Causality in the Specialized Domain of the Environment

Pilar León Araúz and Pamela Faber
Department of Translation and Interpreting, University of Granada
Buensuceso, 11 18002 Granada
E-mail: pleon@ugr.es, pfaber@ugr.es

Abstract

EcoLexicon is a multilingual terminological knowledge base (TKB) that represents environmental concepts and their relations in different formats. In this paper we show how some of the manual processes that we have developed for the extraction and representation of semantic relations can be partially automatized with the help of NLP applications such as NooJ. Focusing on the causal relation, we have designed various graph-based micro-grammars to match and annotate the corpus. This permits the extraction of causal propositions, and identifies the terms that primarily act as causes and effects in environmental contexts. Finally, these grammars can also be used to measure the prototypicality of causal propositions within four different environmental domains.

Keywords: knowledge extraction, causal relations, semantic prototypicality, environment

1. Introduction

EcoLexicon[1] is a multilingual terminological knowledge base (TKB) that represents environmental concepts and their relations in different formats (i.e. ontology, conceptual networks, controlled-language definitions, graphical resources and linguistic contexts, such as knowledeg-rich contexts and concordances). So far, it has 3,343 concepts and 17,413 terms in English, Spanish, German, French, Modern Greek, Russian and Dutch. In this paper we show how some of the manual processes that we have developed in the extraction and representation of semantic relations can be partially automatized with the help of NLP applications such as NooJ (Silberztein, 2003).

2. Semantic relations in EcoLexicon

In addition to hyponymic relations, our inventory of semantic relations also includes six types of meronymy as well as non-hierarchical relations, such as *affects*, *result_of*, *causes*, etc., which best represent the dynamism of the environmental domain (León Araúz and Faber, 2010). Up to the present, all conceptual propositions in EcoLexicon (more than 6,000) have been manually extracted from the corpus (5 million words) and represented in semantic networks. However, knowledge representation would be more objective and efficient if knowledge extraction techniques were more systematic and semi-automatized. Nevertheless, this requires a well-defined set of selection criteria, based on the manual identification of which types of information are useful, why they are useful and how they can be structured.

2.1 Extracting semantic relations from the corpus

According to many research studies, knowledge patterns (KPs) have long been considered one of the most reliable methods for the extraction of semantic relations (Condamines, 2002; Marshman et al., 2002; Barrière, 2004; Barrière and Abago, 2006; Cimiano and Staab, 2006). The term KP was coined by Meyer (2001) to refer to the lexico-syntactic patterns between the terms encoded in a proposition in real texts.

Since Hearst (1992) much as has been written about KPs. Nevertheless, despite their popularity, KPs have never been fully studied and exploited. As Bowker (2004) states, there are still major problems with regards to noise and silence, pattern variation, anaphora, domain and language dependency, etc. Moreover, not all relations have been analyzed in the same depth. Patterns conveying hyponymic relations are the most commonly studied since they play an important role in categorization and property inheritance (Barrière, 2004: 244). Nonetheless, even though non-hierarchical KPs have also been identified by many other authors, they have never been systematically implemented in research studies (Aussenac-Gilles, 2000: 181).

KPs have mostly been used to extract information from general language texts, but they have also been applied in certain specialized domains, such as Medicine (Rosario and Hearst, 2004; Vintar and Buitelaar, 2003; Embarek and Ferret, 2008; Khoo et al., 1999) or Biopharmaceutics (Marshman, 2002). However, to the best of our knowledge, there have been no KP-related studies on the environment.

All approaches seem to agree that the use of KPs for knowledge extraction involves a series of complementary steps. Nevertheless, the order of the steps differs, depending on research objectives (e.g. identification of term pairs, discovery of new KPs, searching for known KPs to discover new term pairs, etc.). In Terminology, Meyer (2001) suggests first identifying an initial set of KPs for each semantic relation. These patterns are then tested and additional patterns are identified. Restrictions are subsequently defined that can be applied to reduce noise and silence. As part of our study, all of this has first

