

Flexible terminological definitions and conceptual frames

Antonio San Martín* and Pilar León-Araúz

LexiCon Research Group. Department of Translation and Interpreting, University of Granada
Calle Buensuceso n.º 11, 18071 Granada, Spain

ABSTRACT

This paper focuses on the manual creation of context-dependent natural-language definitions in EcoLexicon, a terminological knowledge base on the Environment. Given the interdisciplinary nature of the environmental domain, many concepts in EcoLexicon show a high degree of multidimensionality. In other words, this means that concepts can be described from many different perspectives. For such concepts, a single definition that encompasses the whole environmental domain is not informative enough because not all environmental domains describe concepts in the same fashion. For that reason, we propose the creation of flexible definitions.

A flexible definition is a system of definitions for the same concept composed of a general environmental definition with a set of recontextualized definitions (definitions that describe a concept from the viewpoint of a certain subject domain). This approach is based on category definitional templates and conceptual frames that provide a consistent way of managing and representing the dimensions of contextually-variable concepts in terminological definitions.

1 INTRODUCTION

A conceptual system is considered to be multidimensional when its concepts are categorized according to different characteristics, and thus showing their different dimensions (Kageura, 1997). Conceptual representations in terminological knowledge bases tend to be monodimensional. Sometimes, this may be due to the fact that the domain to be described is very constrained and there is no need to represent several dimensions. However, the usual case is that the terminologist prefers to avoid the difficulties associated with managing several dimensions. One of the problems that arise with multidimensional conceptual systems is the writing of natural-language definitions based on feature inheritance, given that the relevance of any conceptual feature can change depending on the dimension being considered and concepts can have more than one hypernym (Bowker, 1996, p. 785).

In a terminological knowledge base for translators, which is the case of EcoLexicon, the framework for this research, multidimensional knowledge representation allows the user to acquire a better insight into a given concept. This is very useful for translators because they may need to translate terms that represent concepts viewed from very different points of view.

For the representation of multidimensionality in terminological definitions, our proposal consists of the creation of several natural-language definitions for a given concept, each one describing the concept from a different

subdomain of the discipline of the Environment. As a consequence, the concept to be defined is situated in different conceptual frames, which also affects which knowledge is represented in the definitions.

2 CONCEPTUAL NETWORKS IN ECOLEXICON

EcoLexicon (<http://ecolexicon.ugr.es>) is a terminological knowledge base on the environment. It is concept-oriented and multilingual. So far, it has 3,533 concepts and 18,798 terms in English, Spanish, German, French, Russian, Modern Greek, and Dutch as well as linguistic and phraseological information for each term. The main target users of EcoLexicon are translators, who must undoubtedly understand what they read and write in subject fields where they are not experts but need to sound like they were. This entails that they need to acquire new specialized knowledge in a very short time. To enhance knowledge acquisition, conceptual information in EcoLexicon is stored and represented in different ways.

On the one hand, specialized knowledge is represented by means of conceptual networks codified in terms of conceptual propositions in the form of a triple (CONCEPT *relation* CONCEPT), for instance, <SAND *type-of* SEDIMENT> or <MORTAR *made-of* SAND>. The conceptual relations used in EcoLexicon include both hierarchical (hypernymic and meronymic) and non-hierarchical relations, some of which are domain-specific. Concept nature (OBJECT, PROCESS, or PROPERTY) determines the combinatorial potential of concepts by means of a closed inventory of conceptual relations (León Araúz & Faber, 2010, p. 14).

On the other hand, conceptual information is also shown in the form of natural language definitions in English and Spanish, which are based on the most prototypical conceptual propositions established by the concept to be defined. Additionally, domain-specific knowledge is also presented in the form of images and videos.

2.1 Frame-based Terminology

The theoretical and methodological framework of EcoLexicon is called Frame-Based Terminology (Faber, 2012). It is a cognitive approach to Terminology inspired in the notion of frame as “any system of concepts related in such a way that to understand any one of them you have to

* To whom correspondence should be addressed: asanmartin@ugr.es

networks restricted to the conceptual propositions that are salient in a certain domain (Fig. 3).

