SEMANTIC RELATIONS, DYNAMICITY, AND TERMINOLOGICAL KNOWLEDGE BASES

Pamela Faber\textsuperscript{a}, Pilar León\textsuperscript{b}, Juan Antonio Prieto\textsuperscript{c}

Abstract The linguistic and conceptual shift in Terminology has led to a more discourse-centered approach with a focus on how terms are used in texts (Temmerman and Kerremans, 2003). This shift has affected the construction of terminological knowledge bases, which have an underlying network of semantic relations. Such a network can be derived from corpus analysis and the extraction of terminological units and semantic relations from knowledge-rich contexts (Meyer, 2001). Until recently, however, semantic relations in termbases were mainly restricted to generic-specific and part-whole relations. This was conducive to static configurations, which are at odds with the representation of dynamic action in domain models (Barrière, 2001: 137). Terminological knowledge bases can acquire greater coherence and dynamicity when: (1) a frame-based structure is used as the top level representation for all concepts; (2) a wider range of conceptual relations are contemplated, some of which may be domain-specific. This paper describes the semantic relations used in the EcoLexicon terminological knowledge base on Environmental Engineering and their representation. This choice of relations is derived from both corpus and definitional analysis, and is also reflected in the graphic resources for each concept. This semantic network can be accessed in the form of a ThinkMap representation.

\textsuperscript{a} Departamento de Traducción e Interpretación, Facultad de Traducción e Interpretación, Universidad de Granada, 18071 Granada, Spain. E-mail: pfaber@ugr.es
\textsuperscript{b} E-mail: pleon@ugr.es
\textsuperscript{c} E-mail:japrieto@ugr.es
The linguistic and cognitive shift in Terminology has led to a more discourse-centred approach with a focus on how terms are actually used in texts (Temmerman and Kerremans, 2003). This shift has also affected the construction of terminological knowledge bases, some of which are based on an underlying network of semantic relations, and make an effort to encode information in the form of knowledge. This is directly related to the construction of ontologies (in the artificial intelligence sense of the term).

2. Ontologies and Terminology
2.1. Ontologies
According to the International Digital Object Identifier (DOI Foundation 2005), an ontology can be defined as “an explicit formal specification of how to represent the entities that are assumed to exist in some area of interest and the relationships that hold among them”.

The purpose of a domain ontology is to eliminate conceptual and terminological confusion. It accomplishes this by specifying a set of generic concepts that characterize the domain as well as their definitions and interrelationships. It is now widely acknowledged that constructing a domain model is crucial to the development of knowledge-based systems. Since ontologies are of interest to linguists, terminologists, computer engineers, and philosophers, they can be designed and built from many different perspectives.

The most important task in creating any ontology is to establish the concepts in the domain, and organize them in a coherent design or upper ontology. In environmental engineering, there are both physical and abstract concepts. Physical concepts are those occupying space and time. They are natural entities, geographic accidents, water bodies,
constructions, and the natural and artificial processes that can affect them in some way. Abstract concepts include theories, equations, and units for measuring physical entities and processes.

One of the ways to extract knowledge for an ontology is through corpus analysis. Both terminological units and semantic relations can be extracted from knowledge-rich contexts (Meyer, 2001). Gillam, Tariq and Ahmad (2005) propose obtaining a conceptual system (ontology) through a systematic examination of texts in the specialist domain. They affirm that in this type of text-based approach, texts may signal changes in concepts that require modifications to the conceptual system. This directly relates ontologies with terminology and terminographic work.

2.2. Terminology

In the same way that ontologies can be constructed from different perspectives, terminology is also a many-splendored thing with approaches that range from traditional approaches (Wüst1968) to communicative approaches (Cabrè, 1999, 2000; Gaudin, 2003) to cognitive approaches (Temmerman, 2000, 2001; Faber et al, 2005, 2007, 2007).

