UNIVERSIDAD DE GRANADA

FACULTAD DE TRADUCCIÓN E INTERPRETACIÓN

Trabajo de Fin de Grado

Grado en Traducción e Interpretación

Extending the conceptual systems in EcoLexicon to enhance multidimensionality

Juan Carlos Gil Berrozpe

Directora: Pamela Faber

29 / 06 / 2016

ACKNOWLEDGEMENTS

With these concise words, I would like to express my most sincere gratitude and appreciation to all those people who, explicit or implicitly, have encouraged me and lent me a helping hand during the development of this final degree project. I do not have words enough to thank them for helping me give the best of myself from the beginning until the very end of this research.

Of these people, I am grateful in particular to Professor Pamela Faber, director and supervisor of this project, for offering me the opportunity to work with her and her excellent team. Without her wise advice and her valuable guidance, none of this work could have been possible.

Likewise, I would also like to thank Professor Pilar León for her time and support with this research. Indeed, I am sure that her outstanding suggestions have shaped and improved the outcome of this work.

INDEX

Abstract

Multidimensionality is the phenomenon by which the characteristics of a certain concept may vary depending on the perspective taken. With no doubt, the representation of multidimensionality is a major challenge in the design of terminological knowledge bases (TKBs), since extracting a few concepts and establishing simple relations between them results in monodimensional systems. EcoLexicon, based on Frame-Based Terminology (FBT), is a multidimensional and dynamic TKB on environmental science that targets user knowledge acquisition through linguistic, conceptual and graphical information. Despite all the advantages that EcoLexicon provides, its vast amount of information is a double-edged sword that occasionally affects its representation of multidimensional knowledge, causing problems that include information overload, excessive noise and redundancy, and transitivity inconsistencies in conceptual relations. To solve these problems, this final degree project proposes an extension of the conceptual systems in EcoLexicon by refining hyponymy in three ways: (i) correcting property inheritance, (ii) implementing umbrella concepts, and (iii) establishing hyponymy subtypes. Moreover, this project also carries out a process of hyponymic extension through corpus extraction to semi-automatically retrieve hyponyms from the EcoLexicon database using customized word sketches, and subsequently complementing and validating the new hierarchies.

Keywords: conceptual relations, multidimensionality, hyponymy, corpus extraction, terminological knowledge bases

Resumen

La multidimensionalidad es el fenómeno por el cual las características de un determinado concepto pueden variar en función de la perspectiva que se adopte. Sin lugar a dudas, la representación de la multidimensionalidad supone un gran desafío en el diseño de bases de conocimiento terminológico (BCT), ya que el hecho de extraer unos pocos conceptos y establecer relaciones muy simples entre ellos acaba dando lugar a sistemas monodimensionales. EcoLexicon, que se basa en la Terminología Basada en Marcos (TBM), es una BCT multidimensional y dinámica sobre el medio ambiente que busca la adquisición de conocimiento por parte del usuario a través de información lingüística, conceptual y gráfica. A pesar de todos los beneficios que proporciona EcoLexicon, su inmensa cantidad de información es un arma de doble filo que, de vez en cuando, afecta a la representación del conocimiento multidimensional y provoca problemas como una sobrecarga de información, demasiado ruido, redundancias en exceso y problemas de transitividad en las relaciones conceptuales. Para solucionar estos inconvenientes, este trabajo de fin de grado propone una extensión de los sistemas conceptuales de EcoLexicon refinando la hiponimia de tres formas: (i) la corrección de la herencia de propiedades, (ii) la implementación de conceptos paraguas y (iii) el establecimiento de subtipos de hiponimia. Además, en este trabajo también se realiza un proceso de extensión hiponímica mediante la extracción de corpus para obtener hipónimos de la base de datos de EcoLexicon de forma semiautomática utilizando gramáticas personalizadas, y así complementar y validar las nuevas jerarquías.

Palabras clave: relaciones conceptuales, multidimensionalidad, hiponimia, extracción de corpus, bases de conocimiento terminológico

1. Introduction

1.1. Summary of work

Multidimensionality, according to several theories in terminology, is the phenomenon by which the characteristics of a certain concept may vary depending on the perspective taken. Moreover, such conceptual multidimensionality can affect a wide range of properties and attributes, from shape to function (Kageura, 1997). The representation of multidimensionality is thus a major challenge in the design of terminological knowledge bases (TKBs), since extracting a few concepts and establishing simple relations between them results in monodimensional systems, which are unrealistic and only permit *in vitro* knowledge acquisition (Dubuc & Lauriston, 1997; Cabré, 1999).

This final degree project was carried out within the research group Lexicon[1]. The head of this group, Professor Pamela Faber, developed a cognitive approach to the study of terminology called Frame-Based Terminology (FBT), whose practical application is a TKB on environmental science known as EcoLexicon[2]. This TKB is a multidimensional and dynamic terminological resource that targets different user groups interested in expanding their knowledge of the environment for text comprehension and generation, such as environmental experts, technical writers, and translators.

Nevertheless, despite all of the advantages that this TKB provides, EcoLexicon can still be improved. Not surprisingly, the vast amount of information contained in this TKB is a double-edged sword, and this occasionally affects its representation of multidimensional knowledge. In fact, its problems include (i) information overload, (ii) excessive noise and redundancy, (iii) transitivity problems in conceptual relations and hierarchies, (iv) lack of property inheritance in terminological definitions, and (v) the presence of different types of cohyponyms at the same hierarchical level.

In order to solve the previously mentioned problems and to enhance multidimensionality as a result, the proposed solution in this final degree project was an extension of the conceptual systems in EcoLexicon by refining hyponymy (Gil-Berrozpe & Faber, 2016)

---

[1] http://lexicon.ugr.es/

[2] http://ecolexicon.ugr.es/

with three complementary methods: (i) correcting property inheritance in concept definitions, (ii) implementing umbrella concepts at intermediate levels of concept hierarchies, and (iii) establishing hyponymy subtypes according to entities and to processes.

In addition to these three ways to refine hyponymy, an additional process of hyponymic extension through corpus extraction was carried out using the tool Sketch Engine[3] (a corpus manager and analysis software). In this way, it was possible to semi-automatically extract hyponyms from the EcoLexicon corpus using customized word sketches, and subsequently improve a particular conceptual system with the new results (Gil-Berrozpe et al., 2016). Without a doubt, the creation of conceptual networks can be greatly facilitated by the use of specialized corpora. Apart from making this process more empirical, corpus analysis generates data that enhance and validate information elicited from experts.

1.2. Objectives

The general objective of this project was the following:

- To extend the conceptual systems in EcoLexicon with a new proposal to refine hyponymy, enhancing the representation of multidimensionality as a result.

To achieve this goal, the following complementary objectives were established:

- To review the fundamental theoretical characteristics behind the construction of TKBs.
- To explore the utility of TKBs for specialized translation purposes.
- To analyze the distinctive characteristics of EcoLexicon for the benefit of specialized translators.
- To review preceding theories and postulates on hyponymy refinement.
- To correct property inheritance in concept definitions so as to ensure coherence and cohesion in the corresponding hierarchies.

[3] https://www.sketchengine.co.uk/

- To establish a series of umbrella concepts that could be introduced as filters for the pertinent concept hierarchies.
- To define a set of hyponymy subtypes according to entities and to processes, relevant for the frame of the environment.
- To use customized word sketches in Sketch Engine so as to semi-automatically extract hyponyms from the EcoLexicon corpus.
- To extend the resulting enhanced conceptual systems in EcoLexicon with the addition of the new hyponyms extracted from corpora.

2. Theoretical framework

2.1. The cognitive shift in terminology

Contemporary theories of terminology have had a major impact on more traditional approaches to conceptual representation and knowledge organization. Whereas the General Theory of Terminology (Wuster, 1968) is mainly based on the univocity principle and the establishment of static standardizing conceptual structures, more recent proposals foreground dynamic phenomena, such as variation and multidimensionality. For this purpose, they take a cognitive perspective that links specialized knowledge representation to cognitive linguistics and semantics (Cabré, 1999; Temmerman, 2000; Faber, 2009, 2012). For example, in Frame-Based Terminology (Faber, 2009; Faber, 2012), frames account for knowledge structures that relate elements and entities associated with culturally embedded situations or events from human experience, and they emphasize both hierarchical and non-hierarchical relations (Faber, 2015). This is especially relevant to terminology work, since an accurate representation of conceptual relations is imperative in the building of a comprehensive knowledge resource, such as a TKB.

In recent years, research in specialized language has begun to acknowledge the need for an interdisciplinary approach and for a set of theoretical premises that will make conceptual modelling more objective (León-Araúz et al., 2010). In fact, the study of terminology and specialized communication is currently experiencing a 'cognitive shift' (Faber, 2009), which is granting greater importance to conceptual organization as reflected in neurological processes (Faber et al., 2014). Terms are specialized knowledge units used to designate the objects, events and processes characteristic of a specialized domain. In the same way as language mirrors the mind, terminological structure can be regarded as a reflection of conceptual structure.

2.2. Conceptual structures and the configuration of specialized concepts

The specification of conceptual structure must be grounded on a set of theoretical assumptions regarding categorization; more specifically, whether and to what extent sensory information is part of semantic representation and processing (Meteyard et al., 2012). In this sense, Patterson et al. (2007), propose a supramodal format for semantic representations, which is modality-invariant though derived from mappings across sensory and motor input. In terminology, the correlate of this supramodal representation is a category schema or template as posited by various authors (Faber et al., 2014; Roche et al., 2009; Leonardi, 2010). This top-level schema constrains perceptual input though, at the same time, it is also derived from sensorimotor mappings. This type of schema facilitates the retrieval of all the information stored, and is the frame for any semantic network.

Not surprisingly, the configuration of specialized concepts in networks with both hierarchical and non-hierarchical or associative relations has proven to be one of the most important aspects of terminology work (León-Araúz et al., 2012). Nevertheless, this task is far from simple because, in certain cases, the semantics of the relations are too vague, as can be observed in many thesauri, conceptual maps, and semantic networks (Jouis, 2006). That is the reason why a wide range of methods for structuring knowledge have been considered in terminology. These include extending non-hierarchical relations, specifying the properties of the relations, and integrating innovative theories from linguistics and artificial intelligence. In order to guarantee high-quality terminological work, it is thus necessary to establish a methodology based on logical properties that will facilitate the accurate organization of conceptual relations.

2.3. Terminological knowledge bases (TKBs)

2.3.1. Introduction to TKBs

Regarded by Meyer et al. (1992) as a hybrid between term banks and knowledge bases, TKBs represent the specialized knowledge of a certain field through related concepts and the terms that designate them in one or various languages. A TKB is thus a product that reflects both linguistic and cognitive processes. Optimally, TKBs should reflect how conceptual networks are established and structured in our minds. They must also be designed to meet the needs of a specific group of users, whether they are experts or lay public.

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.

However, one of the main problems in concept representation derives from the fact that the characteristics of a concept may vary depending on the perspective taken, and this is related to conceptual multidimensionality. In fact, the representation of multidimensionality must also follow rules. In this sense, conceptual (semantic) relations cannot be created on demand, but should systematically be derived from a set inventory (León-Araúz et al., 2012). For this reason, a logical methodology should be followed when extending the existing conceptual relations in a TKB.

2.3.2. TKBs and specialized translation

A complex knowledge acquisition process, which is a necessary prelude to specialized translation, requires optimal terminological resources and an accurate representation of specialized knowledge. Terms are the linguistic units that designate our conceptualization of objects, processes, states and attributes in a specialized domain. Therefore, they play a key role in understanding, representing, transmitting, and acquiring specialized knowledge. However, most translation-oriented terminological resources fail to reflect the complexity and dynamicity of conceptualization (Faber & San Martín, 2010:118), since the vast majority of specialized dictionaries, thesauri, and glossaries are term-based rather than concept-based.

Without a doubt, translators need to be proficient in the use of specialized language in order to translate specialized texts. Moreover, they also need to know how to access different terminological resources and how to acquire a certain amount of their conceptual content. Although it does not necessarily signify that they have as much knowledge in a particular field as a domain expert, they must reach a minimum threshold so as to guarantee high-quality work. For this reason, translators can acquire specialized knowledge in a cost-effective way if they have access to useful knowledge resources and if they have good documentation and search skills.

For example, according to López Rodríguez et al. (2013:50), ontology-based terminological databases can solve different problems related to translation, information retrieval, and knowledge management. This is possible because they link terms that designate the same concept in different languages, showing both conceptual and linguistic information whilst at the same time fomenting data interoperability.

Accordingly, lexicographic and terminographic tools should endeavor to reflect the dynamic way in which we conceptualize the world around us. Tercedor et al. (2012:181) state that from the perspective of professional translation, these resources should simultaneously respond to user needs towards the source text (i.e. information access and knowledge retrieval and acquisition) and the target text (i.e. knowledge transmission and text production). That is the reason why TKBs, which are focused on dynamic and multidimensional knowledge representation, are highly valuable tools in specialized translation.

Furthermore, structuring specialized concepts in networks with both hierarchical and non-hierarchical or associative relations is one of the key elements in modern terminological resources. Nonetheless, this process can give rise to an overly simplistic resource if, in the design stage, methods for structuring knowledge (e.g. establishing subtypes of conceptual relations, extending non-hierarchical relations, and specifying the properties of the relations) are not addressed.

2.3.3. TKBs and knowledge acquisition enhancement

According to Dury (2005:34), there are two categories of specialized translators: (i) scientists or domain experts who became translators for professional reasons and have a good knowledge of the concepts and their organization, but often lack competence and expertise in translation; (ii) specialized translators, who were trained as linguists and have experience in translation, but who often lack information and knowledge of specialized concepts and their organization. Both types of translator need knowledge resources that provide linguistic, conceptual, and contextual information, and which are specifically adapted to their needs.

In this sense, TKBs are terminological resources that store and represent the specialized knowledge in a certain domain, as reflected in the concepts and the terms that designate them in one or several languages. In addition, from a cognitive perspective, Meyer et al. (1992:159) state that the conceptual categories in TKBs should be structured in a similar way to how they are related in the brain. Therefore, they are products that encompass both linguistic and cognitive phenomena.

A conceptually-structured TKB, in which terms are linked to concepts based on non-language-specific criteria, is a useful resource for specialized translators. In this way, not only is there coherent cross-referencing, but also linguistic data can be added and manipulated without altering the quality and consistency of the conceptual design (Giacomini 2014:83). Moreover, their representation of metalinguistic and encyclopedic data contributes to the enhancement of knowledge acquisition for specialized translation purposes, as TKBs enable translators to search for corpus concordances and parallel texts. Not surprisingly, this is an extremely valuable feature for translators since it allows them to avoid extra-lexicographic searches and queries, which can be time-consuming tasks.

Furthermore, according to Giacomini (2015:10), any type of e-lexicographic resource (including TKBs) should respond to three requirements:

- Conceptual structure availability and properties, with a multi-level depth of conceptual structures and multi-vocal relations.
- Ease of access to conceptual data, with direct access via the conceptual structure and the microstructure, and with specified relations.
- Consistency of concept-term correspondences, in the search by concept or by term.

Therefore, these criteria reveal the importance of having a conceptual structure in the form of an ontology. Moreover, any ontology-based terminological database geared to the fulfillment of these requirements should reflect phenomena such as multidimensionality (Kageura, 1997; Bowker, 1997) and natural and contextual knowledge dynamism (León Araúz et al., 2013), which are basic to specialized knowledge representation and acquisition.

On the one hand, multidimensionality can be implemented in a TKB. Methods for doing this include the addition or the deletion of certain concepts or relations in specific nodes or in the system as a whole, the modification of certain characteristics or relations in specific nodes or in the system as a whole, and the implementation of new ways to represent knowledge (linguistic, conceptual, visual, interactive, etc.). On the other hand, contextual dynamism can be achieved by showing how concepts — and therefore, terms — modify their features and use depending on the context in which they are found or used, and depending on the level of user expertise or knowledge.

In fact, regarding specialized translators, the representation of contextual information in a TKB enhances the knowledge acquisition process because it allows them to comprehend how terms are used in real texts and helps them to select the best term for each discourse segment, depending on contextual domains and use situation (León Araúz et al., 2010:140). In line with this, Tercedor et al. (2012:186) also remark on the intimate link between dynamicity and multidimensionality. Since it is now possible to represent concepts from different perspectives or dimensions, lexicographic and terminographic practice should thus envisage the elaboration of more dynamic representations.

Furthermore, apart from features such as multidimensionality or accessibility, another salient characteristic of TKBs is multimodality. It has been shown that linking multimedia information (e.g. images) to the linguistic, conceptual, and contextual information of a TKB helps to satisfy user needs with regard to the reception, production, and translation of specialized texts (Prieto Velasco, 2009:227). In this way, TKBs can be represented as visual thesauri, merging multimodal information and highlighting the multidimensional character of knowledge representations.