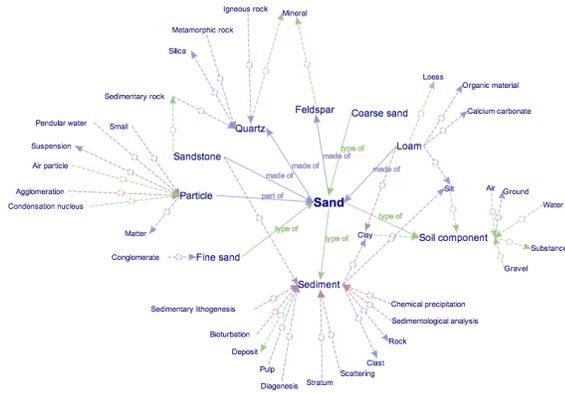


Fig. 3. SAND recontextualized in the SOIL SCIENCES domain.

3 DEFINITIONS IN ECOLEXICON

In EcoLexicon, definitions are regarded as mini-knowledge representations (Faber, 2002, p. 345). As such, they are based on the most representative conceptual propositions established by the concept in EcoLexicon. Each conceptual proposition is considered to be a feature of the concept.

The representativeness of each feature is determined by the category assigned to the concept being defined. Each category has a set of representative conceptual relations that describe it. They are specified in the category definitional template (León Araúz, Faber, & Montero Martínez, 2012, pp. 153–154).

3.1 Category definitional templates

Category definitional templates are schematic representations of the most prototypical relations established by the concepts that are members of the same category. They guide the formulation of definitions. They are encoded in the form of a slot-filler table like Martin’s frame-based definitions (Martin, 1998). In our approach, the slots correspond to conceptual relations and the values to the concepts linked to the definiendum by means of the conceptual relations. When applying a template to a concept, it may only inherit the relation (slot) with the defined concept (value) in the template or activate a more specific concept than the one in the template. An example would be the template for HARD_COASTAL_DEFENCE_STRUCTURE (Table 1), which is applied to the definition of GROIN (Table 2), a member of this category.

HARD_COASTAL_DEFENCE_STRUCTURE	
<i>type-of</i>	CONSTRUCTION
<i>located-at</i>	SHORELINE
<i>made-of</i>	MATERIAL

<i>has-function</i>	COASTAL_DEFENCE
---------------------	-----------------

Table 1. HARD_COASTAL_DEFENCE_STRUCTURE category definitional template (León Araúz et al., 2012, p. 156)

GROIN	
Hard coastal defence structure made of concrete, wood, steel and/or rock perpendicular to the shoreline built to protect a shore area, retard littoral drift, reduce longshore transport and prevent beach erosion.	
<i>type-of</i>	HARD COASTAL DEFENCE STRUCTURE
<i>located-at</i>	PERPENDICULAR TO SHORELINE
<i>made-of</i>	CONCRETE WOOD METAL ROCK
<i>has-function</i>	SHORE PROTECTION LITTORAL DRIFT RETARDATION LONGSHORE TRANSPORT REDUCTION BEACH EROSION PREVENTION

Table 2. Definition of GROIN after the application of the HARD_COASTAL_DEFENCE_STRUCTURE category template (León Araúz et al., 2012, p. 156).

Category definitional templates are created by combining a bottom-up and top-down approach. On the one hand, the top-down approach signifies that the membership in top-level categories partly determines the configuration of the definition. On the other hand, we also take into account the extension of a category (bottom-up approach), because a category is not only determined by its superordinates but also by its members. Consequently, before defining a concept, it is necessary to categorize it and then analyze the other members of the category so as to modulate the template inherited from superordinate categories.

4 FLEXIBLE DEFINITIONS

For concepts with a high level of contextual variation, a single definition that encompasses the whole environmental domain is not sufficiently informative, as is the case of these definitions of SAND in different environmental terminological resources:

- Mineral rock fragments (sediment) which have a particle size between 0.06 millimetres and 2.0 millimetres, which is between -1.0 and 4.0 on the phi (φ) scale. [A Dictionary of Environment and Conservation (Park, 2007)]
- Unconsolidated sediment consisting of mineral granules ranging between about 60 μm and 2 mm in diameter. particles of silica or quartz (SiO₂) are common components of sand. [The Environment Dictionary (Kemp, 1998)]

- A loose material consisting of small mineral particles, or rock and mineral particles, distinguishable by the naked eye; grains vary from almost spherical to angular, with a diameter range from 1/16 to 2 millimeters. [General Multilingual Environmental Thesaurus (GEMET) (European Environment Agency, 2012)]

These definitions of SAND are not very useful for a translator dealing with the concept of SAND in different environmental subdomains. For instance, in *CIVIL ENGINEERING*, it is important to know the different functions of sand, and in *SOIL SCIENCES*, how sand affects the properties of the soils in which it can be found. Furthermore, no consensus seems to exist regarding SAND hypernyms (i.e. FRAGMENT, SEDIMENT, MATERIAL), because they are also source of contextual variation, which shows that knowledge is not naturally structured in clear-cut taxonomies.

