Proceedings of the Computational Linguistics-Applications Conference pp. 9–16
ISBN 978-83-60810-47-7

Environmental knowledge in EcoLexicon

Pilar León Araúz, Arianne Reimerink and Pamela Faber
Department of Translation and Interpreting, University of Granada
Buensuceso 11, 18002, Granada, Spain
Email: {pleon, arianne, pfaber}@ugr.es

***Abstract*—EcoLexicon is a multilingual terminological knowledge base (TKB) on the environment that targets different user groups who wish to expand their knowledge of the environment for the purpose of text comprehension and/or generation. Users can freely access EcoLexicon, and are able to find the information needed, thanks to a user-friendly visual interface with different modules for conceptual, linguistic, and graphical data. The main goal of this TKB is user knowledge acquisition. This paper briefly explains the theoretical premises and methodology applied in EcoLexicon for knowledge extraction and representation. It also shows how environmental concepts are represented, interrelated, and contextualized. EcoLexicon combines the advantages of a relational database, allowing for a quick deployment and feeding of the platform, and an ontology, enhancing user queries. The internal coherence at all levels of a dynamic knowledge representation shows that even complex domains can be represented in a user-friendly way.**

I. Introduction

ECOLEXICON[1] is a multilingual terminological knowledge base (TKB) on the environment. The knowledge base was initially implemented in Spanish, English and German. Currently, three more languages are being added: Modern Greek, Russian and Dutch. So far it has 3,250 concepts and 14,550 terms. It targets different user groups, such as translators, technical writers, environmental experts, etc., who wish to expand their knowledge of the environment for the purpose of text comprehension or generation. These users can freely access EcoLexicon, and are able to find the information needed, thanks to a user-friendly visual interface with different modules for conceptual, linguistic, and graphical data. The main and ultimate goal of EcoLexicon is user knowledge acquisition, which can only be achieved if TKBs account for the natural dynamism of knowledge mainly caused by context and multidimensionality.

II. Theoretical underpinnings of EcoLexicon

EcoLexicon is primarily based on theoretical and methodological premises derived from cognitive linguistics and corpus linguistics. Context and situated cognition are the semantic foundations of our knowledge representation framework, whereas corpus analysis guides our knowledge extraction procedures.

A. Knowledge extraction

According to corpus-based studies, when a term is studied in its linguistic context, information about its meaning and its use can be extracted [1]–[3]. For EcoLexicon, two corpora were created: a textual corpus and a visual corpus. The English textual corpus (5 million words) consists of specialized texts (e.g., scientific journal articles, PhD theses, etc.), semi-specialized texts (textbooks, manuals, etc.), and texts for the general public, all belonging to the multidisciplinary domain of the environment. The visual corpus consists of images selected according to the following criteria: iconicity, abstraction, and dynamism as ways of referring to and representing specific attributes of specialized concepts. Images were classified in terms of the morphological features described by Marsh and White regarding the functional relationship between images and texts [4].

The extraction of conceptual knowledge from the textual corpus combines manual direct term searches and knowledge pattern analysis. According to many research studies, knowlede patterns (KPs) are considered to be one of the most reliable methods for knowledge extraction [5]–[9]. This involves several complementary steps. Normally, the most recurrent knowledge patterns for each conceptual relation identified in previous research are used to find related term pairs [10], [11]. Afterwards, these terms become seed words that are used for direct term searches to find new KPs and relations. The methodology consists of the cyclic repetition of both procedures. Although previous studies propose a semi-automatized annotation-based approach, first of all certain selection criteria must be defined by manually identifying what information is useful, why it is useful, and how it is structured.

Conceptual concordances of EROSION show how different KPs convey different relations with other specialized concepts. The main relations reflected in EROSION concordances are *caused_by*, *affects*, *has_location*, and *has_result*, which highlight the procedural nature of the concept and the important role played by non-hierarchical relations in knowledge representations.

In Figure 1, EROSION is related to various kinds of agents, such as STORM SURGE (1, 7), WAVE ACTION (2, 13), RAIN (3), WIND (4), JETTY (5), CONSTRUCTION PROJECTS (6), MANGROVE REMOVAL (8), SURFACE RUNOFF (9), FLOOD (10), HUMAN-INDUCED FACTORS (11), STORM (12) and MEANDERING CHANNELS (14). They can be retrieved thanks to all KPs expressing the relation *caused_by*, such as *resultant* (1), *agent for* (2, 3), *due to* (6, 7), *responsible for* (11) and *lead to* (13). This relation can also be conveyed through compound adjective phrases, such as *flood-induced* (10) or *storm-caused* (12) and any expression containing *cause* as a

