Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference

Linking a Domain-Specific Ontology to a General Ontology

Pamela Faber[1], Ricardo Mairal[2], Pedro Magaña[3]

[1] University of Granada - Buensuceso, 11 18002 Granada, Spain - pfaber@ugr.es

[2] UNED (Madrid) - Senda del Rey, 7 28040 Madrid, Spain - rmairal@flog.uned.es

[3] CEAMA (Centro Andaluz de Medio Ambiente) - Av. del Mediterráneo s/n 18006 Granada, España - pmagana@ugr.es

Abstract

Ontologies have been criticized because they are not sufficiently flexible, and thus cannot capture the dynamism and complexity of reality. However, they have increasingly come into focus because of the need for knowledge management in both general and specialized knowledge domains. EcoLexicon is a frame-based visual thesaurus on the environment that is gradually evolving towards the status of a formal ontology. For this purpose, the information in its relational database is in the process of being linked to the ontological system of FunGramKB, a multipurpose knowledge base that has been specifically designed for natural language understanding with modules for lexical, grammatical, and conceptual knowledge. This enables the explicitation of specialized knowledge as an extension of general knowledge through its representation in the domain-specific satellite ontology of a main general ontology.

1. Introduction

A domain-specific ontology of concepts within a certain field, along with their relations and properties, is a new medium for the storage and propagation of specialized knowledge (Hsieh et al. 2010). Ontologies reflect a particular conceptualization of reality through the explicit definition of concepts (terms), representing domain entities, and relations. From a more linguistic perspective, Sowa (2000: 492) defines an ontology as "a catalogue of the type of things that are assumed to exist in a domain of interest *D*, from the perspective of a person who uses a language *L* for the purpose of talking about *D*." In this respect, one way to enrich ontology elements is by including linguistic information and structure (Buitelaar et al. 2009). In this sense, a multilingual terminological knowledge base is a valuable knowledge resource since it is composed of signs in various languages that designate concepts, corresponding to mental representations of phenomena in the real world. Multilingual information is in great demand today by institutions (Montiel-Ponsoda et al. 2010), and multilingualism in ontologies benefits society because it helps to reduce confusion regarding conceptual reference in international communication. This has evident implications for e-learning and knowledge acquisition.

EcoLexicon (http://ecolexicon.ugr.es) (Faber et al. 2006; Faber et al. 2007) is a multilingual visual thesaurus on the environment in English, Spanish, and German (currently under expansion to French, Russian, and Modern Greek). Its purpose is knowledge acquisition with a view to specialized text generation by users such as scientific writers, translators, and environmentally-aware sectors of the general public. The conceptualization process of the resource first involved the semi-automatic extraction of concepts and relations from domain-specific documents (Eriksson 2007), based on semantic patterns and lexical markers. Other environmental resources used for information extraction were general upper-level ontologies, such as SIMPLE (Lenci 2000) and environmental ontologies such as SWEET (Raskin and Pan 2005).

The next step was the manual creation of an ontology 'scaffold' made up of basic class hierarchies and relations. The ontology itself is organized around direct representations of physical objects and processes (e.g. alluvial fan, erosion, weathering, etc). This basic set of concepts act as a scaffold, and their natural language descriptions provide the semantic foundation for data querying, integration and inferencing (Samwald et al. 2010). This scaffold could not be generated automatically since some of the structures and entity labels in the database needed to be changed and reinterpreted so that there would be agreement between conceptualizations in different languages. Concepts and relationships are currently in the form of semantic networks or concept maps implemented with ThinkMap. Concept maps have been successfully used in other domains, such as nutrigenomics (García Castro et al. 2006).

In EcoLexicon, environmental concepts are codified in terms of natural language definitions that are visually represented as a network of both hierarchical and non-hierarchical semantic relations extracted from a multilingual corpus of specialized texts. This is in consonance with cognitive semantics (Talmy 2000), which claims that lexical meaning is a manifestation of conceptual structure. The meaning of words thus does not depend on the world itself, but rather on our categorization of the world (Evans Bergen and Zinken 2007). Similarly, Gahegan et al. (2008) affirm that the concepts and relations used to describe the world are constructed by humans. In the vision of conceptual organization offered by cognitive semantics, lexical items are regarded as conceptual categories of distinct yet related meanings that exhibit typicality effects.