[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
```

Figure 1: *Erosion* concordances

been done by manual corpus analysis.

For example, in Figure 1 we show the results of the first step in our approach. We search for specialized terms, such as *erosion*, collect the most meaningful concordances and classify them based on the relations expressed. KPs are then collected, such as those found in Figure 1: *associated with*, *agent for*, *can/may also cause*, *can be due to*, *one of the causes of*, *responsible for*, *lead to*, etc. The next step involves reusing these KPs to discover new term pairs, after which we reinitiate the process with seed terms to discover new KPs. This information is also displayed to users since those who are translators and/or technical writers might find it useful.

During the manual identification of KPs, we encountered certain problems related to the polysemic nature of certain KPs, which did not always convey the same semantic relations (i.e. *formed by*; León Araúz and Reimerink, 2010) or the problem of KPs associated with an incomplete proposition because of anaphora. Nevertheless, we also found that the correct identification of meaningful concordances depends on the semantic and syntactic structure of the text that precedes and follows any KP.

2.2 Representing semantic relations in conceptual networks

The semantic relations between concepts in EcoLexicon are activated depending on the *natural constraints* imposed by a concept's intrinsic nature and its relational power. The activation of relations also depends on the *contextual constraints* stemming from facet incompatibility, which is the result of multidimensionality (see León Araúz and Faber, 2010 for a more detailed explanation). Succinctly put, depending on the type of concepts in a conceptual proposition, only a certain set of relations may apply. For instance, a PHYSICAL ENTITY can only be the *result of* a PROCESS, but not of another ENTITY, and only if the PHYSICAL ENTITY plays the role of PATIENT and not that of AGENT. Furthermore, concepts in the environmental domain have multiple dimensions that are often incompatible because they are context-dependent. For example, despite that WATER is included in propositions such as <CONCRETE *made_of* WATER> and <WATER *causes* EROSION>, these propositions should evidently not be included in the same semantic network. Thus, even though a concept may be part of multiple propositions, only one set of these propositions should be activated in a certain context. Therefore, we have divided the environmental domain into field-specific contextual subdomains, such as HYDROLOGY, GEOLOGY, OCEANOGRAPHY, SOIL SCIENCES, ATMOSPHERIC SCIENCES, etc.

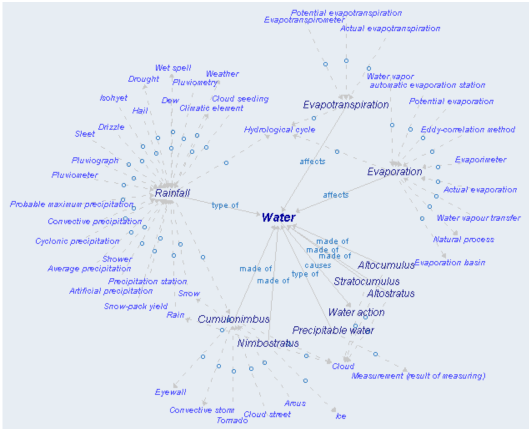

Figure 2: WATER in ATMOSPHERIC SCIENCES

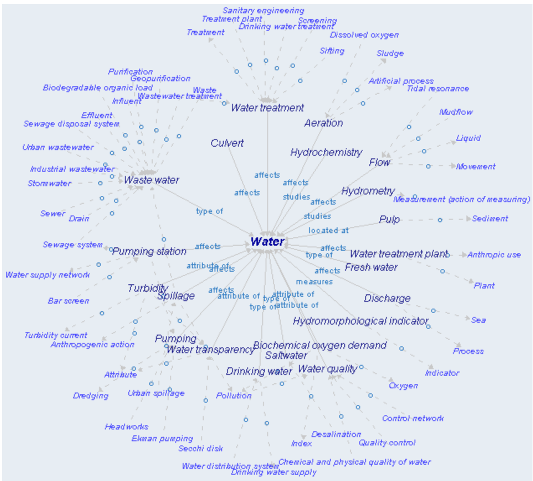

Figure 3: WATER in WATER TREATMENT

Each of these domains provides a frame for conceptual recontextualization. A comprehensive list of all contextual domains can be found in León Araúz and San Martín (in press). Figures 2 and 3, show the different recontextualizations of the semantic networks for WATER in the subdomains of ATMOSPHERIC SCIENCES and WATER TREATMENT. As can be observed, prototypical propositions for WATER (e.g. <WATER *causes* EROSION>), which would generally be activated in a context-free search, do not appear in either network. Instead, it is the context that modulates the prototypicality of propositions. The recontextualization of concepts thus involves decisions about which propositions should be activated within each domain. In EcoLexicon, so far, this has been done manually and intuitively, based on corpus searches and analysis. This time-consuming process has been extremely worthwhile in that it has provided us with the knowledge needed to formalize the structure of KPs for automatic corpus searches and determine the prototypicality of conceptual propositions. Accordingly, the corpus texts are currently being classified in contextual domains.

3. Causal relation