More recently, sociocognitive terminology (Temmerman, 2000) has begun to focus on ontologies as a more viable way of implementing conceptual representations. This combination of terminology and ontology is called termontography, a hybrid term, which is a combination of terminology, ontology, and terminography. Its objective is to link ontologies with multilingual terminological information, and to incorporate ontologies into terminological resources. Temmerman and Kerremans (2003) describe termontography as a multidisciplinary approach in which theories and methods for multilingual terminological analysis (Temmerman, 2000) are combined with methods and guidelines for ontological analysis (Fernandez et al, 1997; Sure and Studer, 2003).

Termontography, as outlined by Temmerman, seems to owe a great deal to the work done by Ingrid Meyer (Meyer et al, 1992; Meyer and McHaffie, 1994; Meyer, Eck and Skuce, 1997; Bowker and L’Homme,
2004) who was one of the first terminologists to perceive that term bases would be even more useful if their organization bore some resemblance to the way concepts are represented in the mind. When term bases become terminological knowledge bases, as conceived by Meyer, they enhance data because the concepts and designations are linked to each other by meaningful relationships. Although the traditional generic-specific and part-whole relationships are contemplated, there is a greater emphasis on other types of relationships that enrich the resulting knowledge structure, such as cause-effect, object-function, etc. (Bowker and L’Homme, 2004).

Frame-based terminology (Faber et al, 2005; Faber et al, 2006; Faber et al, 2007) is another very recent cognitive approach to terminology, which shares many of the same premises as the communicative theory of terminology and sociocognitive terminology. For example, it also maintains that trying to find a distinction between terms and words is no longer fruitful or even viable, and that the best way to study specialized knowledge units is by studying their behavior in texts. Because the general function of specialized language texts is the transmission of knowledge, such texts tend to conform to templates, and are also characterized by a greater repetition than usual of terms, phrases, sentences, and even full paragraphs. Scientific and technical texts are usually terminology-rich because of the quantity of specialized language units in them, and they are also distinctive insofar as the syntactic constructions used.

Specialized language units are mostly represented by compound nominal forms that are used within a scientific or technical field, and have meanings specific of this field as well as a syntactic valence or combinatory value. The concentration of such units in these texts points to the activation of sectors of domain-specific knowledge. As a result, understanding a terminology-rich text requires knowledge of the domain, the concepts within it, the propositional relations within the text as well as the conceptual relations between concepts within the domain.
As its name implies, frame-based terminology uses a modified version of Fillmore’s Frames (Fillmore, 1976, 1982, 1985; Fillmore and Atkins, 1992) coupled with premises from Cognitive Linguistics to structure specialized domains and create non-language-specific representations. Such configurations are the conceptual meaning underlying specialized texts in different languages.

Frames also fall within cognitive linguistic approaches, and are a type of cognitive structuring device based on experience that provide the background knowledge and motivation for the existence of words in a language as well as the way those words are used in discourse. Boas (2005) points out that one of the problems in the creation of multilingual lexical databases is the development of an architecture capable of handling a wide range of linguistic issues such as polysemy, valence information, lexicalization patterns, and translation equivalents. This is also an issue for term bases.

Frames have the advantage of making explicit both the potential semantic and syntactic behavior of specialized language units. This necessarily includes a description of conceptual relations as well as a term’s combinatorial potential. Frame Semantics (Fillmore, 1976, 1982, 1985; Fillmore and Atkins, 1992) and its practical application, the FrameNet Project (Fillmore and Atkins, 1998; Fillmore et al, 2003; Ruppenhofer et al, 2006), assert that in order to truly understand the meanings of words in a language, one must first have knowledge of the semantic frames or conceptual structures that underlie their usage. Evidently, the same can be said for specialized language units.