For that reason, we propose the creation of ‘flexible definitions’. A flexible definition is a system of definitions for the same concept composed of a general environmental definition along with a set of recontextualized definitions derived from it, which situate the concept in different domains.

Since flexible definitions follow the same premises used in the recontextualization of conceptual networks (section 2.2.), they account for the systemic factor in definition building. According to Seppälä (2012, p. 153), as a function of this factor, the relevant features to be included in a definition are determined by the conceptual system in which the concept is inserted.

Recontextualized definitions are standalone, and thus convey all the necessary information to define a concept in a certain domain, independently of the other definitions in the set. Table 3 presents an extract of the flexible definition of SAND²:

General Environmental Definition	Mineral material consisting mainly of particles of quartz ranging in size of 0.05-2 mm.
Geology Definition	Sediment consisting mainly of particles of quartz ranging in size of 0.05-2 mm that is part of the soil and can be found in great quantities in beaches, river beds, the seabed, and deserts.
Soil Sciences Definition	Unconsolidated inorganic soil component consisting mainly of particles of quartz ranging in size of 0.05-2 mm that are the result of weathering and erosion. It renders soils light, acidic, and permeable.

² Not all domains are included in this example.

Civil Engineering Definition	Natural construction aggregate consisting mainly of particles of quartz ranging in size of 0.05-2 mm that is mixed with cement, lime and other materials to produce concrete and mortar.
-------------------------------------	--

Table 3. Extract of the flexible definition of SAND

In a flexible definition, the general environmental definition encodes the basic meaning present in all contextual domains and the recontextualized definitions can be considered a variation of it. For this reason, the general environmental definition includes those propositions shared by all the recontextualized definitions (e.g., in the definition of SAND: <SAND made_of QUARTZ>³).

4.1 Creation of the recontextualized hierarchies

One of the main difficulties posed by flexible definitions is that, contrary to what one might think, even hypernyms are subject to contextual variation. Quite understandably, this can impair feature inheritance in a hierarchy. As shown in Table 3, SAND is categorized in different ways depending on how the concept is prototypically conceived in each domain.

Since a coherent hierarchy needs to be specified before the defining process in order to assure correct feature inheritance⁴, in the case of flexible definitions, each contextual domain requires its own hierarchy. The main information sources that determine how to categorize a concept are the definition of the concept in other terminological resources and KP-based corpus analysis.

On the one hand, extracting the hypernyms of a concept from other resources has its limitations. The first is the fact that it is not usual to find various definitions for the same concept in resources that focus on different domains. For instance, definitions of SAND can be found in Geology and Soil Sciences dictionaries and glossaries. But it is unusual to find an entry for SAND in Water Treatment or Meteorology resources, since the concept is less prototypical in the latter.

On the other hand, however, KP-based corpus analysis (Meyer, 2001) is more useful for the extraction of context-specific hypernymic relations. This method permits the specification of the possible categorizations of a concept in a given contextual domain by applying KP searches to domain-specific corpora.

³ For details on how the general environmental definition is built and the way the recontextualized definitions stem from it, see León Araúz & San Martín (2012).

⁴ Currently EcoLexicon is stored in the form of a relational database. Although it is in the process of becoming a formal ontology, no feature inheritance mechanisms have been implemented yet. However, terminologists manually take feature inheritance into account during conceptual modeling and definition writing.

However, all the hypernym candidates extracted with these two methods can only be used as a guide. Concepts can be categorized in several ways even in the same knowledge domain. In fact, many of the categories that can be extracted with these two methods could be considered ad hoc categories (Barsalou, 1983), constructed for a specific purpose in a certain situation and lacking conventionalization, rather than well-established categories.

The main guidelines for the structuring of recontextualized hierarchies in EcoLexicon are coherence (for correct feature inheritance once all the data is implemented in an ontology) and the activation of the most prototypical underlying conceptual frame.