Broadly speaking, causality is the relation between a cause and its effect. Of the non-hierarchical relations in EcoLexicon, causality is one of the most important. Obviously, the environment is conceived as a process where causes and effects are at the core of any event. Not surprisingly, causal relations are also crucial for other difficult tasks in NLP, such as question answering (Girju, 2003).

The extraction and representation of causality have been studied from a wide range of disciplines and perspectives. These include: (i) Cognitive Linguistics, as reflected in Talmy's Force dynamics (2000), (ii) Artificial Intelligence, in different NLP applications; (iii) Philosophy and Psychology (White, 1990), etc. All these studies affirm that there are many ways to express causation since it can be expressed in passive, active, subject-object, nominal or verbal propositions. Moreover, causes and effects have very diverse syntactic representations. More specifically, causation is not only expressed by constructions such as *due to* or *because of*, but also by causative nouns (*cause* or *consequence*) and verbs. Although there are many causative verbs (e.g. *cause, generate, lead, produce*, etc.), their syntactic behavior can vary. As a result, one single grammar is not sufficient to formalize their complementation structures.

This has led researchers to classify causal relations in different facets. For example, Blanco et al. (2008) classified these relations in *influence*, *condition*, *consequence* and *reason.* In contrast, the classification in Nastase (2003) is based on *cause*, *effect*, *purpose*, *entailment*, *enablement*, *detraction* and *prevention*. For Khoo et al. (2002), causation is also complex and multifaceted. They use templates for each causal category involved in the relation (*cause*, *effect*, *subjects involved*, *condition*, *modality*) and provide a classification of explicit patterns, such as adverbial (*so*, *hence*, *therefore*), prepositional (*because of*) and subordination (*as*, *since*) causal links, clause integrated links (*that's why*, *the result was*), causative verbs (*break*, *kill*), resultative constructions, conditionals and causative adverbs, adjectives, and prepositions.

Girju (2003) also states that causative constructions may be explicit or implicit. Her work focuses on explicit but ambiguous verbal causation patterns. She provides a list of 60 causative verbs and classifies them into simple causatives (*cause*, *lead to*, *bring about*, *generate*, *make*, *force*, *allow*); resultative causatives (*kill*, *melt*, *dry*, etc.) and instrumental causatives (*poison*, *hang*, *punch*, *clean*) This identification of causes and effects is derived from the transitivity of WordNet verbs.

3.1 Causal grammars for EcoLexicon

In EcoLexicon, we have developed a series of KP-based micro-grammars with the help of NooJ, a development environment used to construct large-coverage descriptions of natural languages and apply them to large corpora (Silberztein, 2003). The main advantages of NooJ grammars over manual searches based on regular expressions are recursivity as well as the possibility of annotating the corpus with different tags that can be reused in batch processing tasks. We used NooJ parser to identify causal syntactic structures in a 1,200,000 word corpus. The corpus was manually classified into four contextual domains, each of approximately 300,000 words: ATMOSPHERIC SCIENCES, COASTAL ENGINEERING, OCEANOGRAPHY, and SOIL SCIENCES.

As previously mentioned, causation can be expressed in many different ways. Moreover, the semantic roles and features of the elements in a causal proposition, as well as their syntactic behaviour, can change, depending on the structure and order. For instance, in the proposition <X *causes* Y>, X is the CAUSE and Y the EFFECT, whereas in <X *is caused by* Y>, X is the EFFECT and Y is the CAUSE. This is why we have developed an array of micro-grammars for the causal realizations rather than only one. Apart from searching for the causal KP, we also wanted to

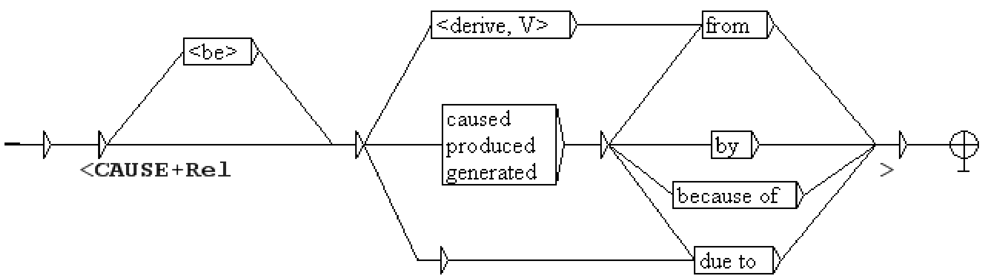

Figure 4: Core grammar of the causal relation