Frame-based terminology focuses on: (1) conceptual organization; (2) the multidimensional nature of terminological units; (3) the extraction of semantic and syntactic information through the use of multilingual corpora. In frame-based terminology, conceptual networks are derived from an underlying domain event, which generates templates for the actions and processes that take place in the specialized field as well as the entities that participate in them.
Our methodology can be used to extract the conceptual system of the domain by means of an integrated top-down and bottom-up approach. The bottom-up approach consists of extracting information from a corpus of texts in various languages, specifically related to the domain. Our top-down approach includes the information provided by specialized dictionaries and other reference material, complemented by the help of experts in the field.

In a parallel way, we specify the underlying conceptual framework of a knowledge-domain event (Faber and Jiménez, 2002; Faber et al, 2006). The most generic or base-level categories of a domain are configured in a prototypical domain event or action-environment interface (Barsalou, 2003). This provides a template applicable to all levels of information structuring. A structure is thus obtained which facilitates and enhances knowledge acquisition since the information in term entries is internally as well as externally coherent (Faber et al, 2007).

For a logical formalism, we plan to use OWL-DL (description logic). As is well known, OWL comes in three varieties: OWL-Lite, OWL-DL, and OWL-Full. OWL-Lite is too basic for our purposes and is not adequate for a knowledge-based system for environmental engineering. In contrast, OWL-Full comes with no computational guarantees, and is not practical for available reasoning systems.

3. Semantic Relations
Evidently, the first step in any knowledge base is to establish the entities in the domain, more specifically, those belonging to the upper levels and link them with conceptual relations or roles. Until recently, however, conceptual relations in term bases, if they existed at all, were mainly restricted to generic-specific and part-whole relations. This was conducive to static configurations, which are at odds with the need to represent dynamic action in domain models (Barrière, 2001: 137). Rogers (2004: 218-219) writes:

“Text-book versions of concept systems tend to focus on the representation of abstract genus-species relations rather than the vast
array of ontological relations which are on the whole — with the exception of part-whole relations—poorly understood and documented.”

However, non-taxonomic relations between concepts are a major building block in common ontology definitions. The determination of non-taxonomic conceptual relationships is not well-researched, and there is no consensus regarding what type of conceptual relationships should be modeled in a particular ontology. In our experience, semantic relations largely depend on the type of entity being described, its nature, and relational power. The top-level identity of the concept influences to a great extent its core inventory of relations.

It is our assertion that terminological knowledge bases can acquire greater coherence and dynamicity when a wider range of conceptual relations are contemplated than the traditional generic-specific and part-whole relations. Some of these relations may be domain-specific.

4. The Ecolexicon Environmental Engineering Knowledge Base
4.1. Conceptual structure
The EcoLexicon knowledge base has a total of 3,147 concepts and 10,541 terms in English, Spanish, and German. Our top-level set of concepts is presently organized in a frame-based event, divided into three macro-templates.
Macro-templates are based on the semantic roles of the components of a prototypical environmental event (AGENT-PROCESS-PATIENT/RESULT). They are the result of a frame-based conceptualization, which reflects the dynamic nature of the environment. The primary conceptual relations that link semantic roles and macrotemplates are non-hierarchical relations such as AFFECTS, CAUSES, RESULTS, which help to resolve the limitations of hierarchical and static taxonomies. In our knowledge base, each concept type is related to other concepts by a set of conceptual relations, some of which are domain-specific. Concepts and relations have the following combinatorial potential:
This combinatorial potential is based on the following criteria:

- **ISA**: this generic-specific relation reflects hierarchical inheritance in the conceptual network of the domain. All entities and events are categorized as instances of a particular class. Classes can, in turn, become instances of superordinate classes, and lead to the most generic categories that are directly linked to the semantic roles of the event structure. Thus, any concept can be linked to its immediate superordinate concept (or several concepts in cases of multidimensionality). For example, sheet pile groyne (*instance*) ISA groyne (*class, instance*) ISA coastal defense structure (*class, instance*) ISA coastal structure (*class, instance*) ISA construction (*category*) ISA result of artificial process (*semantic role*).