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

```
Caused_by

1    , Alabama. Significant storm surge and resultant beach erosion were  associated with Ivan's landfall. However,
2    nd climate  On the Castellón coast, the main agent for erosion is wave action, and this is therefore  responsi
3    f a stream. The first   factor, rain, is the agent for erosion, but the degree of erosion is governed by   oth
4    rts (BW) and semiarid steppe (BS). Wind can also cause erosion and deposition in environments where sediments
5    etty. Reflection of waves from a jetty may  also cause erosion of adjacent shorelines. However, erosion furthe
6    oastal zone management. However, in some cases coastal erosion\par  can be due to construction projects that a
7    tude of about 0,3 M m3 per year.  Acute erosion  Acute erosion due to storm surges  (waves and water levels at
8    er. Mangrove removal is also reported to cause coastal erosion and change sedimentation patterns and shoreline
9    [edit] Erosion  Surface runoff is one of the causes of erosion of the earth's surface. Reduced   crop producti
10   pes. Local disturbances, for instance by flood-induced erosion, redistribution of sediment or accumulation of
11   ors and human-induced factors  responsible for coastal erosion and highlight the time and space patterns withi
12   ocess is typical of a cyclical process of storm-caused erosion in winter, followed by progradation\par  owing
13   can cause excessive wave action that can lead to beach erosion. Trash dumped from boats can be washed up onto
14    that have reached base level develop broad valleys by erosion caused by meandering channels. The stream chann

Affects

15   ing these sensitive creatures.  In some cases, coastal erosion can have adverse effects on water quality and h
16   ine Depositional Coasts  The erosion of coastlines and erosion of sediments being carried to the   shoreline b
17    use of dredged material to restore beaches damaged by erosion. EPA works with the U.S. Coast Guard to regulat
18   reasonable points, though when push comes to shove and erosion threatens buildings, traditional beach maintena
19   ks and arches found on irregular rocky coastlines; and erosion provides the material which forms deltas and ba
20    near the base of the cliff. Constant undercutting and erosion causes the cliffs to retreat landward.

Has_location

21   ed by the position of sand accumulation\par  and beach erosion around littoral barriers. A coastal structure i
22   hes. Kuenen (1950) estimates\par  that beach and cliff erosion along all coasts of the world totals about 0.12g
23   ce and divergence of wave energy over an offshore bar, erosion downdrift of a structure such as a groin, sudde
24   proportional to the longshore transport rate, and\par  erosion takes place downdrift at about the same rate. T

Has_result

25    Excessive loads of silt and other sediments caused by erosion can suffocate bottom-dwelling plants and animal
26   islands or coral reefs.  Primary coasts are created by erosion (the wearing away of soil or rock), deposition
27   \par  transported. Beach material is also derived from erosion of the coastal formations caused by\par  waves
28   ed to the   passage of the ice. Shorelines produced by erosion of glacial till deposits differ   markedly from
29   beaches   and marshes, are being formed as a result of erosion   and transportation of unconsolidated material
30   ion of the seashore and a rise in SLR. The results  of erosion could lead to further seawater intrusion that c
31   fs are developed  in landslide debris. In this cliffs, erosion of softer material has created bays. The  expec
32   s of steep systems, a sea-level rise may cause coastal erosion resulting in profile steepening, and therefore
```

Fig. 1. Non-hierarchical relations associated with EROSION

verb or noun: *one of the causes of* (9), *cause* (4, 5, 8) and *caused by* (14).

EROSION is also linked to the patients it *affects*, such as WATER (15), SEDIMENTS (16), COASTLINES (16), BEACHES (17), BUILDINGS (18), DELTAS (19) and CLIFFS (20). However, the affected entities, or patients, are often equivalent to locations (eg. if EROSION *affects* BEACHES it actually *takes place* at the BEACH). The difference lies in the kind of KPs linking the propositions. The *affects* relation is often reflected by the preposition *of* (10) or by verbs like *threatens* (18), *damaged by* (17) or *provides* (19). In contrast, the *has_location* relation is conveyed through directional prepositions (*around*, 21; *along*, 22; *downdrift*, 23) or spatial expressions, such as *takes place* (24). In this way, EROSION is linked to the following locations: LITTORAL BARRIERS (21), COASTS (22) and STRUCTURES (23). *Result* is an essential dimension in the description of any process since it is not only initiated by an agent affecting a patient in a particular location, but also has certain effects, which can be the creation of a new entity (SEDIMENTS, 25; PRIMARY COASTS, 26; BEACH MATERIAL, 27; SHORELINES, 28; MARSHES, 29; BAYS, 31) or the beginning of another process (SEAWATER INTRUSION, 31; PROFILE STEEPENING, 32).

As can be seen, all these related concepts are quite heterogeneous. They belong to different paradigms in terms of category membership and/or hierarchical range. For instance, some of the agents of EROSION are natural (WIND, WAVE ACTION) or artificial (JETTY, MANGROVE REMOVAL) and others are general concepts (STORM) or very specific ones (MEANDERING CHANNEL). This explains why knowledge extraction must still be performed manually. Nevertheless, it also illustrates one of the major problems in knowledge representation: multidimensionality [12]. This is better exemplified in the following concordances (Figure 2), since multidimensionality is most often codified in the *is_a* relation.

In the scientific discourse community, concepts are not always described in the same way because they depend on perspective and subject-fields. For instance, EROSION is described as a natural process of REMOVAL (33), a GEOMORPHOLOGICAL PROCESS (34), a COASTAL PROCESS (35) or a STORMWATER IMPACT (36). The first two cases can be regarded as traditional ontological hyperonyms. The choice of one or the other depends on the upper-level structure of the representational system and its level of abstraction. However, COASTAL PROCESS and STORMWATER IMPACT frame the concept in more concrete subject-fields and referential settings.

The same applies to subtypes, where the multidimensional nature of EROSION is clearly shown. EROSION can thus be classified according to the dimensions of *result* (SHEET, RILL, GULLY, 37; DIFFERENTIAL EROSION, 38), *direction* (LATERAL, 39; HEADWARD EROSION, 49), *agent* (WAVE, 41; FLUVIAL, 42; WIND, 43, 46; WATER, 44; GLACIAL EROSION; 45) and *patient* (SEDIMENT, 47; DUNE, 48; SHORELINE EROSION, 49). In section III, the consequences of multidimensionality for knowledge representation are shown.