```
             Flash flooding due to extremely heavy rains
     Most storm-related damage was caused by wind, wind-blown rain and tornadoes
Sustained winds of tropical storm force produced by Rita
The most familiar sea level changes are produced by astronomical tides
           Earthquakes are shock waves caused by abrupt movements of the earth's crust
      Local wind patterns (sometimes caused by structures and urban development)
                Sediment fluxes generated by incident waves
            Currents are usually due to tides| and river flows
              Internal waves are generated by wind energy
             Tsunami can also be caused by landslides
           Vapor transfer in soil due to air movement
```

Figure 5: Causal propositions matching <CAUSE+Rel> grammar

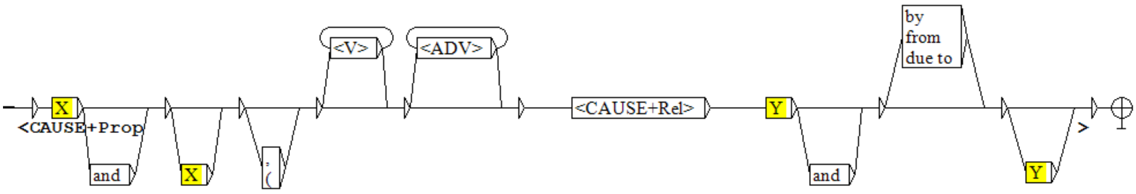

Figure 6: Grammar for causal propositions

extract the elements involved, whether they are causes or effects and regardless of whether they are already stored in our TKB or not.

Thus, when the corpus is matched with the graph-based micro-grammar structures, it is possible to annotate the corpus and extract the entire causal proposition as well as the environmental terms acting as causes and effects.

So far, we have developed five micro-grammars for the following constructions: <X *causes* Y>, <X *caused by* Y>, <X *is the cause of* Y>, <*the cause of* X *is* Y>, and <X *causes* Y *to* Z>. Of course, they are not limited to the verb or noun cause, but also include other causative verbs and nouns. However, we did not include all 60 verbs found in Girju (2003) because each requires a different treatment and will be dealt with separately in the future. Moreover, some of these verbs correspond to other domain-specific relations in EcoLexicon.

This first approach to causation only focuses on the construction <X *caused by* Y>. Despite the many other ways to approach causation in the corpus, this pilot study yielded surprisingly rich results.

For efficiency reasons, the first step was to elaborate a grammar that formalized the most basic sense of causation (Figure 4). This grammar extracts causal links by following different paths. As shown in Figure 4, causation can be expressed by: the participle of *cause*, *produce* and *generate* (optionally preceded by *to be* in any of its inflected forms), and followed by one of the four prepositional constructions. However, it can also be expressed by *derive*, in any of its inflected forms followed by the preposition *from*, or by the adjectival phrase *due to*.

We located all of the occurrences matching this grammar and annotated them with the tag <CAUSE+Rel>. From the entire corpus, we extracted 960 causal occurrences, and thus found meaningful causal sentences such as those in Figure 5.

However, not all of them were found to be valid causal propositions, since sometimes the causal expression did not link two specialized terms, such as those cases where X is expressed as *this*, *that*, etc. Thus, we designed a more complex micro-grammar that reused the annotation <CAUSE+Rel> as the link between X (EFFECT) and Y (CAUSE) (Figure 6).

This grammar contemplates the possibility of having more than one effect and/or cause in the same causal proposition (i.e. ***chemical solution** and **mechanical abrasion** caused by some organisms* or *dune erosion produced by **storm waves** and **water level***). This is why X