These are all basic characteristics of a TKB, which optimize knowledge retrieval in specialized translation. Therefore, when properly designed, TKBs can provide information about term meaning (core meaning, peripheral meaning, metaphorical extensions), term collocations and morphological elements (combinatory potential, derivational potential), and term use and activation in certain texts and contexts (register, genre, dialect, position in associative networks) (Tercedor et al., 2012:182).

2.4. EcoLexicon: an environmental TKB

EcoLexicon (Faber et al., 2005; Faber et al., 2014; Buendía Castro & Faber, 2015; Reimerink et al., 2010), based on the theoretical premises of Frame-Based Terminology, is a multidimensional and dynamic TKB on environmental science. It specifically targets user knowledge acquisition through different types of multimodal and contextualized information, in order to respond to both cognitive and communicative needs. More specifically, its public is any user group interested in broadening its knowledge of the environment for text comprehension and/or generation (environmental experts, technical writers, translators, etc.). This resource is currently available in English and Spanish, though five more languages (German, Modern Greek, Russian, French and Dutch) are being gradually implemented. To date, its database consists of a total of 3,601 concepts and 20,163 terms.

EcoLexicon can be explored through an accessible visual interface with different modules for conceptual, linguistic, and graphical information (see Figure 1). Regarding the interest of a specialized translator, this TKB allows users to perform a concept search query. 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 contextual domain.

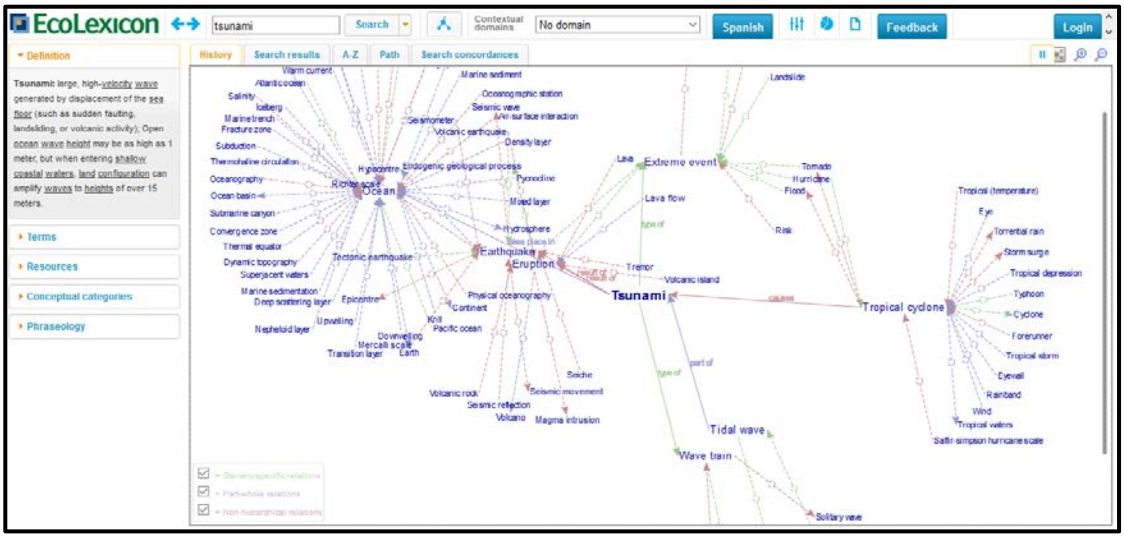

Figure 1. Visual interface of EcoLexicon (conceptual network of TSUNAMI)

In EcoLexicon, conceptual relations are classified in three main groups: generic-specific relations, part-whole relations and non-hierarchical relations (see Figure 2). As can be observed, hierarchical relations are divided into two groups to distinguish between hyponymic relations and meronymic relations. The set of generic-specific relations only comprises *type_of*. In contrast, the set of part-whole relations contains *part_of*, *made_of*, *delimited_by*, *located_at*, *takes_place_in*, and *phase_of*. In the last place, the set of non-hierarchical relations includes *affects*, *causes*, *attribute_of*, *opposite_of*, *studies*, *measures*, *represents*, *result_of*, *effected_by*, and *has_function*. The set of all conceptual relations in EcoLexicon comes to a total of 17. In some cases, these relations are domain-specific (e.g. *measures*), which means that the set of conceptual relations of a TKB may vary from one field of knowledge to another.

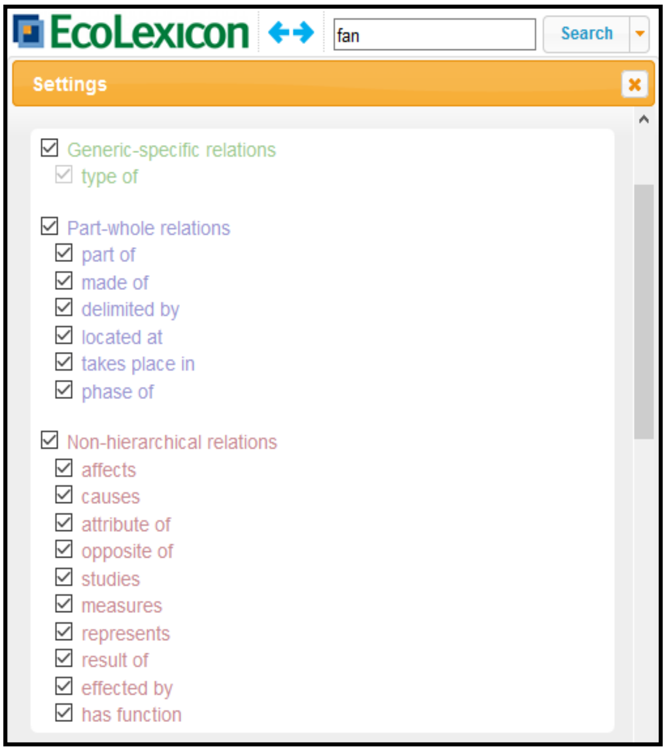

Figure 2. Semantic relations in EcoLexicon

With regard to the microstructure of this TKB, when a concept is selected, five sections are displayed on the left side of the interface:

- **Definition:** this section provides a terminological definition based on the explicitation of the *genus* (hyperonym or superordinate) and one or many *differentiae* (characteristics that vary in each cohyponym), alongside hyperlinks redirecting users to other concepts also contained in the database (see Figure 3). In the case of TSUNAMI, 'wave' is the *genus*, and there are assorted *differentiae* ('large' [*attribute_of*], 'high-velocity' [*attribute_of*], 'generated by displacement of the sea floor' [*result_of*], etc.).

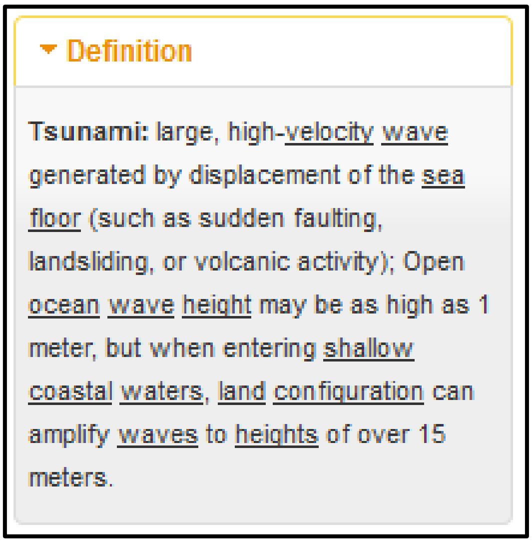

Figure 3. Definition section for TSUNAMI

- **Terms:** it displays the lexical denominations for a concept in the different languages available in EcoLexicon, information regarding the term type and the part of speech, and the option to display a list of corpus concordances (see Figure 4). For example, for TSUNAMI, it shows term variations in English (‘tsunami’ and ‘tidal wave’), in Spanish (‘*maremoto*’ and ‘*tsunami*’), in German (‘*Tsunami*’, ‘*Flutwelle*’ and ‘*Tsunami-Welle*’), in Russian (‘*цунами*’), and in Modern Greek (‘*θαλάσσιο σεισμικό κύμα*’ and ‘*τσουνάμι*’).

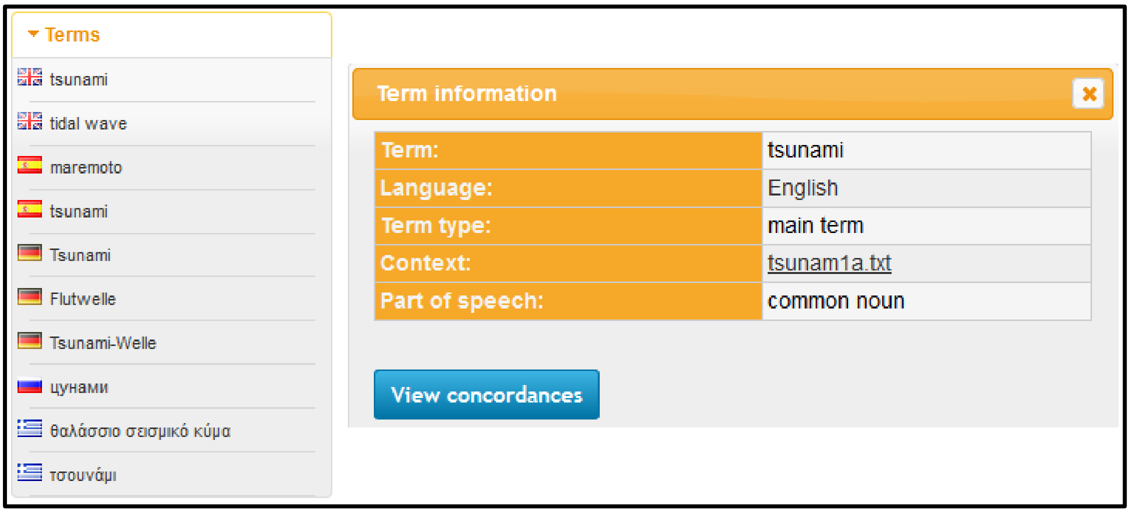

Figure 4. Term section for TSUNAMI

- **Resources:** this section offers a list of multimodal resources (images, videos, hyperlinks to external websites) for the chosen concept (see Figure 5). In this case, TSUNAMI presents a wide variety of resources, including clarifying pictures or diagrams, academic websites with thorough explanations on the topic, and even satellite images of tsunamis.

Figure 5. Resource section for TSUNAMI

- **Conceptual categories:** this module provides a list of different conceptual categories, and classifies the concept as a member of one of the categories (see Figure 6). For example, TSUNAMI is classified at the same time as a 'physical agent' (A.1.5.), as 'movement' (B.1.1.), and as 'part of water mass' (C.1.1.2.1.).

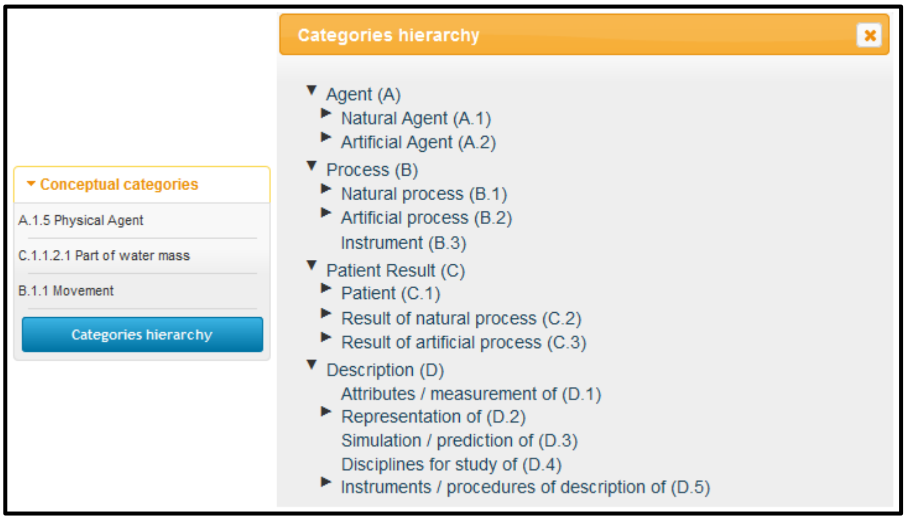

Figure 6. Conceptual category section for TSUNAMI

- **Phraseology:** this section displays the phraseological module, showing the nuclear meaning, the meaning dimension, the phraseological pattern and the verbs related to a certain concept (see Figure 7). TSUNAMI, for example, has a negative semantic prosody, since it is described as a "NATURAL DISASTER that causes a PATIENT to change for the worse". Moreover, it is related to the verb 'destroy', which further increases the negative connotation of the concept.

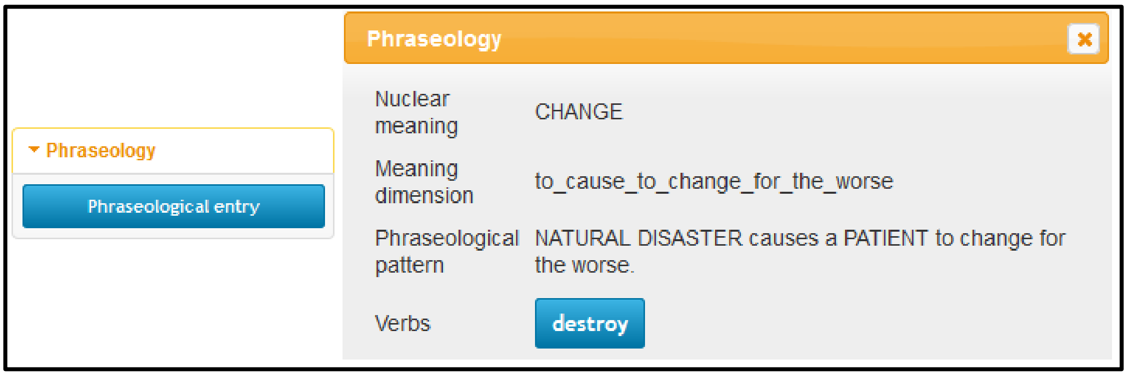

Figure 7. Phraseology section for TSUNAMI

These are the five main features that can be found within the microstructure of EcoLexicon, which highlight its linguistic, conceptual and multimodal nature. Furthermore, for the benefit of specialized translators, EcoLexicon also provides access to contextual information by means of corpus concordances. In this way, with the function 'Search concordances', users can obtain a list of contexts in which a term appears in the texts of the EcoLexicon corpus (see Figure 8).

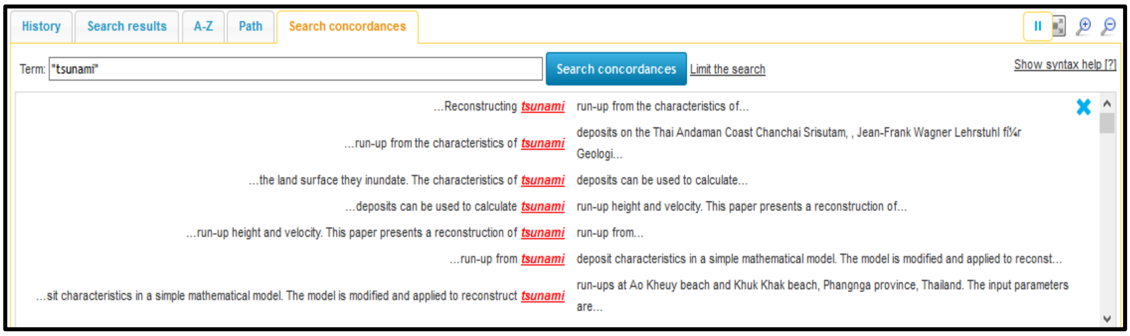

Figure 8. 'Search concordances' function for TSUNAMI

Thanks to all of these features, EcoLexicon is a resource that enhances knowledge acquisition. It also facilitates specialized translation because of its many-faceted knowledge representation: (i) through conceptual relations that codify conceptual propositions (concept-relation-concept) according to hierarchical and non-hierarchical criteria; (ii) through terminological definitions that reflect the salience of those conceptual relations, drawing from a central *genus*; and (iii) through multimodal resources that complement the conceptual and the linguistic information. Furthermore, studies by García Aragón et al. (2014) and Giacomini (2014), corroborate and validate the effectiveness of this terminological resource.

2.5. Refining hyponymy in EcoLexicon

The most basic definition of hyponymy is a relation of inclusion whose converse is hyperonymy (Murphy, 2006:446). According to Murphy (2003:217), hyponymy is central to many models of the lexicon for three reasons: (i) its inference-invoking nature; (ii) its importance in definition; and (iii), its relevance to selectional restrictions in grammar. Therefore, hyponymy plays a key role in ontology-based terminological resources, such as EcoLexicon, since it is the basis of all concept hierarchies.