B. Knowledge representation

According to Meyer et al. [13], TKBs should reflect conceptual structures similarly to how concepts are related in the mind. The organization of semantic information in the brain should thus underlie any theoretical assumption concerning the

```
Is_a

33   vided by the  area (A) of the drainage basin       (1) Erosion is the natural process of removal of soil by wa
34   in the Netherlands, geomorphological processes such as erosion, transport and sedimentation of sandy materials
35   BURY AND DUXBURY, 1996).     Coastal processes such as erosion and accretion are site-specific, season-specifi
36   these  catchments include:  Stormwater impacts such as erosion, channelisation, sediment deposition and sedime

Type_of

37   eroded by shallow overland flow (sheet, rill and gully erosion) and delivered to the drainage network. Channel
38   m the great   local relief, the result of differential erosion by glacier ice. Figure 9-20   includes two sche
39   ing flood events, the dikes are subject to the lateral erosion of the river   trying to reoccupy its former co4
40   d enlarges these small channels and generates headward erosion directed towards the aggrading active channel (
41   out five percent of the material on most beaches. Wave erosion of rocky coasts is usually slow, even where the
42    of the Earth´s land surface is   dominated by fluvial erosion, lakes that do occur are threatened with either
43   ind   climate, topography and surface roughness.  Wind erosion risk applies only when soils are dry and not co
44   oportional to the steepness of the land surface. Water erosion is in proportion to the shear stress exerted by
45   ley to become both wider and deeper over time. Glacial erosion also results in a change in the valley's cross-
46   dominate in periglacial environments: nivation; eolian erosion and deposition; and fluvial erosion and deposit
47    erosion processes. 215 CHAPTER 13 EQUATIONS: SEDIMENT Erosion caused by rainfall and runoff is computed with
48   gineers to simulate  cross-shore beach, berm, and dune erosion produced by storm waves and water levels. The l
49   uctures constructed to date have resulted in shoreline erosion   in their lee. Furthermore, the key environmen
```

Fig. 2. Hierarchical relations associated with EROSION

retrieval and acquisition of specialized knowledge concepts as well as the design of specialized knowledge resources [14].

Furthermore, since categorization itself is a dynamic context-dependent process, the representation and acquisition of specialized knowledge should certainly focus on contextual variation. From a neurological perspective, Barsalou [15] states that a concept produces a wide variety of situated conceptualizations in specific contexts, which clearly determines the type and number of concepts to be related to.

Context has been explored in some depth by disciplines such as psychology, linguistics, and artificial intelligence. Even though all of these approaches have provided valuable insights, there seems to be no consensus on the definition of context since it is invariably conceived for different purposes, depending on the field.

In linguistics, context is especially mentioned in relation to pragmatic and cognitive notions, such as *speech acts* [16], [17], *conventions* [18], *maxims* [19], *Relevance Theory* [20], *framing* [21], and *common ground* [22].

From a computational perspective, contexts are useful to put together a set of related axioms. In this way, contexts are a means for referring to a group of related assertions about which something can be said [23]. Since context, knowledge, and reasoning are closely intertwined [24], artificial intelligence formalizes context to perform automatic inferences and reasoning [23], [25]; to identify relational constraints for context-aware applications [26]; to improve automatic information retrieval; to resolve ambiguities in natural language processing, *inter alia*.

Nevertheless, whatever the approach, context is defined as a dynamic construct. It is thus surprising that term bases are often restricted to generic-specific and part-whole relations, when conceptual dynamism can only be fully reflected through non-hierarchical relations. These are mostly related to the notions of movement, action, and change, which are directly linked to human experience and perceptually salient conceptual features.

Dynamism in the environmental domain comes from the effects of context on the way concepts are interrelated. In EcoLexicon, this is reflected through: (1) the elaboration of category membership templates; (2) the inclusion of multimodal information associated with each entry; (3) the representation of multidimensionality and the situated nature of concepts through an inventory of both hierarchical and non-hierarchical relations (*is_a, part_of, delimited_by, causes, located_in, effected_by, made_of, has_function, result_of, takes_place_in, affects, phase_of, attribute_of*).

In our approach, we consider that a given utterance does not have a meaning, but rather a meaning potential that will always be exploited in different ways that are dependent upon the discourse context [27]. In this sense, we believe that the formalization of context should account for the relational constraints shown by specialized concepts according to their situational nature.

III. A PRACTICAL DEMONSTRATION OF ECOLEXICON

Each entry in EcoLexicon provides a wide range of interrelated information. Figure 3 shows the entry for EROSION. Users are not obliged to view all this information at the same time, but can browse through the interface depending on their needs.

Under the tag Domains, an ontological structure shows the exact position of the concept in the class hierarchy. EROSION, for example, *is_a* natural process of loss (bottom-left corner of the window). The concept definition is shown when the cursor is placed on the concept. Contexts and concordances appear when clicking on the terms, and inform different users about both conceptual and linguistic aspects. Graphical resources are displayed when clicking on the links in the box Resources (in the left-hand margin towards the middle). At a more fine-grained level, conceptual relations are displayed in a dynamic network of related concepts (right-hand side of the window). Users are free to click on any of these concepts and thus further expand their knowledge of this sector of the specialized domain. The terminological units, under the tag Terms provide linguistic information, and show the designations of the concept in English, Spanish, German, and Modern Greek.

In the next sections we will focus step by step on the main resources of EcoLexicon and explain how they are interrelated (relational database and ontology, semantic networks and definitions and images).