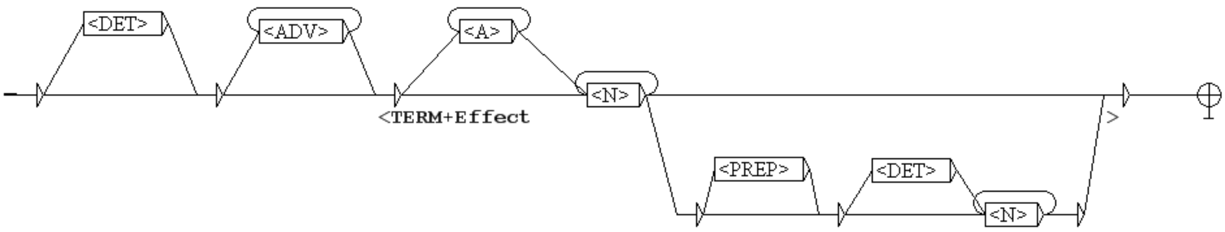

Figure 7: Grammar for <TERM+Effect>

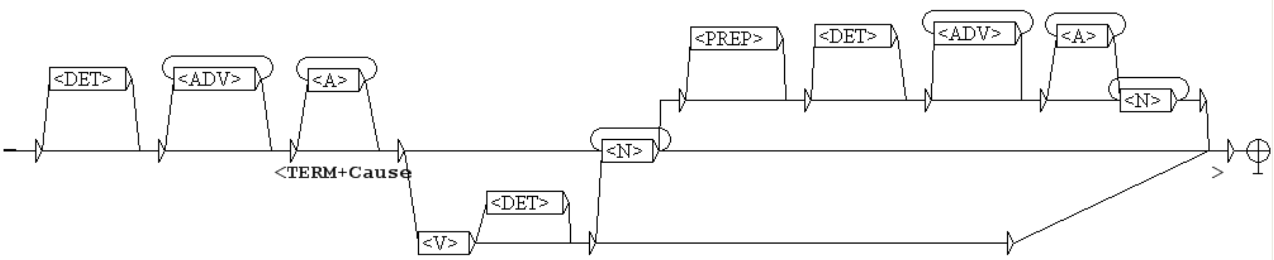

Figure 8: Grammar for <TERM+Cause>

and Y appear twice joined by the conjunction *and*, along with certain prepositions already used in <CAUSE+Rel>. It also includes punctuation marks, such as a comma and a bracket, since they often appear between effects and causal links, as in *local wind patterns (sometimes caused by structures and urban development)*. Moreover, it also accounts for the occurrence of one or more verbs (<V>*) and/or one or more adverbs (<ADV>*) between the effects and <CAUSE+Rel>.

As a result, this grammar is able to identify sentences like *continental glaciers* ***possibly*** *caused by a warming climate*, *coastal erosion* ***may be mainly*** *produced by wave attack*, or *tsunami* ***can also be*** *caused by landslides*. Note that in *can also be*, can also corresponds to <V><ADV> and *be* is matched through the <CAUSE+Rel> grammar. Once identified, they are annotated as <CAUSE+Prop>.

The elements highlighted are two different sub-graphs describing the possible syntactic structure of both X (EFFECT) and Y (CAUSE) as specialized terms (Figures 7 and 8).

As is well known, specialized knowledge units are very often multi-word terms composed of two nouns (*beach erosion*), a combination of adjectives and nouns (*detached breakwater*) or prepositional sentences (*the gravitation of the moon*). Moreover, when they are inserted in a text, they can also be modified by adverbs or adjectives that, strictly speaking, are not part of the terminological phraseme. This is why they are not included in the annotations <TERM+Effect> and <TERM+Cause>, but do appear in the grammar in order to identify the whole proposition.

These structures are capable of identifying various causes and effects as multi-word terms. In ***delta land loss*** *caused by* ***rising sea level***, the effect is identified by following the path <N>* and the cause through <A><N>*. In ***cliff retreat***, *caused by unusually* ***severe winter storms***, the effect and the cause are recovered through the paths <N>* and <A><N>*, respectively. This is possible despite the presence of an adverb (*unusually*) that matches the grammar but is not recovered as part of the term. More complex sentences can also be found, such as ***rates of subsidence*** *caused by* ***compaction of newly deposited sediment***, where the effect now follows the path <N><PREP><N> and the cause <N><PREP><A>*<N>. Furthermore, causes are defined by means of an additional path that includes a verbal proposition in order to identify phrases like ***environmental damages*** *caused by* ***dredging the river*** (<V><DET><N>).

3.2 Causal propositions in EcoLexicon

A search for all <CAUSE+Prop> annotated sentences gave 347 propositions, which were filtered out from the initial 960 occurrences through the formal description of effects and causes as specialized terms (<TERM+Effect> and <TERM+Cause>).

These three tags thus allow the extraction of all meaningful causal propositions for each concept in the corpus and automatically display them to users. Even more interestingly, it is also possible to extract all effect-cause pairs, as well as to measure the prototypicality of certain causal propositions, in each domain.