As in meronymy, hyponymy can also be refined to provide an enhanced representation of generic-specific relations. This can be achieved by specifying subtypes of hyponymy (Murphy, 2003) or by establishing 'facets' and/or 'microsenses' (Cruse, 2002:4-5). Regarding hyponymy subtypes, the most commonly accepted distinction is between taxonomic hyponymy ('is-a-kind-of' relation) and functional hyponymy ('is-used-as-a-kind' relation). 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 2003:219-220).

In relation to the different types of hyponymy, Cruse (2002:4) proposes 'facets'. For instance, he distinguishes two 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).

On the other hand, 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.).

These 'microsenses' are currently represented in EcoLexicon by means of conceptual propositions in contextual discipline-based domains in which they are activated (see Figure 9). For example, based on the information in EcoLexicon, CHLORINE has two 'microsenses', one in the domain of 'Water Treatment and Supply' and the other in 'Chemical Engineering'. This occurs because, apart from being a *type_of* HALOGEN, CHLORINE can also be regarded as a *type_of* WATER DISINFECTANT (see Figure 10). In contrast, in all other domains, CHLORINE is only classified as a *type_of* HALOGEN (see Figure 11).

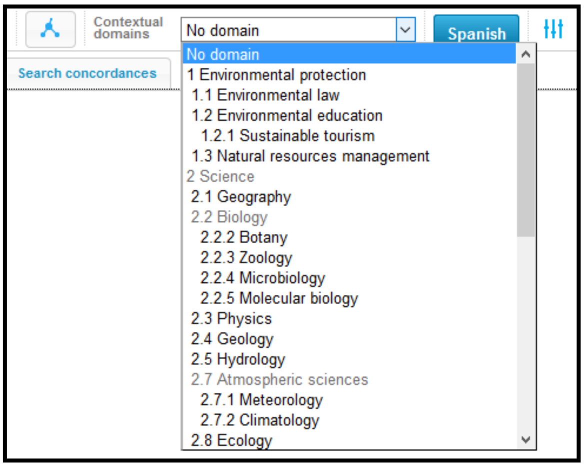

Figure 9. Contextual domains in EcoLexicon

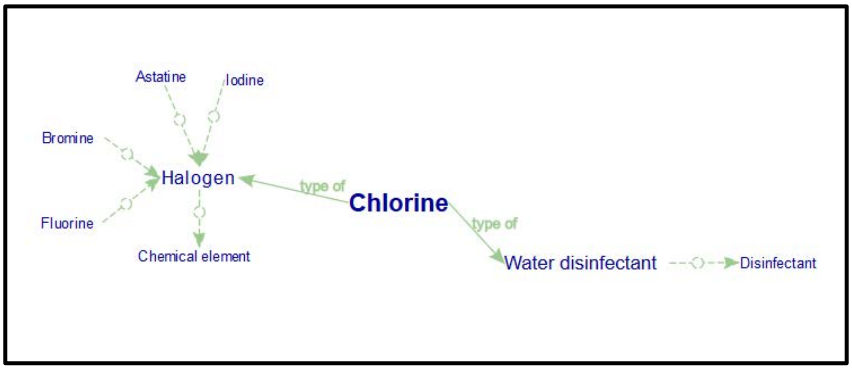

Figure 10. CHLORINE's generic-specific relations in 'Water Treatment and Supply' and 'Chemical Engineering'

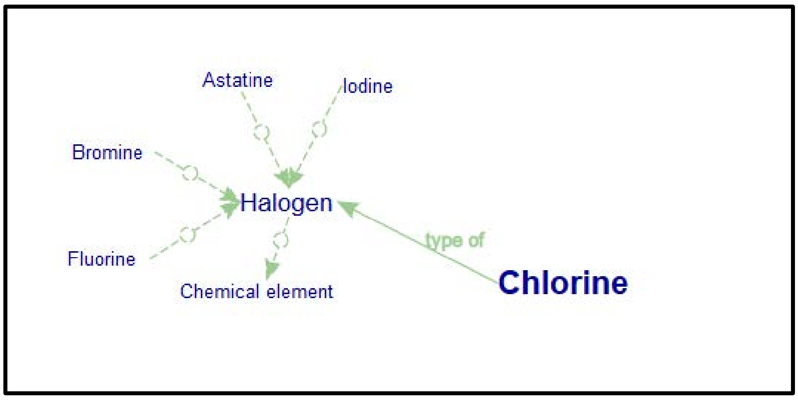

Figure 11. CHLORINE's generic-specific relations in the remaining contextual domains

However, with regard to hyponymy refinement, classifying concepts using contextual domains only makes it possible to filter the query and show context-dependent hyperonyms and hyponyms. In other words, the original sense of hyponymy remains the same and still needs to be decomposed in a certain way so as to guarantee a more accurate representation of generic-specific relations.

Without a doubt, TKBs can acquire greater coherence and dynamicity when the range of conceptual relations is wider than the traditional generic-specific and part-whole relations (León-Araúz et al., 2012), which entails taking into consideration non-hierarchic relations and, in addition, expanding the original sense of both hyponymy and meronymy. In EcoLexicon, the meronymic relation *part_of* was already divided into subtypes, as shown in Figure 2. For example, even though CONDENSATION is *part_of* the HYDROLOGIC CYCLE, it is more accurate to say that CONDENSATION is a *phase_of* the HYDROLOGIC CYCLE. This distinction was made in EcoLexicon because of the following factors: (i) domain-specific needs, (ii) ontological reasoning, and (iii) transitivity-related consistency (León-Araúz & Faber, 2010).

Nevertheless, the *type_of* relation still had not been subdivided. This was the source of a wide range of problems in EcoLexicon, such as the presence of different cohyponyms at the same level (see Figure 12), which produced noise as well as information overload and redundancy. Still another problem lied in transitivity and property inheritance. For example: LIMESTONE was originally represented as a hyponym to both ROCK and SEDIMENTARY ROCK.

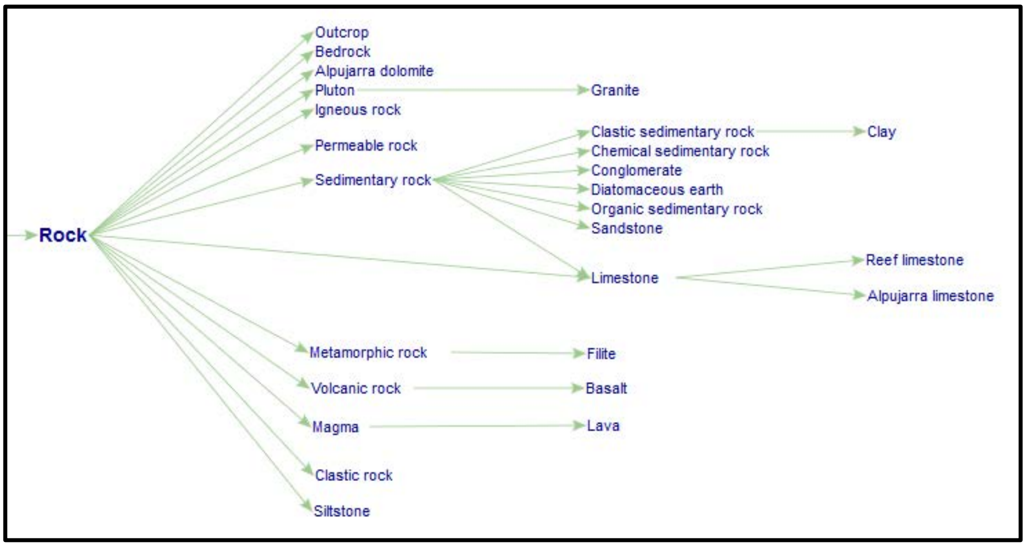

Figure 12. Presence of different dimensions of cohyponyms at the same level, and transitivity problems

Furthermore, Gheorghita & Pierrel (2012) state that the meaning of an input in a TKB can be disambiguated just by adding a domain to the definition. In the case of EcoLexicon, domains are not applied to definitions, but to conceptual relations. However, according to its database, only 2624 (50%) of all relations have been classified using a domain, and this invalidates the possibility of being completely accurate using this method. Therefore, there is still the need to further refine the *type_of* relation in this TKB.

3. Materials and methods

It has been shown that an exceedingly broad definition of hyponymy is the source of a wide range of problems in a TKB, which include information overload, noise, redundancy and transitivity-based inconsistencies (Gil-Berrozpe & Faber, 2016). For this reason, hyponymy, which is the backbone of all hierarchical semantic configurations in a terminological resource, should be refined according to different criteria: (i) corrected property inheritance in concept definitions; (ii) the creation of umbrella concepts, and/or (iii) a more refined set of hyponymy subtypes. This section focuses on the methodological aspects behind the processes of property inheritance correction, the implementation of umbrella concepts, and the semi-automatic corpus-driven extraction of hyponyms using Sketch Engine.

3.1. Correcting property inheritance

Property inheritance correction was regarded as the first step towards dividing the *type_of* relation into subtypes because, before obtaining the final result, it was necessary to guarantee coherence and correction in concept definitions. In this way, it was possible to show how hyponyms inherited the features or traits of their respective hyperonyms.

Indeed, hyponymy is a unidirectional relation where child concepts inherit the properties of their parent concepts, though they also have differentiating properties that make their meaning more specific. In a TKB, property inheritance between hyperonyms and hyponyms can be represented through *genus-differentia* definitions, based on the explicitation of the *genus* (hyperonym or superordinate) and one or many *differentiae* (characteristics that vary between cohyponyms) (Temmerman, 2000).

Despite the fact that most of the concepts in EcoLexicon are defined in this way, transitivity problems still arise (see Figure 12). This final degree project proposes a solution, as exemplified in the analysis of two types of concept: an entity (ROCK) and a process (EROSION).

3.1.1. Property inheritance in the conceptual network of an entity: ROCK

The original *type_of* network of ROCK was initially not accurately defined (see Figure 12). For example, LIMESTONE appeared as a direct hyponym of both ROCK and SEDIMENTARY ROCK, and there were two similar entities that designated ‘clastic rock’ at two different levels (CLASTIC ROCK and CLASTIC SEDIMENTARY ROCK). In order to solve such problems and related issues, the conceptual network of ROCK was enhanced with the addition of new concepts (e.g. SOLID ROCK, MOLTEN ROCK or DOLOMITE), and the property inheritance relations were restructured.

Table 1 shows an example of property inheritance in the original conceptual network. BASALT was defined as a 'rock of igneous origin', but its hyperonym (VOLCANIC ROCK) was also defined as an 'igneous rock'. Furthermore, the hyperonym of VOLCANIC ROCK was assumed to be ROCK, regardless of the fact that the only types of rock mentioned in its definition were 'igneous, sedimentary and metamorphic'.

ROCK: consolidated or unconsolidated aggregate or mass of minerals or organic materials. The three types of rock are igneous, sedimentary, and metamorphic.
VOLCANIC ROCK: extrusive igneous rock solidified near or on the surface of the Earth, resulting from volcanic activity.
BASALT: very hard rock of igneous origin, consisting of augite and triclinic feldspar, with grains of magnetic or titanic iron, and also bottle-green particles of olivine. It is formed by decompression melting of the Earth's mantle.

Table 1. ROCK – BASALT in the former conceptual network (original definitions)

Table 2 shows how property inheritance was improved in the new conceptual network. In this case, it is respected in all senses: BASALT is a *type_of* VOLCANIC ROCK, which is a *type_of* IGNEOUS ROCK, which is a *type_of* SOLID ROCK, which is a *type_of* ROCK. In other words, BASALT in the end reflects the inheritance of the characteristics possessed by all of its hyperonyms.

ROCK: consolidated or unconsolidated aggregate or mass of minerals or organic materials.
SOLID ROCK: rock in solid state, formed by the compression of sediments or the solidification of molten material.
IGNEOUS ROCK: solid rock formed by solidification of molten magma either beneath or at the Earth's surface.
VOLCANIC ROCK: extrusive igneous rock solidified near or on the surface of the Earth, resulting from volcanic activity.
BASALT: very hard volcanic rock, consisting of augite and triclinic feldspar, with grains of magnetic or titanic iron, and also bottle-green particles of olivine. It is formed by decompression melting of the Earth's mantle.

Table 2. ROCK – BASALT in the new conceptual network (enhanced definitions)

Finally, as a result of modifications in the remaining conceptual relations, improved terminological definitions, and the addition of new concepts to fill semantic gaps, the conceptual network of ROCK was enhanced (see Table 3).

<table>
<tr><th>LEVEL 0</th><th>LEVEL 1</th><th>LEVEL 2</th><th>LEVEL 3</th><th>LEVEL 4</th><th>LEVEL 5</th></tr>
<tr><td>ROCK -----------</td><td>solid rock ------------</td><td>sedimentary rock ----</td><td>limestone ------------------</td><td>reef limestone |
Alpujarra limestone |</td><td></td></tr>
<tr><td></td><td></td><td></td><td>clastic rock ------------------</td><td>clay |</td><td></td></tr>
<tr><td></td><td></td><td></td><td>chemical sedimentary rock ---</td><td>dolomite --------------</td><td>Alpujarra dolomite |</td></tr>
<tr><td></td><td></td><td></td><td>organic sedimentary rock |
conglomerate |
diatomaceous earth |
sandstone |
siltstone |</td><td></td><td></td></tr>
<tr><td></td><td></td><td>igneous rock ---------</td><td>plutonic rock ---------------</td><td>granite |</td><td></td></tr>
<tr><td></td><td></td><td></td><td>volcanic rock ---------------</td><td>basalt |</td><td></td></tr>
<tr><td></td><td></td><td>metamorphic rock ---</td><td>filite |</td><td></td><td></td></tr>
<tr><td></td><td></td><td>permeable rock |
bedrock |
outcrop |</td><td></td><td></td><td></td></tr>
<tr><td></td><td>molten rock ----------</td><td>magma ----------------</td><td>lava |</td><td></td><td></td></tr>
</table>

Table 3. Enhanced conceptual network of ROCK (generic-specific relations)

3.1.2. Property inheritance in the conceptual network of a process: EROSION

Property inheritance is also manifested in 'process' type of concepts. In this case, the original conceptual network of EROSION was also analyzed to examine if the inheritance of characteristics between parent and child concepts was accurate. As a result, certain concepts had to be relocated, and the definitions of some hyponyms needed to be enhanced to correct property inheritance.

To portray how property inheritance was corrected in concept definitions, another comparison was made to show the differences in the *type_of* relation established from EROSION to CHANNEL SCOUR. As can be observed in the original conceptual network (see Table 4), CHANNEL SCOUR, located at the third level with respect to EROSION, was defined as 'erosion' when it should have inherited the traits of its direct hyperonym, SCOUR.

EROSION: process by which materials of the Earth's crust are worn away, loosened, or dissolved while being transported from their place of origin by different agents, such as wind, water, bacteria, etc.
FLUVIAL EROSION: erosion of bedrock on the sides and bottom of the river; the erosion of channel banks; and the breaking down of rock fragments into smaller fragments by the flow of water in the channel.
SCOUR: localized erosive action of water in streams, excavating and carrying away material from the bed and banks.
CHANNEL SCOUR: erosion of a stream bed.

Table 4. EROSION – CHANNEL SCOUR in the former conceptual network (original definitions)

In contrast, in the enhanced conceptual network (see Table 5), property inheritance is well expressed, since each hyponym adopts the characteristics of its hyperonym: CHANNEL SCOUR is a *type_of* SCOUR, which is a *type_of* FLUVIAL EROSION, which is a *type_of* WATER EROSION, which is a *type_of* EROSION. Therefore, CHANNEL SCOUR (at the fourth level of hyponymy), after this process, was defined as a type of SCOUR rather than as a type of EROSION.

EROSION: process by which materials of the Earth's crust are worn away, loosened, or dissolved while being transported from their place of origin by different agents, such as wind, water, bacteria, etc.
WATER EROSION: erosion of rocks and sediment by water, involving detachment, transport and deposition.
FLUVIAL EROSION: water erosion of bedrock on the sides and bottom of the river; the erosion of channel banks; and the breaking down of rock fragments into smaller fragments by the flow of water in the channel.
SCOUR: localized fluvial erosion in streams, excavating and carrying away material from the bed and banks.
CHANNEL SCOUR: scour of a stream bed.

Table 5. EROSION – CHANNEL SCOUR in the new conceptual network (enhanced definitions)

Nevertheless, the previously mentioned modifications were not the only changes made to refine the conceptual network, as new concepts (e.g. WATER EROSION, RILL EROSION or STREAMBANK EROSION) were also added. In the end, an enhanced version was obtained (see Table 6). The correction of property inheritance not only enhances content, but also indicates how hyponymy can be decomposed into subtypes, which will be discussed in Section 4.1.