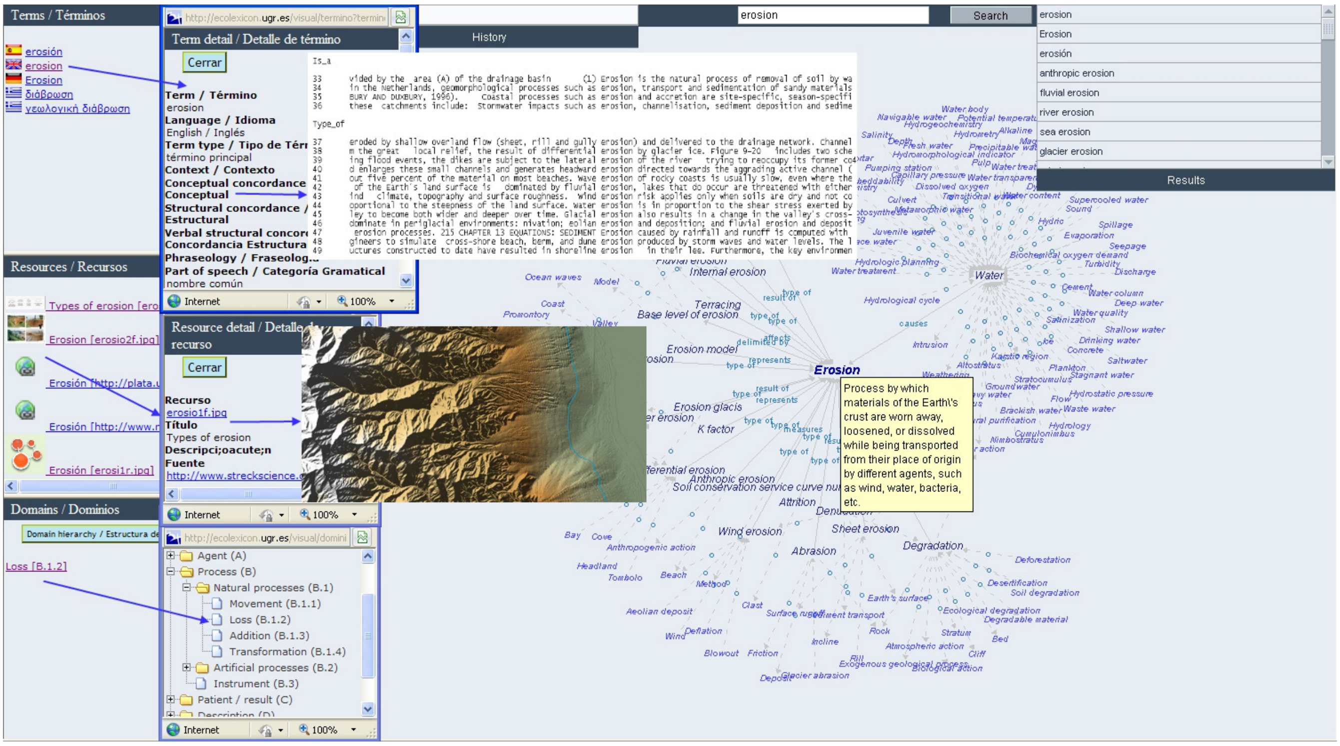


Fig. 3. EcoLexicon user interface

A. Relational Database and Ontology

Data in EcoLexicon are primarily hosted in a relational database (RDB). Nevertheless, relational modelling has some limitations, such as its limited capability to represent real-world entities since natural human implicit knowledge cannot be inferred. Relational models are suited to organize data structure and integrity, whereas ontologies try to specify the meaning of their underlying conceptualization [28]. In EcoLexicon, semantic information is stored in the ontology, while the rest is stored in the relational database. Upper-level classes in our incipient ontology correspond to the basic semantic roles identified for the environmental domain (agent-process-patient-result-location). Conceptual relations in EcoLexicon are enhanced by an additional degree of OWL semantic expressiveness provided by property characteristics. In fact, one of the main advantages of ontologies is that they make reasoning and inferences possible. For example, *part_of* relations can benefit from transitivity, as shown in Figure 4.

In Figure 4, a SPARQL query is made in order to retrieve which concepts are *part_of* Concept 3262, which refers to the concept SEWER. On the right side, DRAINAGE SYSTEM is retrieved as a direct *part_of* relation, whereas SEWAGE COLLECTION AND DISPOSAL SYSTEM and SEWAGE DISPOSAL SYSTEM are implicitly inferred through the Jena reasoner.

However, meronymy cannot always be a transitive relation. This is why six different meronymic relations have been defined. For example, if *located_at* were considered as a

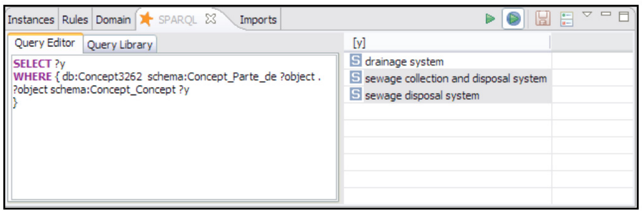


Fig. 4. Concept SEWER in the ontology and inferred transitivity

part_of relation, that would cause fallacious transitivity [29]. If a GABION is *part_of* a GROYNE and a GROYNE *part_of* the SEA, the ontology would infer that GABIONS are *part_of* the SEA, which is false. However, it is true that if a HARD DEFENCE STRUCTURE is *located_at* the BEACH and the BEACH is *part_of* the COAST, then the DEFENCE STRUCTURE is *located_at* the COAST. In this sense, "property chain inclusions", as defined in W3C recommendations, will soon be implemented in EcoLexicon [30].