For instance, Table 1 gives a simplified classification of the most common causes and effects of all four contextual domains.

	CAUSE	EFFECT
ATMOSPHERIC SCIENCES	Tropical cyclones, swells, hurricane, wind, storm, storm surge, heavy rains, floods, typhoon, thunderstorms	Floods, storm surge, waves, tropical storm force winds, rise in ocean level, swells, adiabatic changes
COSTAL ENGINEERING	Glaciers, tides, gravitation, tropical storms, wind, groundwater withdrawal, tectonic movements, dams, rising sea level, changes in wave energy, tidal currents, offshore transport, recession of the beach, seawall, waves, scour, wave action, wave attack, longshore transport, erosion	Fall of water levels, wind, water level changes, eustatic rise in sea level, tsunamis, salt weathering, ocean waves, changes in sea level , antidunes, waves, currents, longshore sand transport, erosion
OCEANOGRAPHY	Tectonic forces, seawater, wind energy, wind, landslides, tidal currents, gravitation, wave swell, faulting	Storm surge, tsunami, waves, tides, wind, estuaries
SOIL SCIENCES	Electrical polarity of the water molecule, vegetation canopy, pressure gradient, gravitation, downward seepage, vapor pressure, osmosis, wind	Rise of the water table, sand columns, intermolecultar forces in liquid water, transpiration, wind

Table 1: CAUSES and EFFECTS in four contextual domains

As can be observed in Table 1, the four domains share many of the same causes and effects detected by the <X *caused by* Y> proposition. Moreover, the multidimensionality of the environmental domain is reflected in certain concepts that can act both as cause and effect even within the same domain (WIND, TIDE, CURRENT, FLOOD, etc.). Interestingly enough, WIND can be cause and effect in all four domains. However, its prototypical role changes across them. Figure 9 and 10 show the standard score of WIND as an effect and as a cause in each of the corpora. The standard score, retrieved thanks to NooJ's statistical module, shows the standard deviations of the occurrences that are above or below the mean. This is similar to the concept of prototypicality used to recontextualize semantic networks in EcoLexicon. Thus, based on Figures 9 and 10, the propositions in which WIND is an effect mostly appear in ATMOSPHERIC SCIENCES texts, whereas those in which WIND is a cause primarily occur in ATMOSPHERIC SCIENCES and OCEANOGRAPHY texts. Therefore, the concept is recontextualized in semantic networks accordingly.

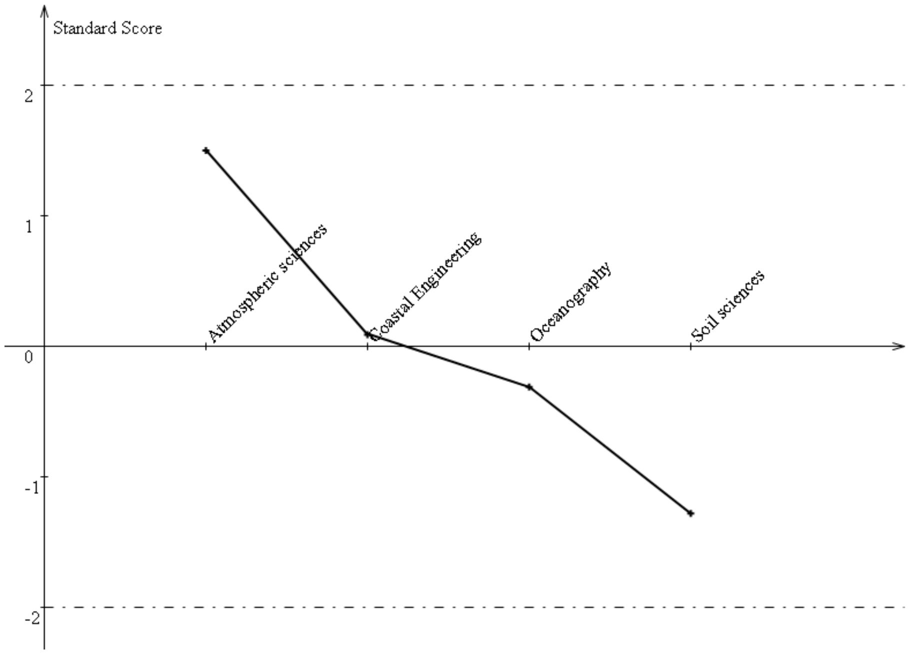

Figure 9: Prototypicality of WIND as an effect

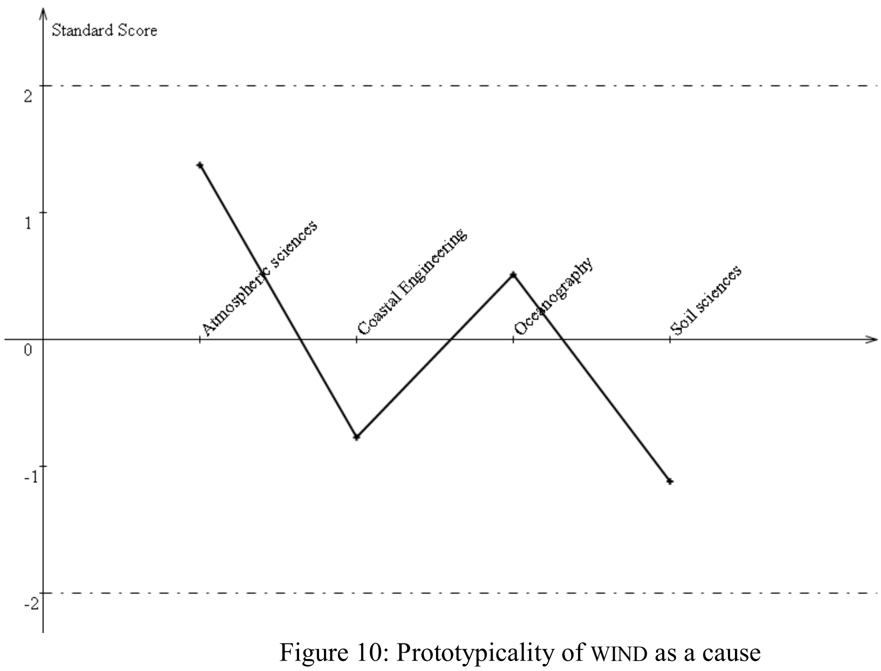

Figure 10: Prototypicality of WIND as a cause

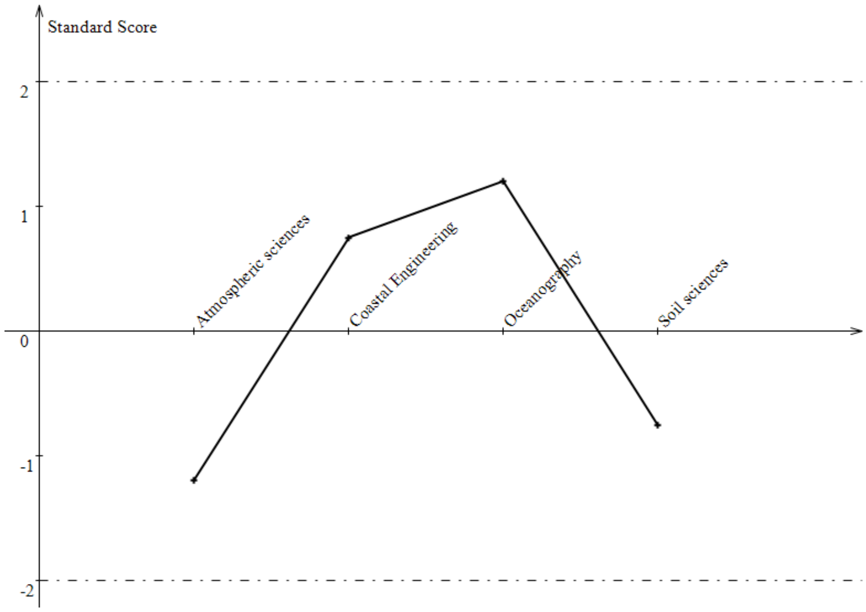

Figure 11: Prototypicality of <STORM SURGE *caused by* WIND>

However, this does not mean that each causal proposition in which WIND is a cause only occurs in ATMOSPHERIC SCIENCES and OCEANOGRAHY. Regarding the concrete WIND-related proposition <STORM SURGE *caused by* WIND>, the results show that it should not only be included in the recontextualized semantic network of OCEANOGRAPHY (and not in that of ATMOSPHERIC SCIENCES), but also in that of COASTAL ENGINEERING. This is why contextual constraints are not applied to individual concepts nor to semantic relations, but to complete and concrete conceptual propositions.