LEVEL 0	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4
EROSION -------------	water erosion -------------	fluvial erosion -----------	scour ------------------	channel scour |
				outflanking |
			sheet erosion |	
			rill erosion |	
			gully erosion |	
			streambank erosion |	
		sea erosion |		
	wind erosion -------------	deflation |		
	abrasion -------------------	glacier abrasion |		
	anthropic erosion |			
	glacier erosion |			
	internal erosion |			
	potential erosion |			
	differential erosion |			
	attrition |			
	denudation |			

Table 6. Enhanced conceptual network of EROSION (generic-specific relations)

3.2. Implementing umbrella concepts

Umbrella concepts are artificial concepts which can be introduced at intermediate levels of a hierarchy to further specify the sense of the expressed hyponymic relation. Therefore, their main objective is to narrow the link that connects parent concepts to child concepts by implementing an intermediate abstract concept, often characterized by an essential adjective (e.g. MEASURING INSTRUMENT).

Most of the umbrella concepts assessed during this project were noun compounds or noun phrases. In fact, this kind of syntactic construction carries a semantic component that makes it possible to classify noun compounds according to different aspects (i.e. composition, formation, agent, etc.). Levi (1978: 50) states that the semantic meaning of complex nominals (noun compounds consisting of a head noun preceded by a modifier, which is either another noun or a nominal adjective) can be extracted from their components. In fact, they imply semantic relationships, since they are all derived from the deletion or the nominalization of the predicate in the underlying sentence. For this reason, Levi (*ibid*: 76-77) highlights nine types of semantic predicates extracted by its deletion (CAUSE, HAVE, MAKE, USE, BE, IN, FOR, FROM, ABOUT) and four types of semantic predicates extracted by its nominalization (ACT, PRODUCT, AGENT, PATIENT).

Furthermore, various authors have also examined the assignment of semantic relations to noun-modifier pairs. For example, Rosario & Hearst (2001) follow Levi's premises and criteria to assign semantic relations to noun compounds in a domain-specific corpus. This resulted in a multi-class classification that greatly diminished raw information content. Moreover, Nastase & Szpaczowski (2003) also explored noun-modifier semantic relations based on semantic and morphological information about words themselves. In their study, they obtained a set of thirty semantic relations, which were subsequently used to represent identifying data (Nastase et al., 2006). More specifically, those thirty fine-grained semantic relations were classified in five classes: causal (e.g. CAUSE, EFFECT, PURPOSE), participant (e.g. AGENT, OBJECT, BENEFICIARY), spatial (e.g. DIRECTION, LOCATION, LOCATION AT), temporal (e.g. FREQUENCY, TIME AT, TIME THROUGH), and quality (e.g. MANNER, MATERIAL, MEASURE). Downing (1977:828) also proposed an inventory of the most common underlying semantic relations in noun compounds, made up of a total of twelve relations: WHOLE-PART, HALF-HALF, PART-WHOLE, COMPOSITION, COMPARISON, TIME, PLACE, SOURCE, PRODUCT, USER, PURPOSE, and OCCUPATION.

Thus, the identification and assignment of semantic relations to noun compounds could be a useful method to define and create umbrella concepts in EcoLexicon. For instance, depending on the semantic predicate underlying a noun compound, we could speak of 'function-based umbrella concepts', 'formation-based umbrella concepts', and 'agent-based umbrella concepts'. In fact, many examples of concept hierarchies that could be enhanced or had already been enhanced by using umbrella concepts were found in EcoLexicon. To be precise, this project focused on assessing the inclusion of umbrella concepts in the hierarchies of INSTRUMENT, PROCESS, CHEMICAL ELEMENT, and REPRESENTATION.

3.2.1. Umbrella concepts in the hierarchy of INSTRUMENT

To begin with, the original concept hierarchy of INSTRUMENT (see Table 7) presented a vast amount of information overload because of the large number of subordinates (68 hyponyms) that were linked to the same superordinate. For this reason, and because of the semantics of the concept INSTRUMENT, a set of five 'function-based umbrella concepts' (e.g. FILTERING INSTRUMENT, MEASURING INSTRUMENT, RECORDING INSTRUMENT) was introduced to provide a more accurate classification of the hyponyms (see Table 8).

INSTRUMENT (68 hyponyms)	ACCELEROMETER | AIR SAMPLER | ALBEDOGRAPH | ALBEDOMETER | ALTI-ELECTROGRAPH | ALTIMETER | ANEMOCLINOMETER | ANEMOGRAPH | ANEMOMETER | ANEROID CAPSULE | ATMORADIOGRAPH | AUTOMATIC SAMPLER | BAR SCREEN | BAROMETER | BATHYMETER | CLINOMETER | CLOUD CHAMBER | COMPASS | CREST GAGE | CTD | CURRENT METER | DEGREASER | DEPTH | FINDER | DREDGE | ECHO SOUNDER | EKMAN WATER BOOTLE | ELECTROSONDE | EMANOMETER | EVAPORIMETER | EVAPOTRANSPIROMETER | FLOWMETER | HYGROMETER | IMPEDOMETER | INCLINED GAUGE | INFILTROMETER | MARIGRAPH | METEOROGRAPH | PERMEAMETER | PHOTOMETER | PIEZOMETER | PLUVIOGRAPH | PLUVIOMETER | PSYCHROMETER | RADAR | SALINOMETER | SAND FILTER | SECCHI DISK | SEDIMENT SAMPLER | SEDIMENT TRAP | SEISMOGRAPH | SEISMOMETER | SEXTANT | SNOW GAUGE | SOLAR CELL | SOLAR PANEL | SOUNDING BALLOON | SOUNDING LEAD | STADIMETER | STAFF GAUGE | TENSIOMETER | THERMOMETER | THICKENER | TIDE STAFF | VENTILATED THERMOMETER | WATER SAMPLER | WATER-LEVEL RECORDER | WATER-TREATMENT PLANT

Table 7. Original concept hierarchy of INSTRUMENT, without umbrella concepts

INSTRUMENT (5 hyponyms)	FILTERING INSTRUMENT	BAR SCREEN | DEGREASER | SAND FILTER | SIEVE
	MEASURING INSTRUMENT	ALBEDOGRAPH | ALTI-ELECTROGRAPH | ANEMOGRAPH | ATMORADIOGRAPH | COMPASS | MARIGRAPH | METEOROGRAPH | PLUVIOGRAPH | RADAR | SEISMOGRAPH | WATER-LEVEL RECORDER
	RECORDING INSTRUMENT	ACCELEROMETER | ALBEDOMETER | ALTIMETER | ANEMOCLINOMETER | ANEMOMETER | ANEROID CAPSULE | BAROMETER | BATHYMETER | CLINOMETER | CREST GAGE | CTD | CURRENT METER | DEPTH FINDER | ECHO SOUNDER | ELECTROSONDE | EMANOMETER | EVAPORIMETER | EVAPOTRANSPIROMETER | FLOWMETER | HYGROMETER | IMPEDOMETER | INCLINED GAUGE | INFILTROMETER | PERMEAMETER | PHOTOMETER | PIEZOMETER | PLUVIOMETER | PSYCHROMETER | SALINOMETER | SECCHI DISK | SEDIMENT TRAP | SEISMOMETER | SEXTANT | SNOW GAUGE | SOUNDING BALLOON | SOUNDING LEAD | STADIMETER | STAFF GAUGE | TENSIOMETER | THERMOMETER | TIDE STAFF | VENTILATED THERMOMETER
	SAMPLING INSTRUMENT	AIR SAMPLER | AUTOMATIC SAMPLER | DREDGE | EKMAN WATER BOTTLE | SEDIMENT SAMPLER | WATER SAMPLER
	TRANSFORMING INSTRUMENT	CLOUD CHAMBER | SOLAR CELL | SOLAR PANEL | THICKENER | WATER-TREATMENT PLANT

Table 8. Enhanced concept hierarchy of INSTRUMENT, with umbrella concepts

3.2.2. Umbrella concepts in the hierarchy of PROCESS

The second example of how umbrella concepts can be implemented to refine the step from a hyperonym to a hyponym is shown in the concept hierarchy of PROCESS. In its original version (see Table 9), the number of hyponyms proposed for the same hyperonym (111 in total) was disproportionate in comparison to what is usually found in EcoLexicon.

However, in the enhanced version of the concept hierarchy (see Table 10), the visual representation of these relations is much clearer thanks to the inclusion of two 'agent-based umbrella concepts' (ARTIFICIAL PROCESS and NATURAL PROCESS).

Furthermore, after umbrella concepts were included in the hierarchy of PROCESS, certain original concepts became new umbrella concepts (e.g. ATMOSPHERIC PROCESS, ENDOGENIC GEOLOGICAL PROCESS, HYDROLOGICAL PROCESS) which also improved the whole hierarchy of the conceptual network. This proposal was limited to the inclusion of those umbrella concepts located at upper levels (i.e. ARTIFICIAL PROCESS and NATURAL PROCESS). However, in future work the use and function of these new sets of umbrella concepts will be examined to consider their inclusion in the TKB.

PROCESS (111 hyponyms)	ABSORPTION | ACCLIMATIZATION | ACCRETION | ADIABATIC PROCESS | AERATION | AFFLUX | AGGLOMERATION | ALIMENTATION | ANOMALOUS PROPAGATION | ANTICYCLOGENESIS | ANTICYCLOLYSIS | ATMOSPHERIC ABSORPTION | ATMOSPHERIC EMISSION | ATMOSPHERIC PROCESS | ATTENUATION | BERGERON PROCESS | BIOACCUMULATION | BIOTIC PROCESS | BLEEDING | BREAKING DROPS | CATCHMENT | CHANGE | CHEMICAL REACTION | CHEMOTAXIS | CLIMATE CHANGE | CLOUD ELECTRIFICATION | CLOUD SEEDING | COAGULATION | COALESCENCE | COASTAL PROCESS | COLLISION | COMPOSTING | CONDENSATION | CONDUCTION | CONSOLIDATION | CONTAMINATION | COOLING (OFF) | CORROSION | CYCLOGENESIS | DECANTATION | DECOMPOSITION | DECREASE | DEPOSITION | DESALINATION | DISCHARGE | DISEASE | DRAIN | DRAW-OFF | EFFLORESCENCE | ELIMINATION OF SOLID WASTE | ELUTRIATION | EMISSION | ENDOGENIC GEOLOGICAL PROCESS | EROSION | ERROR CORRECTION | EUSTATIC SEA LEVEL CHANGE | EUTROPHICATION | EVAPORATION | EXOGENOUS GEOLOGICAL PROCESS | FERTIGATION | FERTILIZATION | FILTRATION | FREEZING | GEOREFERENCING | GULLYING | HYDRODYNAMIC PROCESS | HYDROLOGICAL PROCESS | INLET STABILISATION | IONIZATION | ISOSTASY | LEACHING (SOIL) | LEACHING (WASTEWATER) | LOWERING OF THE WATER TABLE | MANAGEMENT | MELTING | NOURISH | NUCLEAR REACTION | OSMOSIS | OVERFLOW | OZONE DEPLETION | PARTICLE DISPERSION | PERCOLATION | PONDING | PUMPING | RADIATION PROCESS | RECYCLING | RETARD BEACH EROSION | RETARD LITTORAL DRIFT | REVEGETATION | RISE OF THE WATER TABLE | SALINIZATION | SAMPLING | SELECTIVE SORTING | SHOALING | SOLIDIFICATION | SOLUTION | SOUND | STOCHASTIC PROCESS | SUBLIMATION | SUSTAINED DEVELOPMENT | TECHNOLOGY | TEST DRILLING | THAW | TREATMENT | WATER EVACUATION | WATER LEVEL | REGULATION | WATER PONDING | WATERLOGGING | WAVE GENERATION | WAVE PROPAGATION

Table 9. Original concept hierarchy of PROCESS (without umbrella concepts)

PROCESS (2 hyponyms)	ARTIFICIAL PROCESS	AERATION | AFFLUX | ALIMENTATION* | ATMOSPHERIC EMISSION* | ATTENUATION* | CATCHMENT | CHANGE* | CHEMICAL REACTION* | CLIMATE CHANGE* | CLOUD SEEDING | COMPOSTING | CONTAMINATION* | CORROSION* | DECANTATION | DECREASE* | DESALINATION | DISCHARGE | DRAIN | DRAW-OFF | ELIMINATION OF SOLID WASTE | ELUTRIATION | EROSION* | ERROR CORRECTION | EUTROPHICATION* | FERTIGATION | FERTILIZATION | FILTRATION* | GEOREFERENCING | INLET STABILISATION* | IONIZATION* | LEACHING (WASTEWATER) | MANAGEMENT | NOURISH | NUCLEAR REACTION | OZONE DEPLETION | PUMPING | RECYCLING | RETARD BEACH EROSION | RETARD LITTORAL DRIFT | REVEGETATION* | SALINIZATION* | SAMPLING | SOUND | SUSTAINED DEVELOPMENT | TECHNOLOGY | TEST DRILLING | TREATMENT | WATER EVACUATION | WATER LEVEL REGULATION **<u>New umbrella concepts:</u>** ATMOSPHERIC PROCESS* | RADIATION PROCESS* | STOCHASTIC PROCESS
	NATURAL PROCESS	ABSORPTION | ACCLIMATIZATION | ACCRETION | AGGLOMERATION | ALIMENTATION* | ANOMALOUS PROPAGATION | ANTICYCLOGENESIS | ANTICYCLOLYSIS | ATMOSPHERIC ABSORPTION | ATMOSPHERIC EMISSION* | ATTENUATION* | BERGERON PROCESS | BIOACCUMULATION | BLEEDING | BREAKING DROPS | CHANGE* | CHEMICAL REACTION* | CHEMOTAXIS | CLIMATE CHANGE* | CLOUD ELECTRIFICATION | COAGULATION | CONDENSATION | CONDUCTION | CONSOLIDATION | CONTAMINATION* | COOLING (OFF) | CORROSION* | CYCLOGENESIS | DECOMPOSITION | DECREASE* | DEPOSITION | DISEASE | EFFLORESCENCE | EMISSION | EROSION* | EUSTATIC SEA LEVEL CHANGE | EUTROPHICATION* | EVAPORATION | FILTRATION* | FREEZING | GULLYING | INLET STABILISATION* | IONIZATION* | ISOSTASY | LEACHING (SOIL) | LOWERING OF THE WATER TABLE | MELTING | NUCLEAR REACTION | OSMOSIS | OVERGLOW | PARTICLE DISPERSION | PERCOLATION | PONDING | PROGRADATION | REVEGETATION* | RISE OF THE WATER TABLE | SALINIZATION* | SELECTIVE SORTING | SHOALING | SOLIDIFICATION | SOLUTION | SUBLIMATION | THAW | WATER PONDING | WATERLOGGING | WAVE GENERATION | WAVE PROPAGATION **<u>New umbrella concepts:</u>** ADIABATIC PROCESS | ATMOSPHERIC PROCESS* | BIOTIC PROCESS | COASTAL PROCESS | ENDOGENIC GEOLOGICAL PROCESS | EXOGENOUS GEOLOGICAL PROCESS | HYDRODYNAMIC PROCESS | HYDROLOGICAL PROCESS | RADIATION PROCESS*
* These concepts can be hyponyms of ARTIFICIAL PROCESS and NATURAL PROCESS at the same time. Whether they refer to an artificial process or to a natural process will be shown in further hyponymic levels.		

Table 10. Enhanced concept hierarchy of PROCESS (with umbrella concepts)

3.2.3. Umbrella concepts in the hierarchy of CHEMICAL ELEMENT

The third example analyzed was CHEMICAL ELEMENT. Without umbrella concepts, the concept hierarchy was extremely shallow since there was no intermediate level between the hyperonym and the hyponyms (42 in total) that could be used to classify them in more specific subgroups (see Table 11). However, in chemistry, chemical elements can be classified according to different criteria (period, group, block, etc.), so a set of ten 'general umbrella concepts' (e.g. ACTINIDE, NOBLE GAS, TRANSITION METAL) was implemented to refine the hyponymic classification of the concept hierarchy (see Table 12).