4.2 Underlying conceptual frames

According to the principle of cognitive economy (Rosch, 1978, p. 28), categorization serves to mentally store and retrieve the properties generally associated with a concept in a cost-efficient manner. This also applies to the choice of genus in a definition. It follows that by choosing a genus, certain features are assigned to the definiendum (those inherited via the genus) without the need to list them explicitly in the definition.

As for recontextualized definitions, the choice of genus is even more important because by categorizing a concept as a member of a contextual domain, it is inserted into a specific conceptual frame. Such a frame takes the form of a description that relates different conceptual categories. Whereas in FrameNet (Ruppenhofer, Ellsworth, Petruck, Johnson, & Scheffczyk, 2006), frames are described by stating the relation between frame elements, in our proposal we use the categories in the Environmental Event as well as any concept in EcoLexicon. If the frame is an event composed of different stages, the information is expressed sequentially (Table 4).

Unlike in Fillmore's double-decker definitions (2003) or Maks' contextual definitions (2006), the conceptual frame is not part of the definition. It is created in order to guide the creation of the definitional templates of the categories appearing in it. In other words, the definition includes the information of the conceptual frame. As a consequence, the recontextualized definition of a concept is determined by the category to which it belongs and the underlying frame in which it takes part.

When SAND is categorized as SEDIMENT in *GEOLOGY*, SOIL_COMPONENT in *SOIL SCIENCES*, and CONSTRUCTION_AGGREGATE in *CIVIL ENGINEERING* this situates it in the frames of SEDIMENTATION (Table 4), SOIL_PROPERTIES (Table 5), and COMPOSITE_MATERIAL_PRODUCTION (Table 6), respectively.

Frame: SEDIMENTATION
Contextual domain: <i>GEOLOGY</i>
<ol style="list-style-type: none"> 1. A MATERIAL suffers WEATHERING and EROSION and, as a consequence, becomes a SEDIMENT. 2. NATURAL_AGENTS transport the SEDIMENT. 3. The SEDIMENT is deposited in a GEOGRAPHIC_FEATURE.

Table 4. SEDIMENTATION frame

Frame: SOIL_PROPERTIES
Contextual domain: <i>SOIL SCIENCES</i>
SOIL is composed of SOIL_COMPONENTS that determine the SOIL'S PHYSICAL, CHEMICAL, and BIOLOGICAL_PROPERTIES.

Table 5. SOIL_PROPERTIES frame

Frame: COMPOSITE_MATERIAL_PRODUCTION
Contextual domain: <i>CIVIL ENGINEERING</i>
A HUMAN_AGENT produces COMPOSITE_MATERIAL by mixing a CONSTRUCTION_AGGREGATE with a MATRIX so as to use it in CONSTRUCTION.

Table 6. COMPOSITE_MATERIAL_PRODUCTION frame

5 THE CASE OF SAND IN THE CONTEXTUAL DOMAIN OF CIVIL ENGINEERING

In the *CIVIL ENGINEERING* hierarchy, SAND is categorized as a CONSTRUCTION_AGGREGATE, which itself is a subordinate of MATERIAL. Therefore, the category definitional template for MATERIAL affects the category definitional template of CONSTRUCTION_AGGREGATE, and the latter can be used, in turn, for the definition of SAND.

Table 7 and 8 show the category definitional templates for MATERIAL and CONSTRUCTION_AGGREGATE.

MATERIAL	
<i>type-of</i>	PHYSICAL OBJECT
<i>made-of</i>	SUBSTANCE
<i>component-of</i>	PHYSICAL OBJECT

Table 7. MATERIAL category definitional template

CONSTRUCTION_AGGREGATE	
Material consisting of particles that is mixed with a matrix to produce composite material to be used for construction.	
<i>type-of</i>	MATERIAL
<i>made-of</i>	SUBSTANCE_PARTICLES
<i>component-of</i>	COMPOSITE_MATERIAL (when mixed with MATRIX)

<i>has-attribute</i>	NATURAL/ARTIFICIAL
<i>has-function</i>	CONSTRUCTION

Table 8. CONSTRUCTION_AGGREGATE category definitional template with definition

The category definitional template for CONSTRUCTION_AGGREGATE is partly determined by its superordinate concept MATERIAL. Therefore, it activates the relations *made-of* and *component-of*. The underlying COMPOSITE_MATERIAL_PRODUCTION frame (Table 6) is the reason why the concepts COMPOSITE_MATERIAL and MATRIX are part of one of the values in the template. The relations *has-attribute* and *has-function*, as well as the specification that a <CONSTRUCTION_AGGREGATE is *made-of* PARTICLES>, are included in the template because of the subordinate concepts of CONSTRUCTION_AGGREGATE. An analysis of its category members such as SAND, GRAVEL, or SLAG reveals that such information is relevant for the description of the category.