4. Conclusion and future work

In this paper we have shown how KP-based corpus analysis can be enhanced through the formal description of the syntactic structures of KPs and the help of NLP applications. Although manual work is still necessary to discover new patterns that reflect semantic relations in real texts, the knowledge thus acquired can be reused in automatic procedures. Otherwise, knowledge representation in lexical resources would be overly dependent on intuition.

In the near future, these patterns will be applied to the whole corpus in EcoLexicon. Once the corpus is classified in contextual domains, it will be proccessed using these causal micro-grammars, and new ones will be designed for other semantic relations in our TKB. This is a cyclic process since the application of relational micro-grammars to the most prototypical term pairs in each domain will also validate the categorization of the corpus.

A further step will be to identify possible cases of noise and silence and finally measure the precision and recall of the results with a gold standard. The disambiguation of polysemic structures also remains a challenge. Apart from polysemic KPs, specialized terms may also yield confusing results. For instance, when searching for the prototypicality of WAVE-related propositions, the SOIL SCIENCES domain shows false positives. The reason for this is that *wave* is a very common term in this domain, but only in its physics sense and not in its sea-related sense. Our intuition is that these problems could be solved by adding a semantic component to the grammars. As Girju and Moldovan (2002) state, semantic features are essential to constrain which entities will be efficiently linked through causation. Although these authors use a set of features from WordNet for this purpose, we plan to implement a NooJ-based dictionary containing all of the terms in EcoLexicon as well as the semantic features that define our concepts and categories.

5. Acknowledgements

This research has been carried out within the framework of the project RECORD: Representación del Conocimiento en Redes Dinámicas [Knowledge Representation in Dynamic Networks, FFI2011-22397], funded by the Spanish Ministry for Science and Innovation.

6. References

Aussenac-Gilles, N.; Séguela, P. (2000). Les relations sémantiques : du linguistique au formel. *Cahiers de Grammaire* 25, *Sémantique et Corpus*, 175-198.

Barrière, C.; Agbago, A. (2006). TerminoWeb: a software environment for term study in rich contexts, *International Conference on Terminology, Standardization and Technology Transfer,* Beijing, 103-113.

Barrière, C. (2004). Building a concept hierarchy from corpus analysis. *Terminology* 10: 2, 241-263.

Blanco, E.; Castell, N.; Moldovan, D. (2008). Causal Relation Extraction. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco.

Bowker, L. (2004). Lexical knowledge Patterns, Semantic Relations, and Language Varieties: Exploring the Possibilities for Refining Information Retrieval in an International Context. *Cataloging and Classification Quarterly*, 37(1): 153 - 171