CHEMICAL ELEMENT (42 hyponyms)	ALUMINUM | ARGON | ARSENIC | ASTATINE | BERYLLIUM | BORON | BROMINE | CALCIUM | CARBON | CHLORINE | CHROMIUM | COBALT | COPPER | FLUORINE | GALLIUM | GERMANIUM | HELIUM | HYDROGEN | IODINE | IRON | ISOTOPE | LEAD | LITHIUM | MAGNESIUM | MANGANESE | NEON | NICKEL | NITROGEN | OXYGEN | PHOSPHORUS | POTASSIUM | RADON | SCANDIUM | SELENIUM | SILICON | SODIUM | SULFUR | TITANIUM | TRACE ELEMENT | URANIUM | VANADIUM | ZINC

Table 11. Original concept hierarchy of CHEMICAL ELEMENT (without umbrella concepts)

<table>
<tr><td rowspan="11">CHEMICAL ELEMENT
(13 hyponyms)</td><td>ACTINIDE</td><td>URANIUM</td></tr>
<tr><td>ALKALI METAL</td><td>LITHIUM | POTASSIUM | SODIUM</td></tr>
<tr><td>ALKALINE EARTH METAL</td><td>BERYLLIUM | CALCIUM | MAGNESIUM</td></tr>
<tr><td>BASIC METAL</td><td>ALUMINUM | GALLIUM | LEAD</td></tr>
<tr><td>HALOGEN</td><td>ASTATINE | BROMINE | CHLORINE | FLUORINE | IODINE</td></tr>
<tr><td>LANTHANIDE</td><td>-</td></tr>
<tr><td>NOBLE GAS</td><td>ARGON | HELIUM | NEON | RADON</td></tr>
<tr><td>NONMETAL</td><td>CARBON | NITROGEN | OXYGEN | PHOSPHORUS | SELENIUM | SULFUR</td></tr>
<tr><td>SEMIMETAL</td><td>ARSENIC | BORON | GERMANIUM | SILICON</td></tr>
<tr><td>TRANSITION METAL</td><td>CHROMIUM | COBALT | COPPER | IRON | MANGANESE | NICKEL | SCANDIUM | TITANIUM | VANADIUM</td></tr>
<tr><td>HYDROGEN | ISOTOPE | TRACE ELEMENT</td><td></td></tr>
</table>

Table 12. Enhanced concept hierarchy of CHEMICAL ELEMENT (with umbrella concepts)

Nevertheless, as can be observed in Table 12, umbrella concepts are not always reflected in polylexical terms. In fact, in this case, umbrella concepts can be both monolexical (e.g. ACTINIDE, HALOGEN) and polylexical terms (e.g. NOBLE GAS, TRANSITION METAL). The fact that most of the umbrella concepts are noun compounds or noun phrases means that they can be created relatively easily through simple syntactic and semantic constructions. Nevertheless, in certain cases the umbrella concepts carry an essive component which can only be described with a completely different term. These are mostly 'general umbrella concepts' with exactly the same use and function as those noun compounds derived from a process of semantic predicate deletion or nominalization.

3.2.4. Umbrella concepts in the hierarchy of REPRESENTATION

The last example was the concept hierarchy of REPRESENTATION, in which all of the umbrella concepts implemented were monolexical terms. Since the original concept hierarchy of REPRESENTATION (see Table 13) was characterized by a vast amount of hyponyms (204), it was necessary to include an intermediate classification to regroup the cohyponyms in more specific dimensions. Therefore, a set of 18 'general umbrella concepts' related to ways of representing information (e.g. GRAPH, IMAGE, MODEL) was implemented so as to filter the hierarchy and thus obtain a much clearer hyponymic classification (see Table 14).

All these examples indicate the usefulness of umbrella concepts as a means of refining hyponymy in a TKB, complementarily to correcting property inheritance and to decomposing hyponymy into subtypes. However, further research is needed to determine the usefulness of umbrella concepts in EcoLexicon as a whole. Moreover, it will also be necessary to establish objective parameters (i.e. abstraction, information overload, noise, etc.) to decide which conceptual networks should be enhanced with umbrella concepts.

<table>
<tr>
<td>Representation
(204 hyponyms)</td>
<td>adiabatic chart | age of the sea waves | ageostrophic model | air density | aircraft meteorological data relay | analog model | anemogram | atmospheric model | average monthly and yearly precipitation map | axis of anticyclone | axis of depression | backwater curve | balance between the pressure field and the horizontal field of motion | balance equation | baseline cartography | bathymetric chart | Beaufort scale | behavior model | berm line | bifurcation ratio | Bowen ratio | capillary-tube model | cartogram | cartographic information | Celsius temperature scale | chemical system | climate model | coalescence efficiency | coefficient | contour | co-range line | co-tidal line | Darcy’s law | data model | derived model | design storm | diagnostic equation | differential equation | digital elevation model | digital terrain model | discharge curve | distributed model | divergence equation | Douglas scale | dynamic equation | dynamics | Earth’s elliptic orbit | echogram | ecliptic plane | ecosystem | Ekman spiral | electricity potential | empirical flood formula | energy equation | environmental fluid dynamics code | environmental map | equation of state | equator | equilibrium line | equinoctial low tide | erosion model | Fahrenheit temperature scale | falling curve | fall-stage-discharge relation | farming and exploitation map | foam line | Fujita-pearson scale | future and emerging technologies | geographic coordinate system | geologic map | gradient | groin system | groundwater flow model | height above sea level | Hele-shaw model | high latitudes | high water line | hydraulic radius model | hydrograph | hydrologic data | hydrologic model | hydrostatic equation | impedance | index | integrated watershed management model | irrigation system | isobar | isobath | isochron | isochrone map | isohaline | isohyet | isohypse | isopach | isopach chart | isotach | isotherm | jet-stream axis | Kelvin temperature scale | land modelling | land use map | level | limit of backwash | limit of uprush | limnigraph | Linke blue sky scale | longitude | low latitudes | low water line | Lucdeme project | manning equation | map of nutrient sources | Margules equation | marigram | mathematical function | mathematical model | Mercalli scale | meridian | mesoscale | metadata | meteorogram | microscale | mixing ratio | momentum equation | Monin-obukhov equation | nautical chart | Newton’s laws of motion | non-linear equation | numerical model | one-line model | operational model | ordinary high water mark | orthophoto | outflow boundary | photomap | physical system | pluviogram | prediction | pressure-volume-temperature relationship | probable maximum precipitation | profile map | protocol | radiant flux | Rankine temperature scale | raster | rating curve | Réaumur temperature scale | recession curve | reef system | reference evapotranspiration | refraction-diffraction model | relative evaporation | relative humidity | relative permeability | Richter scale | ridge and runnel system | road map | rule | Saffir-simpson hurricane scale | sand fraction | satellite image | shoreline | significant wave | simulation | slope map | soil density | soil map | soil saturation | spatial correlation | spatial interpolation of the average yearly and monthly precipitation | specific groundwater runoff | specific humidity | stage-discharge relation | star chart | state-of-sea scale | step | storage ratio | strand line | stratigraphic column | stream line | stream tube | synoptic scale | tap water system | tendency equation | thalweg | tidal current chart | tidal piezometric efficiency | tide table | topographic map | tropic of cancer | tropic of capricorn | T-s diagram | urban system | Van genuchten model | water density | water line (land) | water line (vessel) | water quality model | water retention curve | wave climate | wave equation | wave frequency | wave steepness | waveform</td>
</tr>
</table>

Table 13. Original concept hierarchy of representation (without umbrella concepts)

REPRESENTATION (18 hyponyms)	AXIS	AXIS OF ANTICYCLONE / AXIS OF DEPRESSION / JET-STREAM AXIS
	EQUATION	BALANCE EQUATION / DIAGNOSTIC EQUATION / DIFFERENTIAL EQUATION / DIVERGENCE EQUATION / DYNAMIC EQUATION / ENERGY EQUATION / EQUATION OF STATE / HYDROSTATIC EQUATION / MANNING EQUATION / MARGULES EQUATION / MOMENTUM EQUATION / MONIN-OBUKHOV EQUATION / NON-LINEAR EQUATION / TENDENCY EQUATION / WAVE EQUATION
	GRAPH	ANEMOGRAM / BACKWATER CURVE / CONTOUR / DISCHARGE CURVE / ECHOGRAM / FALLING CURVE / HYDROGRAPH / LIMNIGRAPH / MARIGRAM / PLUVIOGRAM / RATING CURVE / RECESSION CURVE / T-S DIAGRAM / WATER RETENTION CURVE / WAVEFORM
	IMAGE	SATELLITE IMAGE
	ISOGRAM	ISOBAR / ISOBATH / ISOCHRON / ISOHALINE / ISOHYET / ISOHYPSE / ISOPACH / ISOTACH / ISOTHERM
	LATITUDE	HIGH LATITUDES / LOW LATITUDES
	LIMIT	BERM LINE / CO-RANGE LINE / CO-TIDAL LINE / EQUILIBRIUM LINE / FOAM LINE / HIGH WATER LINE / LEVEL / LIMIT OF BACKWASH / LIMIT OF UPRUSH / LOW WATER LINE / ORDINARY HIGH WATER MARK / OUTFLOW BOUNDARY / SHORELINE / STEP / STRAND LINE / STREAM LINE / THALWEG / WATER LINE (LAND) / WATER LINE (VESSEL)
	LONGITUDE	-
	MAP	ADIABATIC CHART / AVERAGE MONTHLY AND YEARLY PRECIPITATION MAP / BATHYMETRIC CHART / CARTOGRAM / ENVIRONMENTAL MAP / FARMING AND EXPLOITATION MAP / GEOLOGIC MAP / ISOCHRONE MAP / ISOPACH CHART / LAND USE MAP / MAP OF NUTRIENT SOURCES / NAUTICAL CHART / ORTHOPHOTO / PHOTOMAP / PROFILE MAP / ROAD MAP / SLOPE MAP / SOIL MAP / STAR CHART / TIDAL CURRENT CHART / TOPOGRAPHIC MAP
	MERIDIAN	-
	MODEL	AGEOSTROPHIC MODEL / ANALOG MODEL / ATMOSPHERIC MODEL / BEHAVIOR MODEL / CAPILLARY-TUBE MODEL / CLIMATE MODEL / DATA MODEL / DERIVED MODEL / DESIGN STORM / DIGITAL ELEVATION MODEL / DIGITAL TERRAIN MODEL / DISTRIBUTED MODEL / ENVIRONMENTAL FLUID DYNAMICS CODE / EROSION MODEL / GROUNDWATER FLOW MODEL / HELE-SHAW MODEL / HYDRAULIC RADIUS MODEL / HYDROLOGIC MODEL / INTEGRATED WATERSHED MANAGEMENT MODEL / LAND MODELLING / MATHEMATICAL MODEL / NUMERICAL MODEL / ONE-LINE MODEL / OPERATIONAL MODEL / REFRACTION-DIFFRACTION MODEL / SIMULATION / VAN GENUCHTEN MODEL / WATER QUALITY MODEL
	PARALLEL	EQUATOR / TROPIC OF CANCER / TROPIC OF CAPRICORN
	PATH	EARTH'S ELLIPTIC ORBIT / ECLIPTIC PLANE
	PIECE OF DATA	BASELINE CARTOGRAPHY / CARTOGRAPHIC INFORMATION / COEFFICIENT / EKMAN SPIRAL / EQUINOCTIAL LOW TIDE / HEIGHT ABOVE SEA LEVEL / HYDROLOGIC DATA / INDEX / METADATA / METEOROGRAM / PREDICTION / PROBABLE MAXIMUM PRECIPITATION / RASTER / REFERENCE EVAPOTRANSPIRATION / RULE / SAND FRACTION / SIGNIFICANT WAVE / SOIL SATURATION / SPATIAL INTERPOLATION OF THE AVERAGE YEARLY AND MONTHLY PRECIPITATION / STRATIGRAPHIC COLUMN / STREAM TUBE / TIDE TABLE / WAVE CLIMATE / WAVE FREQUENCY
	PROGRAMME	AIRCRAFT METEOROLOGICAL DATA RELAY / FUTURE AND EMERGING TECHNOLOGIES / LUCDEME PROJECT / PROTOCOL
	RATIO	AGE OF THE SEA WAVES / AIR DENSITY / BALANCE BETWEEN THE PRESSURE FIELD AND THE HORIZONTAL FIELD OF MOTION / BIFURCATION RATIO / BOWEN RATIO / COALESCENCE EFFICIENCY / DARCY'S LAW / ELECTRICITY POTENTIAL/ EMPIRICAL FLOOD FORMULA / FALL-STAGE-DISCHARGE RELATION / GRADIENT / IMPEDANCE / MATHEMATICAL FUNCTION / MIXING RATIO / NEWTON'S LAWS OF MOTION / PRESSURE-VOLUME-TEMPERATURE RELATIONSHIP / RADIANT FLUX / RELATIVE EVAPORATION / RELATIVE HUMIDITY / RELATIVE PERMEABILITY / SOIL DENSITY / SPATIAL CORRELATION / SPECIFIC GROUNDWATER RUNOFF / SPECIFIC HUMIDITY / STAGE-DISCHARGE RELATION / STORAGE RATIO / TIDAL PIEZOMETRIC EFFICIENCY / WATER DENSITY / WAVE STEEPNESS
	SCALE	BEAUFORT SCALE / CELSIUS TEMPERATURE SCALE / DOUGLAS SCALE / FAHRENHEIT TEMPERATURE SCALE / FUJITA-PEARSON SCALE / KELVIN TEMPERATURE SCALE / LINKE BLUE SKY SCALE / MERCALLI SCALE / MESOSCALE / MICROSCALE / RANKINE TEMPERATURE SCALE / RÉAUMUR TEMPERATURE SCALE / RICHTER SCALE / SAFFIR-SIMPSON HURRICANE SCALE / STATE-OF-SEA SCALE / SYNOPTIC SCALE
	SYSTEM	CHEMICAL SYSTEM / DYNAMICS / ECOSYSTEM / GEOGRAPHIC COORDINATE SYSTEM / GROIN SYSTEM / IRRIGATION SYSTEM / PHYSICAL SYSTEM / REEF SYSTEM / RIDGE AND RUNNEL SYSTEM / TAP WATER SYSTEM / URBAN SYSTEM

Table 14. Enhanced concept hierarchy of REPRESENTATION (with umbrella concepts)

3.3. Extracting hyponyms with the use of word sketches

The creation of any terminological resource can be improved by means of the automatized extraction of hyponymic pairs from specialized texts (León-Araúz & Reimerink, in press). According to Barrière (2004a), specialized corpora contain knowledge-rich contexts (KRCs) that can be made explicit through knowledge patterns (KPs), allowing the terminographer to look only at the subset of sentences that contain the required information. Therefore, implementing tools for corpus analysis in the creation of TKBs allows for a semi-automatic construction of concept hierarchies (Barrière, 2004b).

On the one hand, simply using the 'concordance' function of Sketch Engine, the EcoLexicon corpora and the corpus query language (CQL), different hyponymic structures can be distinguished. CQL can be used to restrict the queries and find only the types of expressions that are required. In the case of hyponymy, KPs are often presented as "X is a Y", "X such as Y", and "X like Y", to name some examples. In this way, it is possible to make a query to automatically find all relevant hyponymic expressions.

On the other hand, this task can be greatly improved using customized sketch grammars with the 'word sketches' function, which makes it possible to take a look at a word's grammatical and collocational behavior. The default word sketches offered by Sketch Engine represent different relations (verb-object, modifiers, prepositional phrases), but they only account for linguistic relations, so developing new customized sketch grammars focused on the extraction of sematic relations could allow terminologists to perform a more efficient conceptual analysis on any corpus (León-Araúz & San Martín, in press).

For this part of the project, the English EcoLexicon corpus, which currently consists of over 59 million words and is limited to the environmental domain, was analyzed. As previously stated, Sketch Engine was used for corpus querying and for generating word sketches. Nevertheless, as the only semantic relation included in the default English sketch grammar is a very simple hyponymic word sketch, several new fine-grained hyponymic sketch grammars needed to be developed, resulting in a total of 18 (León-Araúz & San Martín, in press; Gil-Berrozpe et al., in press). During their development, different issues specific of generic-specific relations were considered: 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. Likewise, the fact that one single sentence could produce more than one term pair due to the enumerations that are often found at each side of the pattern (e.g. *x, y, z and other types of w*) was also considered. In addition to this, a limitation of the number of possible words at each side of the pattern was regarded. Table 15 shows one example of the 18 sketch grammars that were implemented in Sketch Engine during this project, as well as an explanation of each element.