Table 9 shows the definition of SAND after the application of the category definitional template of CONSTRUCTION_AGGREGATE.

SAND	
Natural construction aggregate consisting mainly of particles of quartz ranging in size of 0.05-2 mm that is mixed with cement, lime and other materials to produce concrete and mortar.	
<i>type-of</i>	CONSTRUCTION_AGGREGATE
<i>made-of</i>	(0.05-2 MM) QUARTZ_PARTICLES
<i>component-of</i>	MORTAR/CONCRETE (when mixed with CEMENT/LIME)
<i>has-attribute</i>	NATURAL

Table 9. SAND category definitional template and definition in the CIVIL ENGINEERING contextual domain

As can be observed in Table 9, the proposition <SAND *has-function* CONSTRUCTION> does not appear either in the definition or in the template for SAND. This is because this proposition is inherited from CONSTRUCTION_AGGREGATE, and it is thus implicit. The other relations are represented in the definition because they are a further specification of the category definitional template of CONSTRUCTION_AGGREGATE.

6 CONCLUSIONS

A single definition is not sufficient to describe multidimensional concepts that participate in many different conceptual frames. Context is an essential factor in the choice of definitional information. Depending on the context, a concept may be categorized differently and

therefore establish a link to a different conceptual frame. This underlying conceptual frame guides the configuration of the category definitional template to be used in the defining process.

Because of the inherent limitations of using a closed inventory of conceptual relations, category definitional templates are not as expressive as natural language. Thus, there is the need to nuance the information in the templates. Although the configuration of category definitional templates and frames can be time-consuming, we plan to streamline these tasks in the future by formalizing all this information in an ontology. Our approach based on category definitional templates and frames provides a consistent way of managing and representing the different dimensions of contextually-variable concepts in terminological definitions. This enhances knowledge acquisition in terminological knowledge bases because it affords users a clearer and more coherent vision of each concept and its contextualized meaning in different knowledge domains.

ACKNOWLEDGEMENTS

This research was funded by the Spanish Ministry of Economy and Competitiveness (Project FFI 2011-22397) and the Spanish Ministry of Education, Culture, and Sports (FPU Program AP2009-4519).

REFERENCES

- Barsalou, L. W. (1983). Ad hoc categories. *Memory cognition*, **11**(3), 211–227.
- Bowker, L. (1996). Learning from Cognitive Science: Developing a New Approach to Classification in Terminology. In M. Gellerstam, J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström, and C. R. Papehl (Eds.), *Euralex '96 Proceedings* (pp. 781–787). Göteborg: EURALEX.
- European Environment Agency. (2012). GEneral Multilingual Environmental Thesaurus (GEMET). <<http://www.eionet.europa.eu/gemet>>
- Faber, P. (2002). Terminographic definition and concept representation. In B. Maia, J. Haller, and M. Ulyrich (Eds.), *Training the Language Services Provider for the New Millennium* (pp. 343–354). Porto: Universidade do Porto.
- Faber, P. (2009). The cognitive shift in terminology and specialized translation. *MonTI. Monografías de Traducción e Interpretación*, **1**, 107–134.
- Faber, P. (Ed.). (2012). *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin, Boston: De Gruyter Mouton.
- Faber, P., León Araúz, P., and Prieto Velasco, J. A. (2009). Semantic Relations, Dynamicity, and Terminological Knowledge Bases. *Current Issues in Language Studies*, **1**, 1–23.
- Fillmore, C. J. (1982). Frame Semantics. In The Linguistic Society of Korea (Ed.), *Linguistics in the Morning Calm* (pp. 111–137). Seoul: Hanshin.
- Fillmore, C. J. (2003). Double-Decker Definitions: The Role of Frames in Meaning Explanations. *Sign Language Studies*, **3**(3), 263–295. doi:10.1353/sls.2003.0008