B. Semantic networks: context and dynamism

According to corpus-based information, concepts in EcoLexicon appear related to others in the form of multidimensional semantic networks. Multidimensionality is commonly regarded as a way of enriching traditional static representations by enhancing knowledge acquisition through different points of view in the same semantic network [30]. However, multidimensionality in the environmental domain

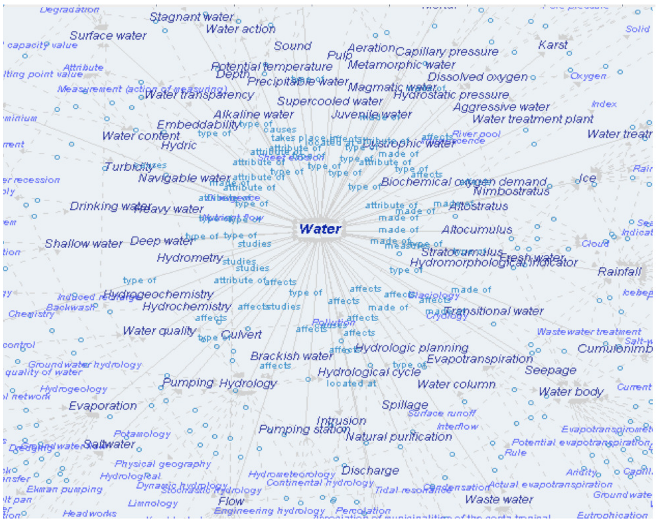


Fig. 5. Information overload in the network of WATER

has caused a great deal of information overload, which ends up jeopardizing knowledge acquisition.

This is mainly caused by versatile concepts, such as WATER (Figure 5), which are usually top-level general concepts involved in a myriad of events. For instance, in its conceptual network, WATER is linked to the same extent to diverse natural and artificial processes, such as EROSION or DESALINATION. Corpus data has provided 72 conceptual relations for the first hierarchical level of WATER.

However, WATER rarely, if ever, activates those relations at the same time, as they evoke completely different situations. Our claim is that any specialized domain contains sub-domains in which conceptual dimensions become more or less salient, depending on the activation of specific contexts. As a result, a more believable representational system should account for re-conceptualization according to the situated nature of concepts. In EcoLexicon, this is done by dividing the global specialized environmental field in different discipline-oriented contextual domains: HYDROLOGY, GEOLOGY, METEOROLOGY, BIOLOGY, CHEMISTRY, ENGINEERING, WATER TREATMENT, etc. In conceptual modelling, facets and contexts can be established in a myriad of different ways. However, in EcoLexicon, a discipline-oriented approach was found the most appropriate. After all, multidimensionality in the environmental domain is often caused by the fact that each discipline deals with the concepts in different terms.

Contextual constraints are neither applied to individual concepts nor to individual relations. Instead, they are applied to each conceptual proposition. For instance, CONCRETE is linked to WATER through a *made_of* relation, but this proposition is not relevant if users only want to know how WATER naturally interacts with landscape. Consequently, that proposition will only appear in an ENGINEERING context [31]. Nevertheless, not only versatile concepts, such as WATER, are constrained, since information overload can also affect any other concept that is linked to versatile ones. For instance,

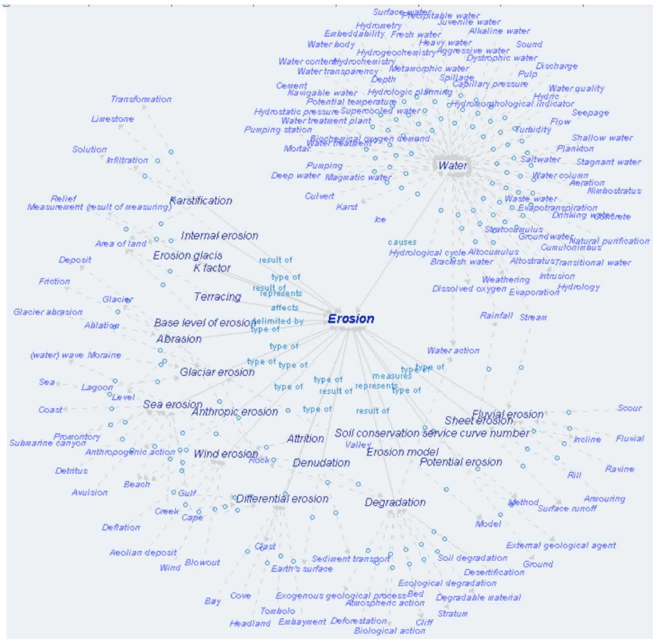


Fig. 6. EROSION context-free network

EROSION takes the following shape in a context-free network (Figure 6), which appears overloaded mainly because it is closely linked to WATER as one of its most important agents.

When contextual constraints are applied, EROSION only appears linked to propositions belonging to the context of GEOLOGY (Figure 7) or HYDROLOGY (Figure 8).