<table>
<tr><td colspan="2">2:"N.*" [tag!="V.*"]{0,5} "MD"? [word!="not"]? [lemma="be|,|\("] [word!="not"]? [word="defined |classified|categori.ed|regarded"] [word="as"] "DT.*|RB.*|JJ.*"* ([lemma="type|kind|example|group|class| sort|category|family|species|subtype|subfamily|subgroup|subclass|subcategory|subspecies"] [word="of"])? [tag!="V.*"]{0,2} 1:[tag="N.*" & lemma!="type|kind|example|group|class|sort |category|family|species| subtype|subfamily|subgroup|subclass|subcategory|subspecies"]</td></tr>
<tr><td>2:"N.*"</td><td>The hyponym is a noun.</td></tr>
<tr><td>[tag!="V.*"]{0,5}</td><td>From 0 to 5 words that are not verbs. This allows to capture enumerations and allows for the presence of adverbs, prepositions, etc.</td></tr>
<tr><td>"MD"?</td><td>Optional modal verb</td></tr>
<tr><td>[word!="not"]?</td><td>Optional word that is not not. This filters out negative sentences.</td></tr>
<tr><td>[lemma="be|,|\("]</td><td>The lemma be, comma or opening parenthesis.</td></tr>
<tr><td>[word!="not"]?</td><td>Optional word that is not not. This filters out negative sentences.</td></tr>
<tr><td>[word="defined|classified|categori.ed| regarded"]</td><td>The words defined, classified, categorized, categorised or regarded.</td></tr>
<tr><td>[word="as"]</td><td>The word as.</td></tr>
<tr><td>"DT.*|RB.*|JJ.*"*</td><td>From 0 to 100 determiners, adverbs or adjectives. This allows for phrases such as “the most important”, “a very special”, etc.</td></tr>
<tr><td>([lemma="type|kind|example|group|class|sort|category|family|species|subtype|subfamily|subgroup|subclass|subcategory|subspecies"] [word="of"])?</td><td>The lemma type, kind, example, group, class, sort, category, family, species, subtype, subfamily, subgroup, subclass, subcategory or subspecies followed by the word of (both optional).</td></tr>
<tr><td>[tag!="V.*"]{0,2}</td><td>From 0 to 2 words that are not verbs. This allows for the presence of determiners, adjectives, adverbs, etc.</td></tr>
<tr><td>1:[tag="N.*" & lemma!="type|kind| example|group|class|sort|category |family|species|subtype|subfamily|subgroup|subclass|subcategory|subspecies"]</td><td>The hypernym is a noun that does not have type, kind, example, group, class, sort, category, family, species, subtype, subfamily, subgroup, subclass, subcategory or subspecies as lemma.</td></tr>
</table>

Table 15. CQL representation of a generic-specific KP followed by its explanation (León-Araúz & San Martín, in press; Gil-Berrozpe et al., in press)

Moreover, Table 16 shows a sample of the various concordances that can be extracted with different generic-specific or hyponymic grammars:

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

Table 16. Concordances extracted with generic-specific grammars (León-Araúz & San Martín, in press)

This process was applied to the lemma ROCK in order to systematically analyze hyponymy as reflected in the specialized texts contained in the EcoLexicon corpus, and subsequently extend and correct its concept hierarchies represented in EcoLexicon. The default *modifier* word sketch and the new *"X" is the generic of…* word sketch (containing the fine-grained 18 hyponymic sketch grammars) were used for this purpose (see Figure 13).

modifier	18,175	0.53	"rock" is the generic of...	2,818	0.08
sedimentary +	1,939	11.44	limestone +	138	10.38
igneous +	1,260	10.99	granite +	106	10.10
metamorphic +	796	10.34	basalt	64	9.42
volcanic +	624	9.56	sandstone	58	9.26
molten +	324	9.12	shale	53	9.15
old +	323	8.78	schist	32	8.49
carbonate +	261	8.69	marble	28	8.30
hard +	215	8.38	gabbro	23	8.03
solid +	254	8.33	gneiss	22	7.96
country +	173	8.15	mineral	29	7.89
plutonic +	155	8.11	andesite	21	7.89
parent +	150	7.93	rhyolite	20	7.81

Figure 13. Hyponymic ROCK word sketches

As can be observed in Figure 13, the *modifier* word sketch offers a list of hyponyms that are exclusively compound nouns (e.g. SEDIMENTARY ROCK, IGNEOUS ROCK, METAMORPHIC ROCK, VOLCANIC ROCK, MOLTEN ROCK, etc.). On the contrary, the *"X" is* *the generic of…* word sketch provides a series of monolexical hyponyms to the hyperonym (e.g. LIMESTONE, GRANITE, BASALT, SANDSTONE, SHALE, etc.), and thus expanding the results of the query for generic-specific relations. In the same way, Figure 14 also shows the outcome of applying this process to IGNEOUS ROCK, with similar results. In addition to this, Figure 15 shows a list of concordances for IGNEOUS ROCK that can be obtained using the new word sketches.

igneous rock

(rock-n filtered by igneous-j)

rock: modifier	1,260	1.00	rock: "rock" is the generic of...	53	0.04
intrusive	55	10.13	gabbro	7	11.00
extrusive	44	10.02	diorite	5	10.93
coarse-grained	23	8.97	granite	19	10.69
plutonic	18	8.60	basalt	10	10.12
ultramafic	12	8.06	biotite	5	10.08
mafic	9	7.56	andesite	4	10.06
silicic	8	7.52	peridotite	3	10.02
felsic	8	7.51	olivine	3	9.35
ultrabasic	7	7.42	amphibole	3	9.21

Figure 14. Hyponymic IGNEOUS ROCK word sketches

crystals to form. Thus, intrusive	**igneous rocks**	, such as granite, diorite, and gabbro,
composition, areas underlain by	**igneous rocks**	such as granite are particularly susceptible
and slate respectively, while	**igneous rock**	such as granite may be metamorphosed into
. It is characteristic of acid	**igneous rocks**	such as granite, but because of its resistance
Granite and basalt are types of	**igneous rocks**	. Metamorphic rock is formed by the action
volcanic breccia differ from other	**igneous rocks**	such as granite and basalt? FIGURE 3.16
chemical weathering. Thus, when	**igneous rocks**	such as granite are attacked by weathering
other hand granite, a type of	**igneous rock**	that is generally strong and durable, is

Figure 15. Hyponymic IGNEOUS ROCK concordances

Finally, after combining the results of these corpus queries with the existing concept hierarchies in EcoLexicon, the conceptual systems were extended with new concepts.

4. Results and discussion

Once property inheritance was corrected in the definitions of the analyzed examples and certain conceptual systems were enriched through the implementation of umbrella concepts, it was possible to establish different subtypes of hyponymy depending on entities and processes (Gil-Berrozpe & Faber, 2016). However, this project only dealt with the specification of fine-grained hyponymy subtypes in the examples shown in Section 3.1.: the conceptual networks of ROCK and EROSION. Afterwards, the results of the semi-automatic corpus-driven extraction of ROCK hyponyms through the use of word sketches (Section 3.3.) were also assessed, and they were included in the corresponding hierarchy to extend the conceptual system and to expand the outcome of refining hyponymy with subtypes (Gil-Berrozpe et al., 2016).

4.1. Establishing subtypes of hyponymy

According to Murphy (2003, 2006), hyponymy can be divided into subtypes, such as taxonomic hyponymy and functional hyponymy (see Section 2.5). In this case, after correcting property inheritance and enriching the hierarchies with new concepts, it was possible to specify a more fine-grained set of subtypes in the previously analyzed conceptual networks. This subdivision of generic-specific relations is initially based on whether the concept is an entity (ROCK) or a process (EROSION).

4.1.1. Hyponymy subtypes in the conceptual network of an entity: ROCK

Based on the improved tree-mode network of ROCK (see Figure 16) and the new concept definitions, up to five different entity-related subtypes of hyponymy could be established:

- **State-based hyponymy:** a *type_of* relation dependent on the state of matter of the hyponyms.
- **Formation-based hyponymy:** a *type_of* relation dependent on the formation process or the origin of the hyponyms.
- **Composition-based hyponymy:** a *type_of* relation dependent on the components or the constituents of the hyponyms.
- **Location-based hyponymy:** a *type_of* relation dependent on the physical situation or location of the hyponyms.
- **Attribute-based hyponymy:** a *type_of* relation dependent on the traits or features of the hyponyms.

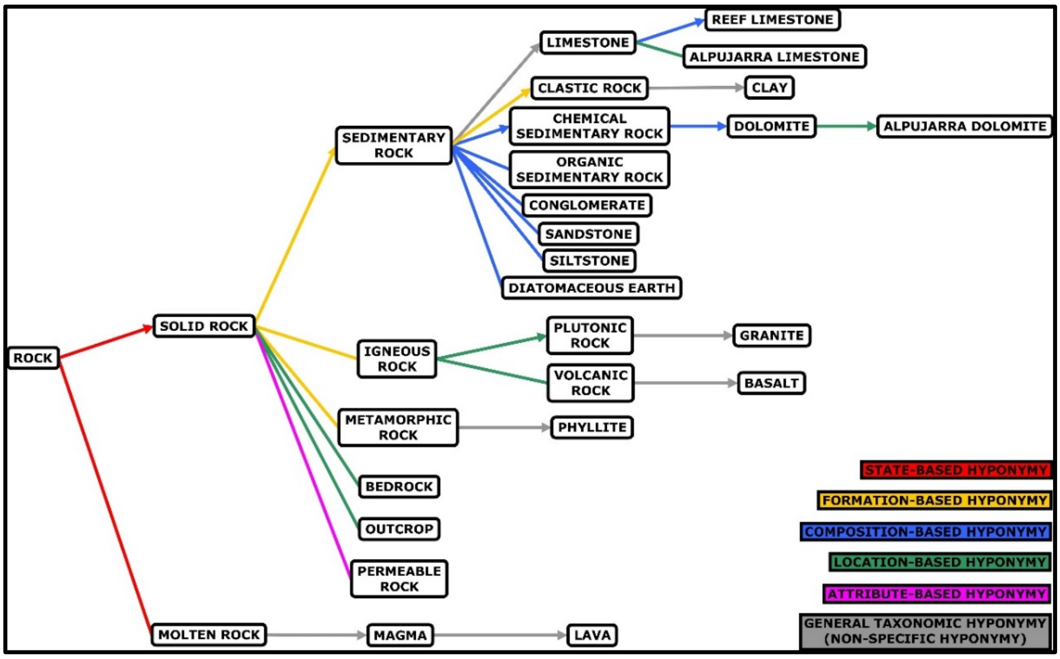

Figure 16. Enhanced tree-mode network of ROCK

Moreover, Table 17 offers some examples of these subtypes found in the conceptual network of ROCK. According to the definitions contained in the EcoLexicon database, for instance, IGNEOUS ROCK is considered to be a *formation-based_type_of* SOLID ROCK because it is "formed by solidification of molten magma"; REEF LIMESTONE is presented as a *composition-based_type_of* LIMESTONE since it is "composed of the remains of sedentary organisms"; and VOLCANIC ROCK is represented as a *location-based_type_of* IGNEOUS ROCK because it is "solidified near or on the surface of the Earth".

It should be noted that this classification for subtypes is, in a certain way, related to the different types of umbrella concept that have been discussed in Section 3.2. (e.g. 'formation-based', 'agent-based', 'function-based'). Not surprisingly, noun compounds or complex nominals are usually located at the first levels of hierarchization, and they can be easily classified with a hyponymy subtype due to their semantic component.

However, not all hyponymic relations can be classified using a subtype. There are certain child concepts whose differentiating features make it impossible to determine only one hyponymy subtype. For example, GRANITE is a *type_of* PLUTONIC ROCK based on its attributes ("coarse-grained, light-colored, hard"), its composition ("consisting chiefly of quartz, orthoclase or microline, and mica") and its function ("used as a building material"). Such cases will remain classified as general taxonomic hyponymy, or as a non-specific *type_of* relation.

Formation-based hyponymy: **(X *formation-based_type_of* Y)**
▪ SEDIMENTARY ROCK < SOLID ROCK ▪ IGNEOUS ROCK < SOLID ROCK ▪ CLASTIC ROCK < SEDIMENTARY ROCK
Composition-based hyponymy: **(X *composition-based_type_of* Y)**
▪ DOLOMITE < CHEMICAL SEDIMENTARY ROCK ▪ ORGANIC SEDIMENTARY ROCK < SEDIMENTARY ROCK ▪ REEF LIMESTONE < LIMESTONE
Location-based hyponymy: **(X *location-based_type_of* Y)**
▪ BEDROCK < SOLID ROCK ▪ VOLCANIC ROCK < IGNEOUS ROCK ▪ ALPUJARRA LIMESTONE < LIMESTONE
State-based hyponymy: **(X *state-based_type_of* Y)**
▪ SOLID ROCK < ROCK ▪ MOLTEN ROCK < ROCK
Attribute-based hyponymy: **(X *attribute-based_type_of* Y)**
▪ PERMEABLE ROCK < SOLID ROCK

Table 17. Examples of hyponymy subtypes found in the conceptual network of ROCK

Nonetheless, this list of subtypes is not a closed inventory of hyponymic relations, but only those which were distinguished so far in the conceptual network of ROCK and similar entities. As future work in this line of research, a minimum number of coincidences will eventually be established to confirm the validity (and usefulness) of a hyponymy subtype.

4.1.2. Hyponymy subtypes in the conceptual network of a process: EROSION

In reference to the improved tree-mode network of EROSION (see Figure 17), up to four process-related subtypes of hyponymy were established:

- **Agent-based hyponymy:** a *type_of* relation dependent on the agent or the promoter that causes the hyponyms.
- **Patient-based hyponymy:** a *type_of* relation dependent on the entity or location affected by the hyponyms.
- **Result-based hyponymy:** a *type_of* relation dependent on the results and effects of the hyponyms.
- **Attribute-based hyponymy:** a *type_of* relation dependent on the traits or features of the hyponyms.

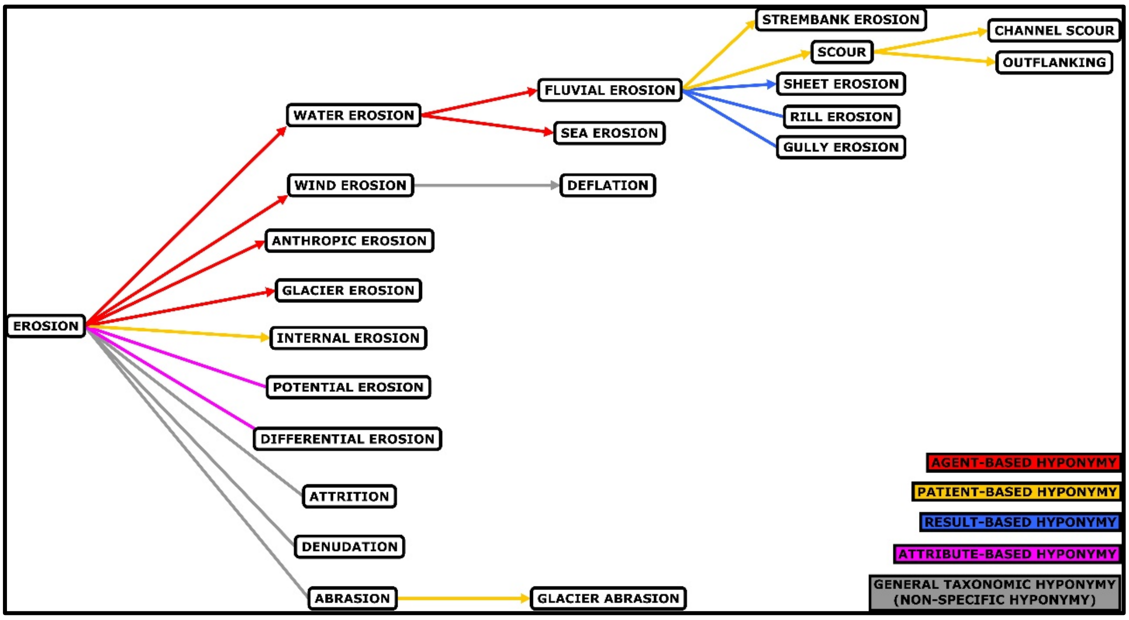

Figure 17. Enhanced tree-mode network of EROSION

Furthermore, Table 18 contains some examples of these hyponymy subtypes found in the conceptual network of EROSION. Based on the definitions contained in the EcoLexicon database, for example, ANTHROPIC EROSION is considered to be an *agent-based_type_of* EROSION because it is “caused by human activities”; GLACIER ABRASION is regarded as a *patient-based_type_of* ABRASION since it is the abrasion “of a glacier bed”; and RILL EROSION is a *result-based_type_of* FLUVIAL EROSION because it “forms small channels”.

Agent-based hyponymy (X *agent-based_type_of* Y)
▪ SEA EROSION < EROSION ▪ ANTHROPIC EROSION < EROSION ▪ FLUVIAL EROSION < WATER EROSION
Patient-based hyponymy (X *patient-based_type_of* Y)
▪ STREAMBANK EROSION < FLUVIAL EROSION ▪ GLACIER ABRASION < ABRASION ▪ CHANNEL SCOUR < SCOUR
Result-based hyponymy (X *result-based_type_of* Y)
▪ SHEET EROSION < FLUVIAL EROSION ▪ RILL EROSION < FLUVIAL EROSION ▪ GULLY EROSION < FLUVIAL EROSION
Attribute-based hyponymy: (X *attribute-based_type_of* Y)
▪ POTENTIAL EROSION < EROSION ▪ DIFFERENTIAL EROSION < EROSION

Table 18. Examples of hyponymy subtypes found in the conceptual network of EROSION

As shown in Table 18, the process-related subtypes of hyponymy are different from those of an entity (except for attribute-based hyponymy, which is common to both). A process is generally a nominalization of a verb, and thus it often involves an agent, a patient, and a result. This differs from formation, composition, and state, which are typical of entities. Moreover, in the case of processes, patient-based hyponymy sometimes overrides location-based hyponymy, as the patient can be a physical location (e.g. CHANNEL SCOUR affects a stream bed, and therefore takes place in a stream bed).