- Kageura, K. (1997). Multifaceted/Multidimensional Concept Systems. In S. E. Wright and G. Budin (Eds.), *Handbook of Terminology Management. Volume 1: Basic Aspects of Terminology Management* (pp. 119–132). Amsterdam/Philadelphia: John Benjamins.
- Kemp, D. (1998). *The Environment Dictionary*. London, New York: Routledge.
- León Araúz, P. (2009). *Representación multidimensional del conocimiento especializado: el uso de marcos desde la macroestructura hasta la microestructura*. PhD Thesis. Universidad de Granada.
- León Araúz, P., and Faber, P. (2010). Natural and contextual constraints for domain-specific relations. In V. Barbu Mititelu, V. Pekar, and E. Barbu (Eds.), *Proceedings of the Workshop Semantic Relations. Theory and Applications* (pp. 12–17). Valletta.
- León Araúz, P., Faber, P., and Montero Martínez, S. (2012). Specialized language semantics. In P. Faber (Ed.), *A Cognitive Linguistics View of Terminology and Specialized Language* (pp. 95–175). Berlin, Boston: De Gruyter Mouton.
- León Araúz, P., Reimerink, A., and Faber, P. (2009). Knowledge Extraction on Multidimensional Concepts: Corpus Pattern Analysis (CPA) and Concordances. *8ème conférence internationale Terminologie et Intelligence Artificielle*. Toulouse.
- León Araúz, P., and San Martín, A. (2011). Distinguishing Polysemy from Contextual Variation in Terminological Definitions. In M. L. Carrió, J. Contreras, F. Olmo, H. Skorczynska, I. Tamarit, and D. Westall (Eds.), *Actas del X Congreso de la Asociación Europea de Lenguas para Fines Específicos: La investigación y la enseñanza aplicadas a las lenguas de especialidad y a la tecnología* (pp. 173–186). Valencia: Universitat Politècnica de València.
- León Araúz, P., and San Martín, A. (2012). Multidimensional Categorization in Terminological Definitions. In R. V. Fjeld and J. M. Torjusen (Eds.), *Proceedings of the 15th EURALEX International Congress* (pp. 578–584). Oslo: EURALEX.
- Maks, I. (2006). Frame-based definitions in a Learners' Dictionary for Dutch Business Language. In P. Ten Hacken (Ed.), *Terminology, Computing and Translation* (pp. 191–206). Tübingen: Narr.
- Martin, W. (1998). Frames as definition models for terms. *Proceedings of the International Conference on Professional Communication and Knowledge Transfer. Vol. 2.* (pp. 189–220). Vienna: TermNet.
- Meyer, I. (2001). Extracting knowledge-rich contexts for terminography. In D. Bourigault, C. Jacquemin, and M.-C. L'Homme (Eds.), *Recent advances in computational terminology* (pp. 279–302). Amsterdam/Philadelphia: John Benjamins.
- Park, C. (2007). *A Dictionary of Environment and Conservation*. Oxford, New York: Oxford University Press.
- Reimerink, A., and Faber, P. (2009). EcoLexicon: A Frame-Based Knowledge Base for the Environment. In J. Hřebíček, J. Hradec, E. Pelikán, O. Mírovský, W. Pillmann, I. Holoubek, and T. Bandholtz (Eds.), *European conference of the Czech Presidency of the Council of the EU TOWARDS eENVIRONMENT Opportunities of SEIS and SISE: Integrating Environmental Knowledge in Europe*. Brno: Masaryk University.
- Rosch, E. (1978). Principles of Categorization. In E. Rosch and B. B. Lloyd (Eds.), *Cognition and Categorization* (pp. 27–48). Lawrence Erlbaum Associates.
- Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., and Scheffczyk, J. (2006). *Framenet II: Extended theory and practice*. Unpublished manuscript.
- Seppälä, S. (2012). *Contraintes sur la sélection des informations dans les définitions terminographiques: vers des modèles relationnels génériques pertinents*. PhD Thesis. Université de Genève.
- Tercedor, M., and López Rodríguez, C. I. (2008). Integrating corpus data in dynamic knowledge bases The Puertoterm project. *Terminology*, **14**(2), 159–182. doi:10.1075/term.14.2.03ter