- **PART-OF**: this relation also reflects the hierarchical structure of the domain. In the case of physical objects, this relation directly refers to the parts of each concept. In the case of mental objects or processes, this reflection generally refers to phases. In the
same way that objects are incomplete and can even lose their identity without one or more of their constituent parts, processes are incomplete without one or more of their phases (e.g. piping PART-OF dredging).

- **MADE-OF**: this relation links both artificial and natural objects to the material they are made of, and thus bears a certain resemblance to the PART-OF relation without being the same. Even though the material of an object is part of it, this relation is different from the PART-OF relation since material is variable. For example, a groyne head is PART-OF all groynes, but the same cannot be said of the material used to make this type of construction since groynes can be MADE-OF stone, concrete, or wood.

- **DELIMITED-BY**: this relation is used for physical objects, and marks the boundaries, dividing one object from another. This is a domain-specific relation, mainly for geographic entities, such as the different layers of the atmosphere or the Earth. For example, the stratosphere and mesosphere are DELIMITED-BY stratopause, and the earth’s crust and mantle are DELIMITED-BY Mohorovicic’s discontinuity.

- **LOCATED-AT**: this relation is relevant when the location of a physical object is an essential characteristic for its description. For instance, a groyne is not a groyne if it is not located on the coast. This relation sometimes seems to converge with the PART-OF relation. In such cases, the PART-OF relation overrides the LOCATED-AT relation. For example, a river bed is PART-OF a river instead of LOCATED-AT a river, because a river cannot exist without its bed.
TAKES-PLACE-IN: this relation describes the context of processes which have spatial and temporal dimensions. The distinction between this relation and LOCATED AT is based on the fact that processes are not as bounded in space as objects, and also have a temporal dimension. For example, littoral drift TAKES-PLACE-IN the sea; and thermal low TAKES-PLACES-IN summer.

ATTRIBUTE-OF: this relation is only useful for concepts designated by specialized adjectives, such as isotropic, alluvial, abyssal, etc., or nouns that designate the properties of other concepts, such as altitude, capacity, coefficient, etc. (e.g. littoral ATTRIBUTE-OF coast; permeability ATTRIBUTE-OF permeable groyne).

RESULT-OF: this relation is relevant to either processes or entities that are derived from other processes. Even though processes and entities can be the result of another process, a process cannot be the result of an object. For example, accretion is the RESULT-OF sedimentation (process), but it cannot be regarded as the RESULT-OF sediments (object).

AFFECTS: this relation, along with RESULT-OF, are crucial conceptual relations in dynamic systems since both have a high combinatorial potential and can relate all kinds of concepts to changing environments. Actually, they are the only ones that can be shared by concepts belonging to all three macro-templates. They link processes or objects that cause a change in any other object or process without producing a final result (e.g. groyne AFFECTS littoral drift). Moreover, complex conceptual relations such as AFFECTS, can generate a hierarchy of domain-specific relations such as RETARDS (beach nourishment RETARDS beach erosion), CHANGE-STATE-OF (temperature CHANGE-STATE-OF water), ERODES (water ERODES rocks), etc.
HAS-FUNCTION: this relation not only links objects or processes that are artificially created or carried out with a specific function, but also natural entities which, despite not being goal-directed, can be used for human profit. Natural concepts with a function are aquifer (HAS-FUNCTION water supply), sand (HAS-FUNCTION beach nourishment), etc. As in the case of AFFECTS, HAS-FUNCTION can also be associated with other domain-specific subordinate relations, such as MEASURES for instruments (a pluviometer MEASURES precipitation); STUDIES, for sciences (potamology STUDIES surface currents); and REPRESENTS for graphics, maps and charts (a hydrograph REPRESENTS rate of water flow).

EFFECTED-BY: this relation is only used for *instruments* that carry out some *process* or create an *entity*. For example, dredging is EFFECTED-BY a dredger and a marigram is EFFECTED-BY a tide gauge. This relation is especially meaningful in those domains where human interaction plays an essential role as is the case of environmental contexts.

