminologies Nouvelles* 19, 29–32.

Bowker, Lynne & Pearson, Jennifer (2002). *Working with Specialized Language: A Practical Guide to Using Corpora*. London: Routledge.

Buendía Castro, Miriam (2013). *Phraseology in Specialized Language and its Representation in Environmental Knowledge Resources*. PhD Thesis. Granada: Universidad de Granada.

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

Cabré, María Teresa (2000). Elements for a Theory of Terminology: Towards an Alternative Paradigm. *Terminology* 6(1), 35–57. Amsterdam/Philadelphia: John Benjamins.

Cabré, María Teresa (2003). Theories of Terminology: Their Description, Prescription and Explanation. *Terminology* 9(2), 163–199. Amsterdam/Philadelphia: John Benjamins.

Condamines, Anne (2002). Corpus Analysis and Conceptual Relation Patterns. *Terminology* 8(1), 141–162. Amsterdam/Philadelphia: John Benjamins.

Cruse, D. Alan (1995). Polysemy and Related Phenomena from a Cognitive Linguistic Viewpoint. In P. Saint-Dizier & E. Viegas (eds.), *Computational Lexical Semantics*. Cambridge: Cambridge University Press.

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

Edo Marzá, Nuria (2012). Lexicografía especializada y lenguajes de especialidad: fundamentos teóricos y metodológicos para la elaboración de diccionarios especializados / Specialised Lexicography and Specialised Languages: Theoretical and Methodological Foundations for the Elaboration of Specialised Dictionaries. *Lingüística* 27, 98–114.

Evans, Vyvyan & Green, Melanie (2006). *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.

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

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

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

Faber, Pamela; León-Araúz, Pilar; & Reimerink, Arianne (2014). Representing Environmental Knowledge in EcoLexicon. In E. Bárcena, T. Read, & J. Arhus (eds.), *Languages for Specific Purposes in the Digital Era*, *Educational Linguistics* 19, 267–301. Berlin/Heidelberg: Springer.

Faber, Pamela; León-Araúz, Pilar; & Reimerink, Arianne (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*, 73–80. Portorož, Slovenia.

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

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

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

Gaudin, François (1993). *Pour une socioterminologie: des problèmes pratiques aux pratiques institutionnelles*. Rouen: Publications de l'Université de Rouen.

Gaudin, François (2003). *Socioterminologie: une approche sociolinguistique de la terminologie*. Brussels: Duculot.

Geeraerts, Dirk (2010). *Theories of Lexical Semantics*. Oxford: Oxford University Press.

Gil-Berrozpe, Juan Carlos & Faber, Pamela (2016). Refining Hyponymy in a Terminological Knowledge Base. In *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)*, 8–15. Portorož, Slovenia.

Gil-Berrozpe, Juan Carlos & Faber, Pamela (in press). The Role of Terminological Knowledge Bases in Specialized Translation: The Use of Umbrella Concepts. *Temas actuales en terminología y estudios sobre el léxico*. Granada: Comares.

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

Girju, Roxana; Badulescu, Adriana; & Moldovan, Dan (2003). Learning Semantic Constraints for the Automatic Discovery of Part-Whole Relations. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 1–8. Edmonton, Canada.

Hearst, Marti A. (1998). Automated Discovery of Word-Net Relations. In C. Fellbaum (ed.), *WordNet: An Electronic Lexical Database*, 132–152. Cambridge, MA: MIT Press.

ISO (2000). ISO Standard 704:2000. *Terminology Work: Principles and Methods*. International Organization for Standardization. Geneva, Switzerland.

ISO (2009). ISO Standard 704:2009. *Terminology Work: Principles and Methods*. International Organization for Standardization. Geneva, Switzerland.

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

Kilgarriff, Adam; Rychlý, Pavel; Smrz, Pavel; & Tugwell, David (2004). The Sketch Engine. In G. Williams & S. Vessier (eds.), *Proceedings of the Eleventh EURALEX International Congress*, 105–16. Lorient, France.

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

León-Araúz, Pilar; Reimerink, Arianne; & García Aragón, Alejandro (2013). Dynamism and Context in Specialized Knowledge. *Terminology* 19(1), 31–61. Amsterdam/Philadelphia: John Benjamins.

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

Liu, Lei; Cao, Cungen; & Wang, Haitao (2006). Extracting Hyponymic Relations from Chinese Free Corpus. In *Proceedings of the 5th WSEAS International Conference on Applied Computer Science*, 963-969. Hangzhou, China.

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

Marshman, Elizabeth; Morgan, Tricia; & Meyer, Ingrid (2002). French Patterns for Expressing Concept Relations. *Terminology* 8(1), 1–29. Amsterdam/Philadelphia: John Benjamins.

Meyer, Ingrid (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*, 279–302. Amsterdam/Philadelphia: John Benjamins.

Meyer, Ingrid; Bowker, Lynne; & Eck, Karen (1992). COGNITERM: An Experiment in Building a Knowledge-based Term Bank. In *Proceedings of the Fifth EURALEX International Congress*, 159–172. Tampere, Finland.

Miller, George A. (1998). Nouns in WordNet. In C. Fellbaum (ed.), *WordNet: An Electronic Lexical Database*, 23–46. Cambridge, MA: MIT Press.

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

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

Nuopponen, Anita (2005). Concept Relations: An Update of a Concept Relation Classification. In B.N. Madsen & H.E. Thomsen (eds.), *7th International Conference on Terminology and Knowledge Engineering (TKE 2005)*, 127–138. Copenhagen, Denmark.

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

Rosch, Eleanor (1978). Principles of Categorization. In E. Rosch & B.B. Lloyd (eds.), *Cognition and Categorization*, 27–48. Hillsdale, NJ: Erlbaum.

Sager, Juan Carlos (1990). *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia: John Benjamins.

Schumann, Anne-Kathrin (2012). Knowledge-rich Context Candidate Extraction and Ranking with KnowPipe. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*, 3626–3630. Istanbul, Turkey.

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

Temmerman, Rita (2001). Sociocognitive Terminology Theory. In M.T. Cabré Castellví & J. Feliú (eds.), *Terminología y Cognición*, 75–92. Barcelona: Universitat Pompeu Fabra.

Temmerman, Rita; Kerremans, Koen; & Vandervoort, Veerle (2005). La termontographie en contexte(s). In *Actes des Septièmes Journées Scientifiques du Réseau Lexicologie, Terminologie, Traduction*. Brussels, Belgium.

Winston, Morton E.; Chaffin, Roger; & Herrmann, Douglas (1987). A Taxonomy of Part-Whole Relations. *Cognitive Science*, 11, 417–444.

Wüster, Eugen (1968). *The Machine Tool. An Interlingual Dictionary of Basic Concepts*. London: Technical Press.

Wüster, Eugen (1979). *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*. Wien: Springer. (Translation into Spanish directed by Cabré, María Teresa: *Introducción a la teoría de la terminología y a la lexicografía terminológica*. Barcelona: IULA, 1998.)