Cimiano, P., Staab, S. (2005) Learning Concept Hierarchies from Text with a Guided Agglomerative Clustering Algorithm, *Workshop on Learning and Extending Lexical Ontologies with Machine Learning Methods*, Bonn.

Condamines, A. (2002). Corpus analysis and conceptual relation patterns. *Terminology* 8:1, 141-162.

Embarek, M.; Ferret, O. (2008). Learning Patterns for Building Resources about Semantic Relations in the Medical Domain. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 08).* Marrakech, Morocco: ELRA.

Girju, R. (2003). Automatic Detection of Causal Relations for Question Answering. In the proceedings of the *41st Annual Meeting of the Association for Computational Linguistics (ACL 2003), Workshop on "Multilingual Summarization and Question Answering - Machine Learning and Beyond"*.

Girju, R.; Moldovan, D. (2002). Text Mining for Causal Relations. In *Proceedings of the International Florida Artificial Intelligence Research Society (FLAIRS 2002)*, Pensacola, Florida.

Hearst, M.A. (1992). Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING-1992).*

Khoo, C.; Chan, S.; Niu, Y.; Ang, A. (1999). A method for extracting causal knowledge from textual databases. *Singapore Journal of Library and Information Management*, 28, 48-63.

Khoo, C.; Chan, S.; Niu, Y. (2002). The many facets of the cause-effect relation. In R. Green, C.A. Bean and S.H. Myaeng (eds.), *The semantics of relationships: an interdisciplinary perspective*. 51-70.

León Araúz P.; Faber, P. (2010). Natural and contextual constraints for domain-specific relations. In *Proceedings of the Workshop Semantic Relations. Theory and Applications*, ed. Verginica Barbu Mititelu, Viktor Pekar, and Eduard Barbu, 12-17. Valletta.

León Araúz, P; Reimerink, A. (2010). Knowledge extraction and multidimensionality in the environmental domain. In *Proceedings of the Terminology and Knowledge Engineering (TKE) Conference 2010*. Dublin: Dublin City University.

León Araúz, P.; San Martín, A. (in press). Distinguishing polysemy from contextual variation in terminographic definitions. In *Proceedings of the 10th International Conference of AELFE*, Valencia: AELFE.

Marshman, E.; Morgan, T.; Meyer I. (2002). French patterns for expressing concept relations. *Terminology* 8:1, 1-29.

Marshman, E. (2002). The cause relation in Biopharmaceutical Texts: Some English Knowledge Patterns. In *Proceedings of Terminology and Knowledge Engineering, TKE 2002*, 89-94.

Meyer, I. (2001). Extracting knowledge-rich contexts for terminography: A conceptual and methodological framework. In Bourigault, D., Jacquemin, C. and L'Homme, M.C. (eds.), *Recent Advances in Computational Terminology*. Amsterdam: John Benjamins. 279-302.

Nastase, V.(2003). *Semantic Relations Across Syntactic Levels*. Ph.D. Dissertation, University of Ottawa.

Rosario, B. and Hearst, M. (2004). Classifying Semantic Relations in Bioscience Text. In *Proceedings of ACL 2004*.

Silberztein, M. (2003). *NooJ Manual*. Available at: http://www.nooj4nlp.net/NooJManual.pdf

Talmy, L. (2000). *Toward a cognitive semantics*. MIT Press

Vintar, S.; Buitelaar, P. (2003). Semantic relations in concept-based cross-language medical information retrieval. In *ECML/PKDD Workshop on Adaptive Text Extraction and Mining (ATEM)*, Germany.

White, P. (1990). Ideas about causation in philosophy and psychology. *Psychological Bulletin*, 108 (1): 3-18