Evidently, each of the above relations have its inverse relation (GENERIC-OF ↔ IS-A; RESULT-OF ↔ CAUSES; PART-OF ↔ HAS-PART, etc), which in our database depends on the direction of the arrow linking concepts.

4.2. Ontological Structure
The conceptualization of any domain depends on the task to be accomplished. In this case, our task is to achieve interoperability for descriptions in Spanish, English, and German of environmental entities (concepts and roles). A major problem in modeling any domain is the fact that languages can reflect different conceptualizations and construals. A case in point is the concept of *canal*, whether natural or artificial. The
terminological designations for the generic concept and for subtypes in English and Spanish do not provide us with a one-to-one match, despite referring to the same physical reality.

Relations in OWL-DL are conceived as roles, which are used instead of binary predicates. Concepts correspond to unary predicates. They represent categories, and can be instantiated by individuals. For example, INSTRUMENT, within the DESCRIPTION macro-template, is a primitive conceptual category, which represents a simple or complex object used to carry out a specific purpose.
This type of network can be regarded as an incipient ontology, which provides the raw material to build a maximally coherent knowledge base, whose objective is to produce new knowledge, prove the consistency of existing knowledge, and to enhance searches. Minimally, an ontology includes classes of domain entities, their instances, and the relations holding among the instances. However, this is the most basic type. Evidently, placing more restrictions on entities makes the resulting ontology more sophisticated and capable of automated reasoning. One of the elements necessary for such reasoning are axioms or logical sentences used to make explicit assertions about domain entities.

For example, the top-level categories in INSTRUMENT [‘Instrumento’] represent the principal types of function that a scientific instrument can have in this domain through its subtypes: (i) MEASURING INSTRUMENT [‘Instrumento de medición’]; (ii) RECORDING INSTRUMENT [‘Instrumento registrador’]; (iii) SAMPLING INSTRUMENT [‘Instrumento de muestreo’]. The relations here include the fact that all these concepts are a type of instrument and the different instances of each class. The same is true if we focus on the subordinate level of RECORDING INSTRUMENT [‘Instrumento registrador’]:

Figure 3. EcoLexicon representation: Instrumento [Instrument]
Again, only ISA relations are meaningful enough to represent this concept. However, from the third level of specialization (TIDE GAUGE ['Mareógrafo'], domain-specific relations such as EFFECTED-BY ['se hace con'] and MEASURES ['mide'] are displayed:
This is due to the fact that both macro-roles and top level categories are based on generic concepts from general language. They act like semantic primitives modeling the domain from an ontological perspective. Nevertheless, specialized subcategories must be organized in a much more sophisticated way than a static hierarchical structure. Domain-specific and non-hierarchical relations make knowledge representation more meaningful and connected to reality since they show both multidimensionality and dynamism.

This is in line with the levels of conceptual specificity described in prototype theory. According to Rosch (1979), both superordinate and subordinate categories have fewer defining attributes than basic level categories. On the contrary, the basic level is the most inclusive level at
which there are characteristic patterns of behavioral interaction and for which a clear visual image can be formed (Cruse and Croft, 2004: 83).

In our case study, RECORDING INSTRUMENT [‘Instrumento registrador’] represents the superordinate level, TIDE GAUGE [‘Mareógrafo’] the basic level, and FLOAT GAUGE [‘Mareógrafo de caja’] the subordinate level. Accordingly, domain-specific relations are only shown when the concept TIDE GAUGE [‘Mareógrafo’] is activated since it is in the basic level where concepts have a greater degree of intracategorial similarity as well as a lesser degree of intercategorial similarity. RECORDING INSTRUMENT [‘Instrumento registrador’] is too generic to have more than the most basic hierarchical relations since it is not linked to any specific mental image.