Comparing both networks and especially focusing on EROSION and WATER, the following conclusions can be drawn. The number of conceptual relations changes from one network to another since EROSION is not equally relevant in both domains. EROSION is a prototypical concept of GEOLOGY, and thus participates in more propositions in that domain than in HYDROLOGY. Nevertheless, since it is also strongly linked to WATER, HYDROLOGY is also an essential domain in the representation of EROSION. Relation types do not substantially change from one network to the other, but the GEOLOGY domain shows a greater number of *type_of* relations. This is due to the fact that HYDROLOGY only includes types of EROSION whose agent is WATER, such as FLUVIAL EROSION and GLACIER EROSION. In contrast, GEOLOGY includes those propositions as well as others, such as WIND EROSION, SHEET EROSION, ANTHROPIC EROSION, etc. GEOLOGY, on the other hand, also includes concepts that are not related to HYDROLOGY such as ATTRITION because there is no WATER involved.

However, WATER displays more relations in HYDROLOGY than in GEOLOGY. This is caused by the fact that WATER is a much more prototypical concept in HYDROLOGY. Therefore, its first hierarchical level shows more concepts. For example, in GEOLOGY, there are fewer WATER subtypes because the network only shows those that are related to the geological cycle (MAGMATIC WATER, METAMORPHIC WATER, etc.). In

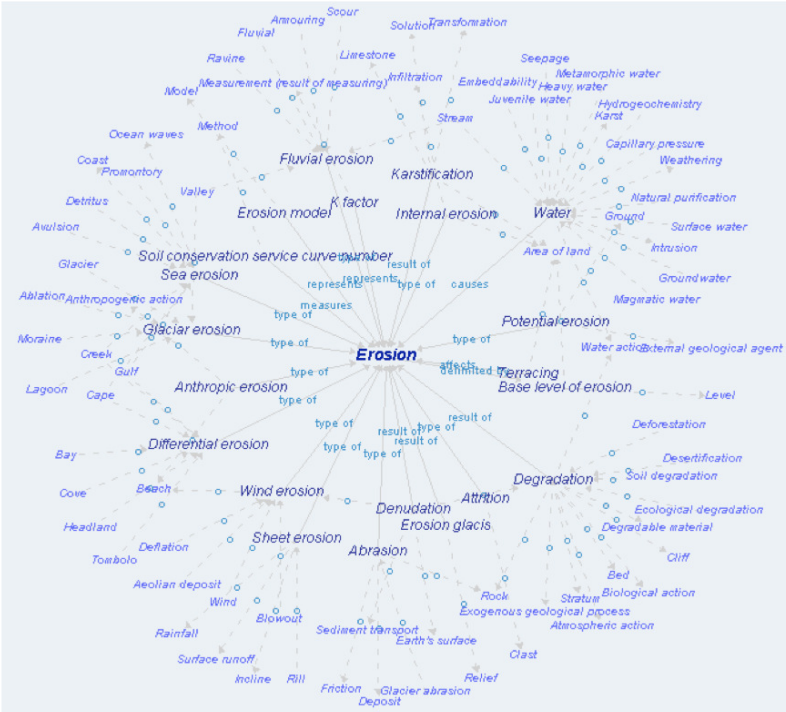


Fig. 7. EROSION in the GEOLOGY contextual domain

HYDROLOGY, there are more WATER subtypes related to the hydrological cycle itself (SURFACE WATER, GROUNDWATER, etc.). Even the shape of each network illustrates the prototypical effects of WATER or EROSION. In Figure 7, EROSION is displayed in a radial structure that reflects the fact that it is a central concept in GEOLOGY, whereas in Figure 8, the asymmetric shape of the network implies that, even more than EROSION, WATER is the prototypical concept of HYDROLOGY.

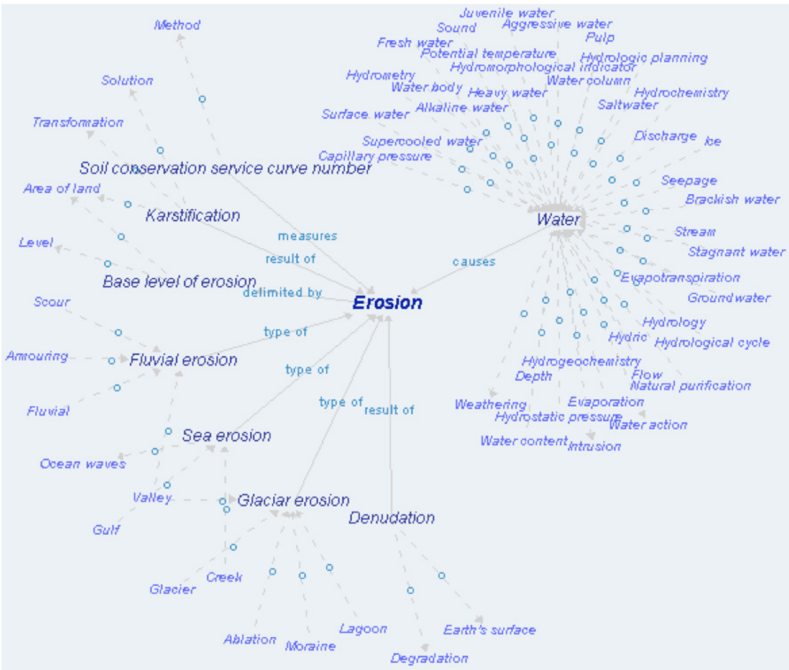


Fig. 8. EROSION in the HIDROLOGY contextual domain

C. Definitions and images

Definition construction follows a top-down and a bottom-up approach. This means that definitional elements are extracted from other resources' definitions combined with our corpus information. They are elaborated following the constraints imposed by the basic ontological classes and the inventory of conceptual relations. Similar concepts are grouped together in different templates according to category membership. These definitional templates are combined with images, which are selected from the web to further explain the relations expressed in the templates. In Figure 9, the natural geological process template is shown. It contains the four basic relations typical of any natural process: *is_a*, *has_agent*, *affects* and *has_result*. As this concept is at a high level in the ontology, the fillers for the conceptual relations are very general ones. *Has_result* does not even have any filler at all, as no constraints can be applied at this general level. Only the *agent* dimension is constrained to a geological entity. As this level of the hierarchy is very general and therefore rather abstract, the image chosen is general and abstract as well. The geological cycle describes how all natural geological processes, such as EROSION, the WATER CYCLE, ROCK FORMATION, etc. interact and are all interdependent.