Furthermore, the general taxonomic hyponymy (*type_of*) is also present in processes. In fact, various examples of it can be found in the conceptual network of EROSION. For instance, DENUDATION is a *type_of* EROSION based on its agents ("caused by the action of water, ice, wind and waves"), its patient ("the Earth's surface") and its result ("redistribution of Earth surface material").

In the same way as for entities, these process-related subtypes of hyponymy do not constitute a closed set, since further research is needed to determine the extension and scope of process-related subtypes of hyponymy. Nevertheless, it can be observed that generic-specific relations can be refined to establish subtypes of hyponymy through the analysis of the concepts in a network and their definitions. This helps to solve problems related to redundancy and transitivity-based inconsistencies. It also enhances specialized knowledge representation and acquisition since users can easily determine the specific subsense by means of which two or more concepts are linked with a generic-specific relation.

4.2. Extending the conceptual systems through a semi-automatic corpus-driven extraction

After carrying out the semi-automatic corpus-driven extraction of ROCK hyponyms with the customized word sketches in Sketch Engine (see Section 3.3.), a whole new set of concepts was retrieved and subsequently implemented in EcoLexicon. Figure 18 shows, in red color, the introduced concepts in the hierarchy of ROCK with respect to the previous conceptual system (recently seen in Figure 16).

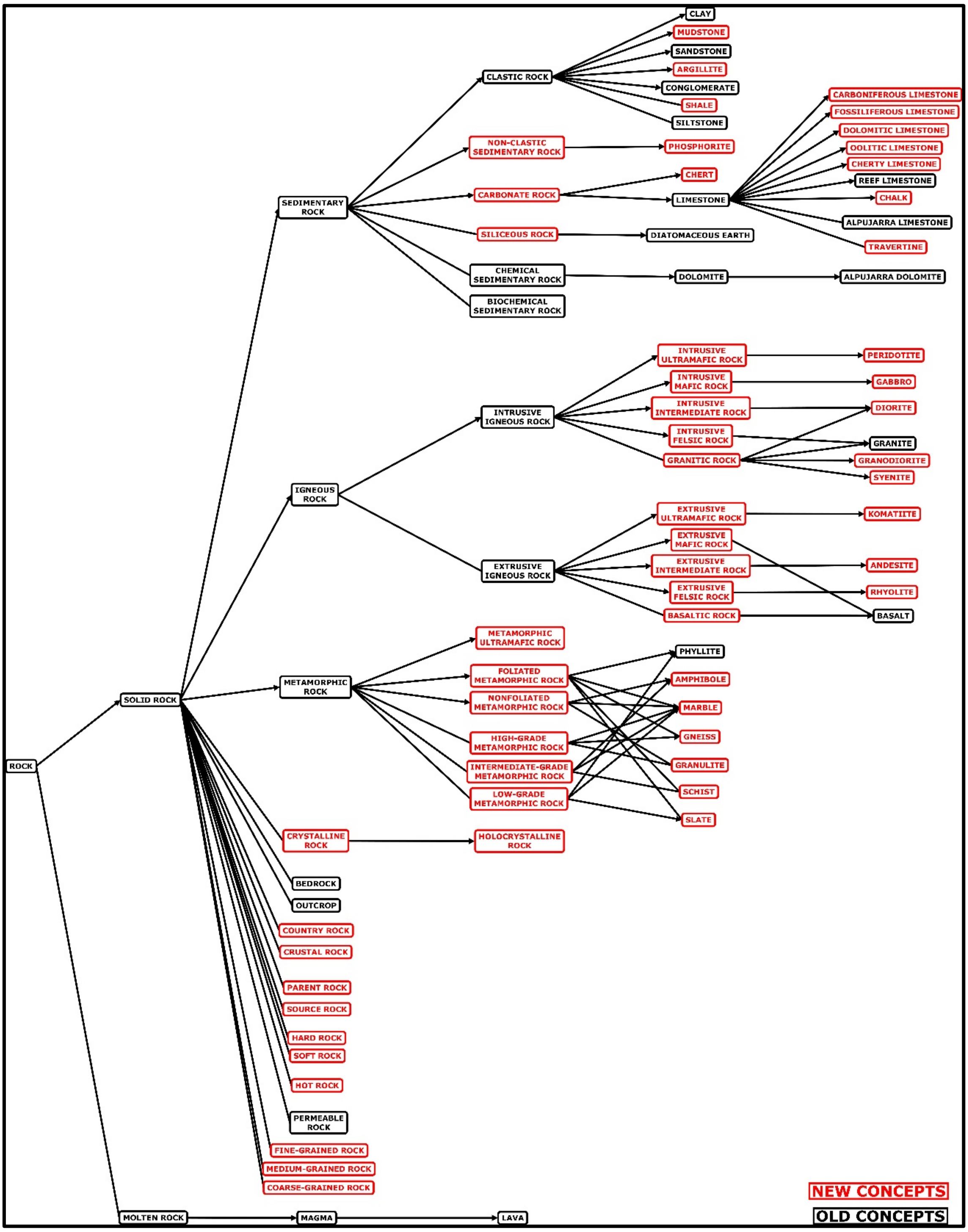

Figure 18. Extended tree-like network of ROCK with the new concepts

In this way, a total of 57 new concepts (including CARBONATE ROCK, HIGH-GRADE METAMORPHIC ROCK, INTRUSIVE ULTRAMAFIC ROCK, HOLOCRYSTALLINE ROCK, and MEDIUM-GRAINED ROCK, to name a few examples) were implemented and located in their corresponding place within the conceptual network of ROCK. Moreover, the lexical representations of three concepts were replaced with preferred terms (with regard to the number of occurrences), as found in the EcoLexicon corpus: ORGANIC SEDIMENTARY

ROCK > BIOCHEMICAL SEDIMENTARY ROCK, VOLCANIC ROCK > EXTRUSIVE IGNEOUS ROCK, and PLUTONIC ROCK > INTRUSIVE IGNEOUS ROCK.

Afterwards, based on the corpus concordances and specialized dictionaries definitions, the new generic-specific relations were classified according to hyponymy subtypes (see Figure 18).

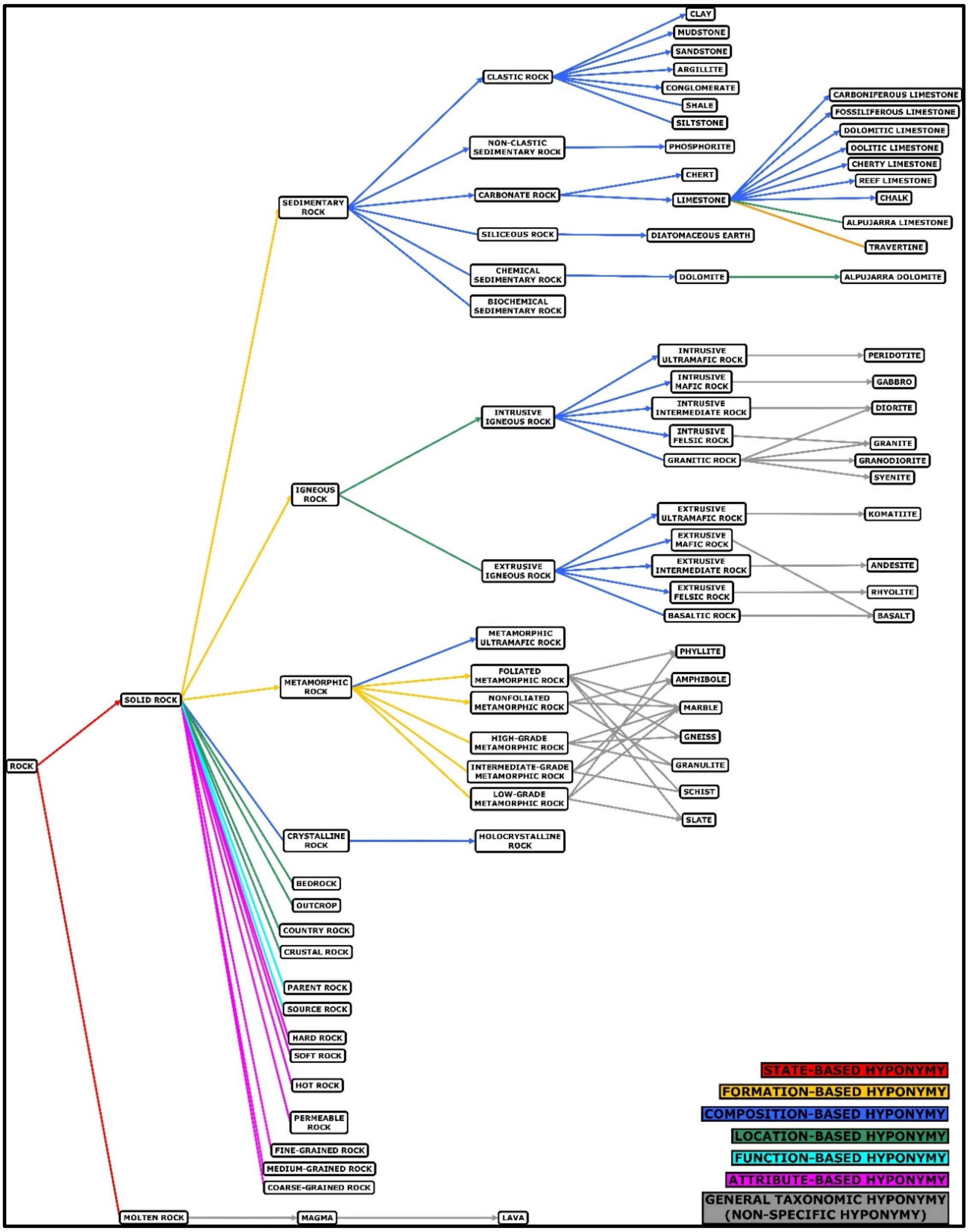

Figure 19. Extended tree-like network of ROCK with hyponymy subtypes

Several remarks can be pointed out from this extended conceptual network, as it reveals a hierarchization of hyponymic subtypes according to their location at different levels. It can be observed that the first hyponymic level, made from ROCK to SOLID ROCK and MOLTEN ROCK, implies state-based hyponymy. Regarding the child concepts to SOLID ROCK, the next step in the hierarchy implies various hyponymy subtypes that represent the belonging of the cohyponyms to different dimensions within the own conceptual system: there are three formation-based hyponyms (SEDIMENTARY ROCK, IGNEOUS ROCK, and METAMORPHIC ROCK), a composition-based hyponym (CRYSTALLINE ROCK), four location-based hyponyms (BEDROCK, OUTCROP, COUNTRY ROCK and CRUSTAL ROCK), two function-based hyponyms (PARENT ROCK and SOURCE ROCK), and seven attribute-based hyponyms (HARD ROCK, SOFT ROCK, HOT ROCK, PERMEABLE ROCK, FINE-GRAINED ROCK, MEDIUM-GRAINED ROCK, and COARSE-GRAINED ROCK). Let it be noted that a new subtype of hyponymy, function-based hyponymy, was revealed after this process of hyponymic extension with a corpus-driven extraction of hyponyms (for instance, SOURCE ROCK refers to those "from which hydrocarbons are capable of being generated"). Moreover, within the attribute-based hyponyms, a subcategorization of this subtype of hyponymy (according to hardness, temperature, permeability and texture) could be established if enough concordances were found in the corpus (see Figure 20).

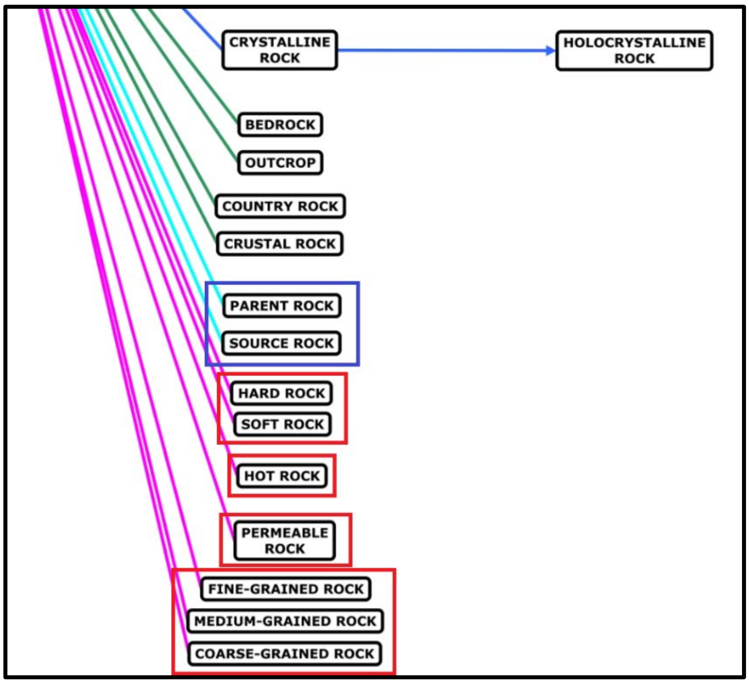

Figure 20. New function-based hyponymy (blue rectangle) and subcategorization of attribute-based hyponymy (red rectangle)

As for the child concepts to SEDIMENTARY ROCK (see Figure 21), it can be observed that virtually at all hyponymic levels the predominant relation is composition-based hyponymy, except for just three concepts (ALPUJARRA LIMESTONE and ALPUJARRA DOLOMITE, which are location-based hyponyms, and TRAVERTINE, which is a formation-based hyponym). In fact, sedimentary rocks are mainly characterized by the nature of their sediments or constituents, so it is not surprising that the main type of generic-specific relation in this conceptual system is based on composition.

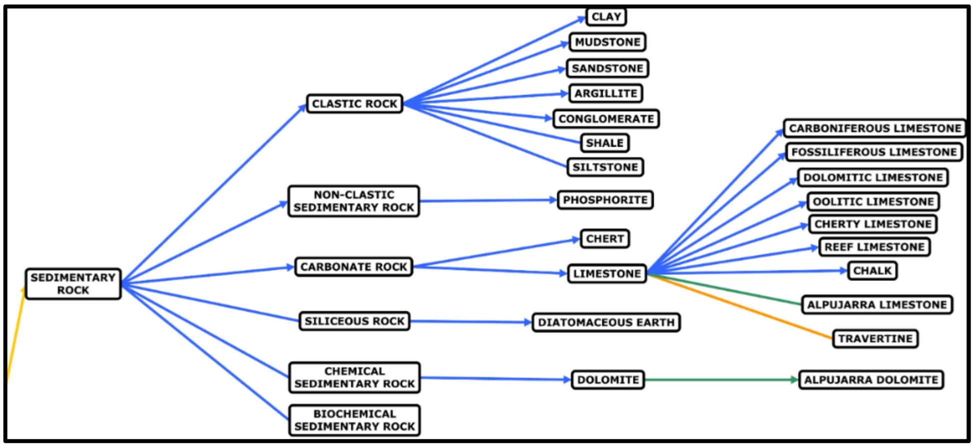

Figure 21. Detail of the SEDIMENTARY ROCK hierarchy

With regard to the concept hierarchy of IGNEOUS ROCK (see Figure 22), an interesting phenomenon of hyponymy subtypes according to levels is observed: location-based hyponymy appears at the first level, composition-based hyponymy at the second level, and a non-specific hyponymy at the third and last level. In Geology, igneous rocks are first classified depending on whether they are formed on the Earth's surface (EXTRUSIVE IGNEOUS ROCK) or within the Earth (INTRUSIVE IGNEOUS ROCK); then, they are classified according to their characteristic components (INTRUSIVE MAFIC ROCK, INTRUSIVE FELSIC ROCK, EXTRUSIVE ULTRAMAFIC ROCK, etc.); and finally, they show their general taxonomic hyponyms (PERIDOTITE, SYENITE, BASALT, etc.)