As for the FLOAT GAUGE [‘Mareógrafo de caja’] conceptual network, it does not need any other relations, since it inherits those belonging to TIDE GAUGE [‘Mareógrafo’]. A mental image of this subordinate level also matches its superordinate, and its differentiating attributes can only be shown in its definitional structure.

5. **Graphical Information**

Graphical information is essential in the representation of specialized concepts in a knowledge base. The images in EcoLexicon\(^2\) complement the linguistic information provided and focus on the conceptual relations that are most important for the concept type and its level of specificity. This is possible, thanks to the classification of images in terms of iconicity, abstraction and dynamism.

Iconicity refers to the degree of resemblance between the image and what it depicts. Accordingly, the degree of iconicity of an image depends on the number of characteristics that it shares with its referent. Abstraction refers to the effort needed by users to understand the image and recognise the concept represented. Finally, dynamism involves the implicit or explicit representation of movement.
Furthermore, certain types of concepts and their relations are best represented by specific types of images. The following examples illustrate the representation of instrument concepts of greater or lesser specificity: RECORDING INSTRUMENT, TIDE GAUGE, and FLOAT GAUGE.

Figure 6 shows the Global Observing System, which records data from devices such as ocean buoys, satellite soundings, and weather radars. Images for such concepts are difficult to find because RECORDING INSTRUMENT, for example, is so general and has few defining attributes. In our database, this type of generic concept is generally designated by general language words.

---

**RECORDING INSTRUMENT**

Figure 6: Graphical representation of RECORDING INSTRUMENT

In this case, the image selected focuses on HAS-FUNCTION, the most central conceptual relation of RECORDING INSTRUMENT. The function of any recording instrument is to obtain data and send it to receivers. The image has a low level of abstraction, and this facilitates understanding. In this case, the function of the concept is reflected as a process, which can be divided into phases. This requires a dynamic representation, which is represented by arrows, showing the transmission flow of scientific information.
Figure 7 shows the image selected for TIDE GAUGE, an instrument concept at the basic level. Basic-level concepts in specialized knowledge structures are linked to images. As such, they reflect a wider set of generic-specific and associative domain-specific conceptual relations. In this case, an iconic image is the most suitable given that this type of images depicts shapes, colours, proportions, size, etc. The image in Figure 7 is called a ghost view, because it shows hidden elements that otherwise could not be seen. Dynamism is also conveyed by an arrow to represent the result of the recording instrument (i.e. a mareogram.).

![Figure 7: Graphical Representation of TIDE GAUGE](image)

At a more specific level, FLOAT GAUGE is a type of TIDE GAUGE based on the movement of a float which records changes in water level. As a result, it inherits the characteristics and relations of the TIDE GAUGE, and adds a new feature *measuring water level through a float*. 
Figure 8 shows a very realistic image (photograph) along with a schematic view of the different components of a float gauge. It represents the same concepts and relations as Figure 7, but with a higher level of abstraction.

Iconicity is particularly useful for the representation of the hierarchical relations ISA and PART-OF, and the non-hierarchical relations DELIMITED-BY or LOCATED-AT. Abstraction increases communicative efficacy and facilitates understanding of more complex domain-specific relations like AFFECTS or ATTRIBUTE-OF. Dynamism highlights the procedural character of concepts evidenced by relations like EFFECTED-BY, RESULT-OF or HAS-FUNCTION. These considerations are fundamental for the selection of graphical information for the representation of concepts in a specialized knowledge base in which conceptual, linguistic, and graphical information converges in maximally coherent knowledge representation.
Notes
[1] This research is part of the project *Marcos de conocimiento multilingües en la gestión integrada de zonas costeras* (MarcoCosta) (P06-HUM-01489), funded by the Andalusian Regional Government.

[2] All images contained in this paper have been extracted from the EcoLexicon multimedia terminological database, available at: http://manila.ugr.es/visual/index.html.

[3] This paper was presented at the XVIII FIT World Congress and full acknowledgment and copyright is given to the source. Permission of the publication has been granted by the author(s).

References