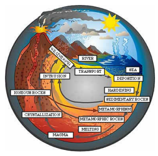
NATURAL GEOLOGICAL PROCESS		
[IS_A]	Natural process	
[HAS_AGENT]	Natural geological agent	
[AFFECTS]	(Entity, Process)	
[HAS_RESULT]	(Entity, Process)	

Fig. 9. NATURAL GEOLOGICAL PROCESS template

EROSION is the next level in the hierarchy and constrains the natural geological agent of the process to GRAVITY, WIND, WATER, ICE, and ANIMALS, with all their subtypes. The filler of the *affects* relation is the Earth's surface and all its subparts. The *result* dimension only includes several of the many results of the process. A new dimension is added at this level: *has_phases*. The images chosen at this level of abstraction combine all the agents of EROSION and another that shows all types of possible landscapes resulting from EROSION (Figure 10).

These images cannot give detailed information given that EROSION contains many subtypes, depending on the agent involved and the result obtained. For example, if we keep going down in the hierarchy, the level of specificity increases and the patients and results are closer to their real world referents. This is in consonance with Rosch's basic level and prototype theory [32]. According to prototype theory, basic level concepts belong to the first level of abstraction for which we can develop a concrete mental image. Although EROSION seems to be at the basic level (according to general language dictionaries, for example), when dealing with specialized knowledge, the basic level moves downwards in the hierarchy. This is why WATER EROSION can be better illustrated (Figure 11).

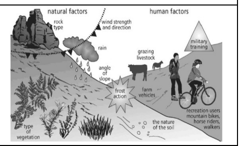
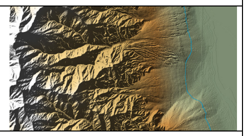
EROSION		
[IS_A]	Natural geological process of reduction	
[HAS_AGENT]	Gravity Water (river, stream, rain) Ice (glacier) Wind Animals	
[AFFECTS]	Earth's surface (beaches, mountains, soil...)	
[HAS_RESULT]	Landslide Rill Gully Sheet …	
[HAS_PHASES]	Weathering Transport Deposition	

Fig. 10. EROSION template

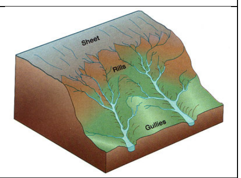
WATER EROSION		
[IS_A]	Erosion	
[HAS_AGENT]	Water (river, stream, rain, wave, current…)	
[AFFECTS]	Earth's surface (beaches, mountains, soil...)	
[HAS_RESULT]	Rill	
	Gully	
	Sheet	
	Cliff	
	Beach	
[HAS_PHASES]	Weathering	
	Transport	
	Deposition	

Fig. 11. WATER EROSION template

The template of WATER EROSION constrains the agent dimension further to WATER. The *patient* and *phases* dimensions are the same as the superordinate EROSION, but the *result* dimension shows clear examples at this level of description. Here three images have been added showing the three phases of WATER EROSION. Moreover, two images show the *result* dimension. Going even further down in the hierarchy, other dimensions become more specific. For example, COASTAL EROSION and SURFACE EROSION are types of WATER EROSION that constrain the *location* dimension.

The linguistic description of the concepts in EcoLexicon follows these templates insofar as type, quantity, and configuration of information are concerned. In this way, definitions show a uniform structure that complement the information encoded in conceptual networks, and directly refer to and evoke the underlying conceptual structure of the domain. These templates can be considered a conceptual grammar which thus ensures a high degree of systematisation.

IV. CONCLUSIONS

In this paper we have briefly explained the methodology applied in EcoLexicon for knowledge extraction and representation. Corpus analysis, combining direct term searches and knowledge pattern analysis, has fed the EcoLexicon knowledge base with reliable information. However, this information has to be represented coherently and systematically. EcoLexicon combines the advantages of a relational database, allowing for a quick deployment and feeding of the platform, and an ontology, enhancing user queries. The internal coherence at all levels of a dynamic knowledge representation shows that even complex domains can be represented in a user-friendly way. This methodology solves two challenges derived from multidimensionality: (1) it offers a qualitative criterion to represent specialized concepts according to recent research on situated cognition [15] both in dynamic networks and multimodal definitions; (2) it is a quantitative and efficient solution to the problem of information overload. Further steps in EcoLexicon will be the automatization of some of its extraction procedures, as well as the evaluation of the resource through usability tests.

ACKNOWLEDGMENT

This research has been funded by project FFI2008-06080-C03-01/FILO, from the Spanish Ministry of Science and Innovation.

REFERENCES

[1] I. Meyer and K. Mackintosh, "The corpus from a terminographer's viewpoint," *International Journal of Corpus Linguistics*, vol. 1, no. 2, pp. 257–285, 1996.