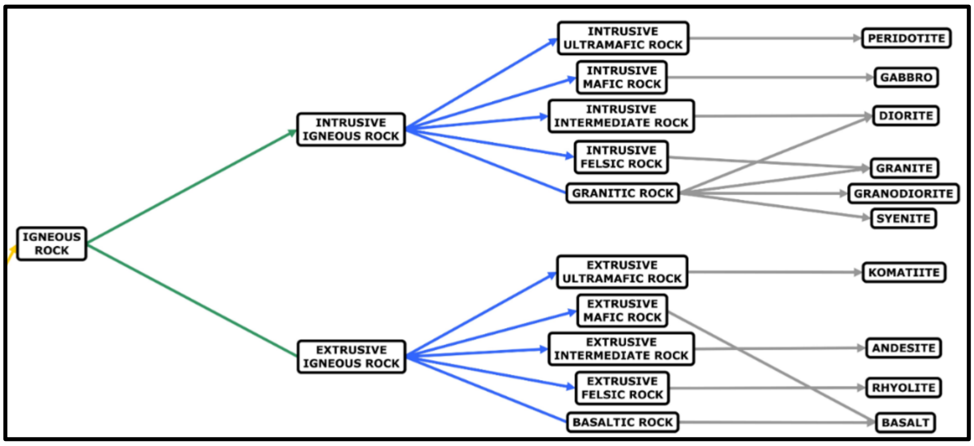

Figure 22. Detail of the IGNEOUS ROCK hierarchy

Regarding the child concepts to METAMORPHIC ROCK (see Figure 23), only one composition-based hyponym (METAMORPHIC ULTRAMAFIC ROCK) was found in the EcoLexicon corpus. Contrary to the two previous concept hierarchies, in this case the main type of generic-specific relation is based on formation, since metamorphic rocks are characterized by a transformation process known as 'metamorphism'. In this way, the first hyponymic level is virtually represented with formation-based hyponymy. However, it should be noted that a subcategorization could be established within the process subtype, as there are formation-based hyponyms that depend on the result of the process (FOLIATED METAMORPHIC ROCK and NONFOLIATED METAMORPHIC ROCK) and on its intensity (HIGH-GRADE METAMORPHIC ROCK, INTERMEDIATE-GRADE METAMORPHIC ROCK, and LOW-GRADE METAMORPHIC ROCK). As in the case of the IGNEOUS ROCK hierarchy, the last hyponymic level of the METAMORPHIC ROCK hierarchy is characterized by general taxonomic hyponyms, but in this network the hyponyms also show multiple inheritance.

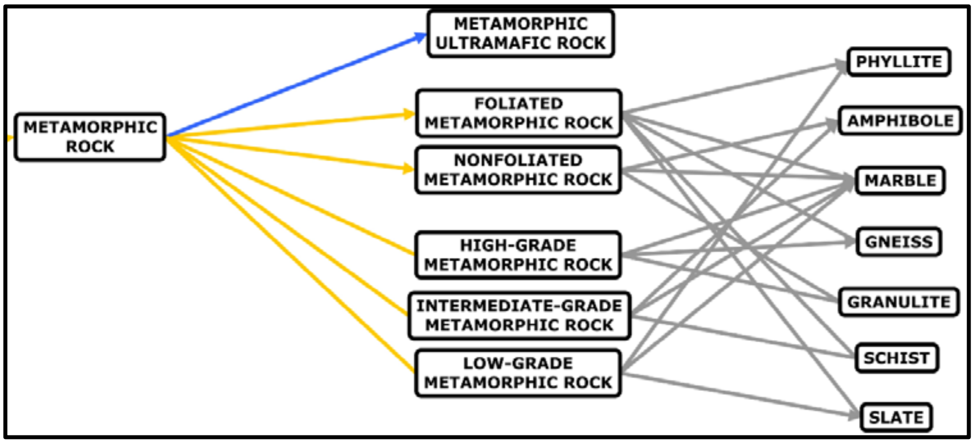

Figure 23. Detail of the METAMORPHIC ROCK hierarchy

5. Conclusion

This final degree project has analyzed how to refine hyponymy in EcoLexicon, a multilingual terminological knowledge base on the environment, so as to extend the conceptual systems and subsequently enhance its representation of multidimensionality.

Contrary to more traditional theories of terminology, which sought the establishment of static standardizing conceptual structures, the contemporary theories of terminology are increasingly placing its focus on dynamic conceptual representation and knowledge organization. Linking specialized knowledge representation to cognitive linguistics and semantics, the study of terminology and specialized communication is, indeed, experiencing a ‘cognitive shift’ that takes into consideration neurological processes and implies an interdisciplinary approach. Likewise, in terminology a series of methods for structuring knowledge according to logical properties have recently been considered, and they seek to facilitate the accurate organization and representation of information. Therefore, it could be stated that this discipline is in constant evolution and that, after the introduction of the cognitive approaches, it is currently experiencing a new golden age.

Furthermore, this final degree project has also discussed the importance of terminological resources in specialized translation, which implies a complex knowledge acquisition process. Not only do translators need to master the writing, understanding, and use of specialized discourse in order to carry out their work, but they also need to have access to effective tools for knowledge acquisition. For this reason, the usefulness of TKBs has been examined, based on their main features: accessibility, dynamism, multidimensionality, and multimodality. When TKBs accurately represent contextual information, they allow users to understand how terms are used in real situations. This makes them an excellent resource for translators.

In this line, this project has assessed the benefits of EcoLexicon, a multilingual TKB on environmental science that displays an accessible visual interface and provides access to conceptual, linguistic, and graphical information. EcoLexicon shows its content in the form of a visual thesaurus (linking concepts through semantic relations), five main features (definition, terms, resources, conceptual categories, and phraseology) and corpus concordances searches.

Nonetheless, EcoLexicon still has aspects that could be improved regarding its representation of specialized knowledge, and that is the reason why this project has dealt with a refinement of hyponymy. Moreover, before doing so, this project has also discussed the main theories on hyponymy by Cruse (2002) and Murphy (2003, 2006).

Throughout this research, the correction of property inheritance in concept definitions has been considered a preliminary though essential phase in the refinement of the *type_of* relation. Another important element which is preliminary to hyponymy refinement is the implementation of umbrella concepts at intermediate levels of a hierarchy, enriching the represented knowledge. After explaining how noun compounds and noun phrases contain semantic meanings that allow them to be classified according to different nuances, this process was used to create and introduce several umbrella concepts in various concept hierarchies in EcoLexicon. Nevertheless, it was observed that umbrella concepts are not only represented with polylexical terms, but also with monolexical terms. Therefore, further research is necessary to determine when they are strictly necessary and the extent to which they should be implemented in the TKB.

After discussing the correction of property inheritance and the enrichment of hierarchies with umbrella concepts, this project has shown how to refine generic-specific relations and establish hyponymy subtypes through the analysis of the concepts in a network and their definitions. In this way, several subtypes of hyponymy have been distinguished for entities (e.g. formation-based hyponymy), for processes (e.g. agent-based hyponymy) and for both types (e.g. attribute-based hyponymy). It has also been demonstrated how this type of refined hyponymy can be implemented in EcoLexicon, thus increasing its informativity for users. However, further research is required to verify the existence of these subtypes of hyponymy in other fields of knowledge, to establish systematic parameters for the creation of new subtypes, and to explore in depth how semantic relations are expressed in nominal clauses and compound nouns (Downing, 1977; Levi, 1978; Rosario & Hearst, 2001; Nastase & Szpakowicz, 2003).

Moreover, it has also been shown how a semi-automatic corpus-driven analysis can provide a faster population and restructuring of conceptual networks. After carrying out this process to enrich and extend the assessed concept hierarchies, it was possible to discover that hyponymy subtypes are usually activated according to hierarchical levels and concept nature. In addition to this, regarding the word sketches that were used during this process, it was noticed that knowledge patterns are not as reliable for process as they are for entities, since their taxonomies are not explicit.

Furthermore, and as future work in this line of research, this final degree project opens the door to verify the existence of hyponymy subtypes in other fields of knowledge, apart from environmental science. Moreover, further research will also involve correcting and validating the proposed enhanced conceptual systems by domain experts. Finally, the sketch grammars will also be refined and expanded so as to offer more accurate results in the queries.

In conclusion, this final degree project has demonstrated that refining hyponymy is a promising approach to the extension of conceptual systems and, in the end, to the enhancement of multidimensional knowledge representation.

6. References

Barrière, C. (2004a). Knowledge-Rich Contexts Discovery. In *Proceedings of the 17th Conference of the Canadian Society for Computational Studies of Intelligence on Advances in Artificial Intelligence* (*Canadian AI 2004)*. London, United Kingdom.

Barrière, C. (2004b). Building a concept hierarchy from corpus analysis. *Terminology*, 10 (2), pp. 241-263. Amsterdam/Philadelphia: John Benjamins.

Bowker, L. (1997). Multidimensional classification of concepts and terms. In Wright, S.E. & Budin, G. (Eds.) *Handbook of Terminology Management: Basic Aspects of Terminology Management*, pp. 133-143. Amsterdam/Philadelphia: John Benjamins.

Buendía Castro, M., & Faber, P. (2015). EcoLexicon como asistente en la traducción. In Corpas Pastor, G., Seghiri Domínguez, M., Gutiérrez Florido, R., & Urbano Mendaña, M. (Eds.) *VII Congreso Internacional de la Asociación Ibérica de Estudios de Traducción e Interpretación: Nuevos horizontes en los Estudios de Traducción e Interpretación (Comunicaciones completas) / New Horizons in Translation and Interpreting Studies (Full papers)*, pp. 195-203. Geneva: Tradulex.

Cabré, M. T. (1999). *La Terminología: Representación y Comunicación*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

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

Downing, P. (1977). On the Creation and Use of English Compound Nouns. *Language*, 53 (4), pp. 810-842.

Dubuc, R. & Lauriston, A. (1997). Terms and Contexts. In Wright, S. E., & Budin, G. (Eds.) *Handbook of Terminology Management*, 1, pp. 80-88. Amsterdam/ Philadelphia: John Benjamins.

Dury, P. (2005). Terminology and Specialized Translation: The Relevance of the Diachronic Approach. *LSP and Professional Communication*, 5 (1), pp. 31-42.

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

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

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

Faber, P., & San Martín, A. (2010). Conceptual Modeling in Specialized Knowledge Resources. *Information Technologies & Knowledge*, 4 (2), pp. 110-121.

Faber, P., León-Araúz, P., & Reimerink, A. (2014). Representing Environmental Knowledge in EcoLexicon. In Bárcena, E., Read, T., & Arús, J. (Eds.) *Languages for Specific Purposes in the Digital Era*, 19, pp. 267-301. Berlin/Heidelberg: Springer.

Faber, P., Márquez Linares, C., & Vega Expósito, M. (2005). Framing Terminology: A Process-Oriented Approach. *Meta : journal des traducteurs / Meta: Translators' Journal*, 50 (4).

García Aragón, A., Buendía Castro, M., & López Rodríguez, C.I. (2014). Evaluación de una base de conocimiento terminológica sobre el medio ambiente en el aula de la traducción especializada. In Vargas Sierra, C. (Ed.) *TIC, trabajo colaborativo e interacción en Terminología y Traducción*, pp. 477-487. Granada: Comares.

Gheorghita, I., & Pierrel, J. (2012). Towards a Methodology for Automatic Identification of Hypernyms in the Definitions of Large-Scale Dictionary. In Nicoletta Calzolari (Conference Chair) et al. (Eds.) *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC2012)*, pp. 2614-2618. Istanbul, Turkey: ELRA.

Giacomini, L. (2014). Testing User Interaction with LSP e-Lexicographic Tools: A Case Study on Active Translation of Environmental Terms. In *Proceedings of the 12th edition of the Konvens Conference*, pp. 77-85. Hildesheim, Germany.

Giacomini, L. (2015). Management and exploitation of conceptual data and information in technical termbases: the electrotechnical vocabulary. In Kosem, I., Jakubiček, M., Kallas, J., Krek, S. (Eds.) *Electronic lexicography in the 21st century: linking lexical data in the digital age*. *Proceedings of the eLex 2015 conference*, pp. 186-197. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd.

Gil-Berrozpe, J.C., & Faber, P. (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)*, pp. 8-15. Portorož, Slovenia.

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

Jouis, C. (2006). Hierarchical Relationships “is-a”: Distinguishing Belonging, Inclusion and Part/of Relationships. In Nicoletta Calzolari (Conference Chair) et al. (Eds.) *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC2006)*, pp. 571-574. Genoa, Italy: ELRA.

Kageura, K. (1997). Multifaceted/Multidimensional concept systems. In Wright, S.E. & Budin, G. (Eds.) *Handbook of Terminology Management: Basic Aspects of Terminology Management*, pp. 119-132. Amsterdam/Philadelphia: John Benjamins.

León-Araúz, P., & Faber, P. (2010). Natural and Contextual Constraints for Domain-Specific Relations. In Mititelu, V. B., Pekar, V., & Barbu, E. (Eds.) *Proceedings of the Workshop Semantic Relations, Theory and Applications*, pp. 12-17. Valletta, Malta.

León-Araúz, P., & Reimerink, A. (in press). Multidimensional categorization in corpus-based hyponymic structures. In Alonso, A., Cruz, L., González, V. (Eds.), *Corpus-*

based studies on language varieties, Linguistic Insights, 210. Bern/Berlin/Bruxelles/Frankfurt am Main/New York/Oxford/Wien: Peter Lang.

León-Araúz, P., & San Martín, A. (in press). Knowledge Patterns as Word Sketches in Sketch Engine. In *Proceedings of META-FORUM 2016 (4-5 July 2016)*. Lisbon, Portugal.

León-Araúz, P., Faber, P., & Montero Martínez, S. (2012). Specialized Language Semantics. In Faber, P., (Ed.) *A Cognitive Linguistics View of Terminology and Specialized Language*, pp. 95-175. Berlin/Boston: De Gruyter Mouton.

León-Araúz, P., Reimerink, A., & García Aragón, A. (2010). Context-based Modelling of Specialized Knowledge. *Information Technologies & Knowledge*, 4 (2), pp. 122-142.

León-Araúz, P., Reimerink, A., & García Aragón, A. (2013). Dynamism and Context in Specialized Knowledge. *Terminology*, 19 (1), pp. 31-61. Amsterdam/Philadelphia: John Benjamins.

Leonardi, P. M. (2010). Digital Materiality? How Artifacts Without Matter, Matter. *First Monday*, 15 (6). Available from: http://firstmonday.org/article/view/3036/2567

Levi, J. (1978). *The syntax and semantics of complex nominals*. New York: Academic Press.

López Rodríguez, C.I., Prieto Velasco, J.A., & Tercedor Sánchez, M. (2013). Multimodal representation of specialized knowledge in ontology-based terminological databases: the case of EcoLexicon. *The Journal of Specialized Translation*, 20, pp. 49-67.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of Age: A Review of Embodiment and the Neuroscience of Semantics. *Cortex*, 48, pp 788-804. doi: 10.1016/j.cortex.2010.11.002

Meyer, I., Bowker, L. & Eck, K. (1992). COGNITERM: An Experiment in Building a Knowledge-Based Term Bank. In Tommola, H., Varantola, K., Salmi-Tolonen, T.,

& Schopp, J., (Eds.) *Proceedings of the Fifth EURALEX International Congress (EURALEX '92)*, pp. 159-172. Tampere, Finland: Tampereen Yliopisto.

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

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

Nastase, V., & Szpakowicz, S. (2003). Exploring noun-modifier semantic relations. In *Fifth International Workshop on Computational Semantics (IWCS-5)*, pp. 285-301, Tilburg, The Netherlands.

Nastase, V., Sayyad-Shirabad, J., Sokolova, M., & Szpakowicz, S. (2006). Learning noun-modifier semantic relations with corpus-based and WordNet-based features. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06)*, pp. 781-786.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where Do You Know What You Know? The Representation of Semantic Knowledge in the Human Brain. *Nature Reviews Neuroscience*, 8, pp. 976-T988.

Prieto Velasco, J.A. (2009). To see or not to see: concept visualization in terminological knowledge bases. In *8th International Conference on Terminology and Artificial Intelligence*, pp. 220-231. Toulouse, France.

Reimerink, A., León-Araúz, P., Magaña Redondo, P. J. (2010). EcoLexicon: An Environmental TKB. In Nicoletta Calzolari (Conference Chair) et al., *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC2010)*, pp. 2322-2329. Valletta, Malta.

Roche, C., Calberg-Challot, M., Damas, L., & Rouard. P. (2009). Ontoterminology: A New Paradigm for Terminology. In Dietz, J. L. G. (Ed.) *Proceedings of the International Conference on Knowledge Engineering and Ontology Development (KEOD 2009)*, 2626–2630. Madeira, Portugal.

Rosario, B., & Hearst, M. (2001). Classifying the Semantic Relations in Noun Compounds via a Domain-Specific Lexical Hierarchy. In *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing (EMNLP-01)*, pp. 82-90.

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

Tercedor, M., López Rodríguez, C. I., & Faber, P. (2012). Working with Words: Research Approaches in Translation-oriented Lexicographic Practice. *TTR: Traduction, Terminologie, Rédaction,* 25 (1), pp. 181-214. Association canadienne de traductologie.

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

Annex

1. Simplified KP-based sketch grammars

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