[2] L. Bowker and J. Pearson, *A practical guide to using corpora*. London / New York: Routledge, 2002, ch. Working with Specialized Language.

[3] A. Reimerink, M. García de Quesada, and S. Montero Martínez, "Contextual information in terminological bases: a multimodal approach," *Journal of pragmatics*, vol. 42, no. 7, pp. 1928–1950, 2010.

[4] E. Marsh and M. White, "A taxonomy of relationships between images and text," *Journal of Documentation*, vol. 59, no. 6, pp. 647–672, 2003.

[5] C. Barrière, "Knowledge-rich contexts discovery," in *Proceedings of the 17th Canadian Conference on Artificial Intelligence (AI'2004)*, Ontario, Canada, May 2004.

[6] C. Barrière and A. Agbago, "Terminoweb: A software environment for term study in rich contexts," in *Proceedings of the International Conference on Terminology, Standardisation and Technology Transfer (TSTT 2006)*, Beijing, China, 2006.

[7] P. Cimiano and S. Staab, "Learning concept hierarchies from text with a guided agglomerative clustering algorithm," in *Proceedings of the Workshop on Learning and Extending Lexical Ontologies with Machine Learning Methods*, Bonn, Germany, 2005.

[8] P. León Araúz and A. Reimerink, "Knowledge extraction and multidimensionality in the environmental domain," in *Proceedings of the TKE conference*, Dublin, Ireland, August 2010.

[9] E. Marshman and M. L'Homme, *Modern Approaches to Terminological Theories and Applications, Linguistic Insights*. Bern, Switzerland: Peter Lang, 2006, ch. Disambiguation of Lexical Markers of Cause and Effect, pp. 261–285.

[10] A. Auger and C. Barrière, "Pattern-based approaches to semantic relation extraction: A state-of-the-art," *Special Issue on Pattern-Based Approaches to Semantic Relation Extraction, Terminology*, vol. 14, no. 1, pp. 1–19, 2008.

[11] A. Condamines and J. Rebeyrolle, "Searching for and identifying conceptual relationships via a corpus-based approach to a terminological knowledge base (ctkb)," *Special Issue on Recent Advances in Computational Terminology, Terminology*, no. 127-148, 2001.

[12] M. Rogers, "Multidimensionality in concepts systems: a bilingual textual perspective," *Terminology*, vol. 10, no. 2, pp. 215–240, 2004.

[13] I. Meyer, L. Bowker, and K. Eck, "Cogniterm: An experiment in building a knowledge-based term bank," in *Proceedings of Euralex '92*, 1992, pp. 159–172.

[14] P. Faber, "Conceptual modelling in specialized knowledge resources," in *Proceedings of the XII International Conference Cognitive Modelling in Linguistics*, Dubrovnik, Croatia, September 2010.

[15] L. Barsalou, "Simulation, situated conceptualization and prediction," *Philosophical Transactions of the Royal Society of London: Biological Sciences*, vol. 364, pp. 1281–1289, 2009.

[16] J. Austin, *How to do things with words*. Oxford, UK: Clarendon, 1962.

[17] J. Searle, *Speech acts: An essay in the philosophy of language*. Cambridge, UK: Cambridge University Press, 1969.

[18] H. Gadamer, *Truth and Method*. New York, USA: Continuum, 1995.

[19] H. Grice, *Syntax and semantics: Speech acts*. New York, USA: Academic Press, 1975, vol. 3, ch. Logic and conversation, pp. 41–58.

[20] D. Sperber and D. Wilson, *Relevance: Communication and Cognition*, 2nd ed. Oxford, UK: Blackwell, 1995.

[21] E. Goffman, *Frame Analysis*. New York, USA: Harper and Row, 1974.

[22] H. Clark, *Using Language*. Cambridge, UK: Cambridge University Press, 1996.

[23] R. Guha, "Contexts: A formalization and some applications," Ph.D. dissertation, Stanford University, 1991.

[24] P. Brézillon, "Task-realization models in contextual graphs," in *Proceedings of Modeling and Using Context (CONTEXT-05)*, A. Dey, B. Kokinov, D. Leake, and R. Turner, Eds. Springer Verlag, 2005.

[25] D. Lenat, "Cyc: a large-scale investment in knowledge infrastructure," *Communications of the ACM*, vol. 38, no. 11, pp. 33–38, 1995.

[26] A. Dey, "Understanding and using context," *Personal Ubiquitous Computing*, vol. 5, no. 1, pp. 4–7, 2001.

[27] V. Evans, "Cognitive linguistics," in *Encyclopedia of pragmatics*, L. Cummings, Ed., 2009.

[28] J. Barrasa, "Modelo para la definición automática de correspondencias semánticas entre ontologías y modelos relacionales," Ph.D. dissertation, UPM, Madrid, 2007.

[29] M. Murphy, *Semantic relations and the lexicon: Antonym, synonymy, and other paradigms*. Cambridge, UK: Cambridge University Press, 2003.

[30] P. León Araúz and P. Faber, "Natural and contextual constraints for domain-specific relations," in *Proceedings of Semantic relations. Theory and Applications*, Valetta, Malta, May 2010.

[31] A. Reimerink, P. León Araúz, and P. Magaña, "Ecolexicon: an environmental tkb," in *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2010)*, Valetta, Malta, May 2010.

[32] E. Rosch, *Cognition and Categorization*. Hillsdale, NJ, USA: Erlbaum, 1978, ch. Principles of categorization, pp. 27–48.