Even though this representation still needs to be further enriched and systematized so as to allow more sophisticated reasoning processes, it permits EcoLexicon to be connected to other ontologies and resources. Accordingly, this paper describes the integration of EcoLexicon into FunGramKB, a multipurpose general knowledge base, which uses COREL as a representation language. It explains how the general concepts in FunGramKB can be extended and reused in deep semantic representations in a domain-specific ontology.

This paper is organized as follows Sections 2 and 3 describe the conceptual organization and design of EcoLexicon and FunGramKB, respectively. Section 4 explains the integration process, the semantic representation as applied to specialized knowledge, the promotion/demotion of general language concepts in the domain-specific ontology, and the advantages gained from contextualizing a specialized knowledge resource within a general ontology. Section 5 presents the conclusions derived, and Section 6 lists the references cited.

2. EcoLexicon

EcoLexicon is a specialized knowledge resource on the environment. It is hosted in a relational database, which is currently being linked to an ontology for reasoning techniques and user queries (Leon, Magaña, and Faber 2008; León and Magaña 2010). As such, it focuses on: (1) conceptual organization; (2) the multidimensional and multilingual nature of terminological units; (3) the extraction of semantic and syntactic information through the use of multilingual corpora. Based on cognitive semantics and situated cognition (Barsalou 2008), the information stored in EcoLexicon is structured in terms of propositions and knowledge frames (Fillmore 1985) that are organized in an ontological structure, which focuses on perceptual information and semantic relations.

EcoLexicon can be regarded as a linguistically-based ontology since its conceptual design is derived from information semi-automatically extracted from specialized texts and the structure of terminological definitions. Its top-level concepts are object, event, and attribute categories. In EcoLexicon, abstract concepts include theories, equations, and units for measuring physical entities. In contrast, physical or concrete concepts are those occupying space and occurring over a period of time. They include natural entities, geographic landforms, water bodies, constructions, and the natural and artificial process events in which they can potentially participate.

In Ecolexicon, the most generic or top-level categories of a domain are configured in a prototypical domain event or action-environment interface (Barsalou 2008), called the Environmental Event (EE) (Faber, Márquez, and Vega 2005) (see Figure 1).

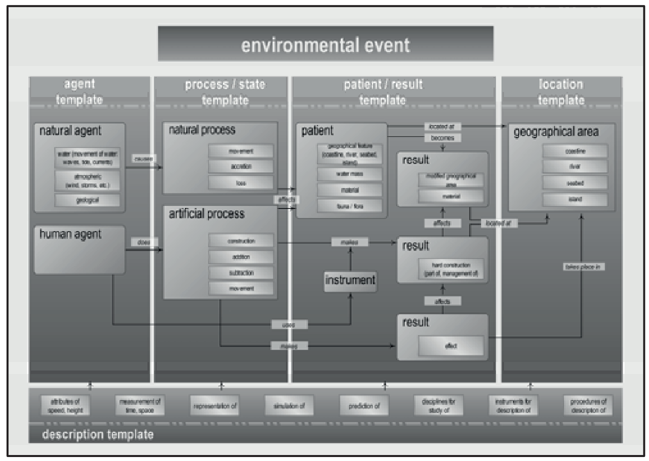

Figure 1. Environmental Event (EE)

The EE has two types of agent that can initiate processes. Such agents can be inanimate (natural forces) or animate (human beings). Natural agents, such as water movement (e.g. waves, tides, and currents) and atmospheric phenomena (e.g. winds and storms) cause natural processes such as littoral drift and erosion in a geographic area such as the coast. These processes affect other entities or PATIENTS (e.g. beaches, sea ports, and seabed) which as a RESULT, may suffer changes (e.g. loss/deterioration/ creation of beaches, and modifications in seabed composition). HUMAN AGENTS can also implement ARTIFICIAL PROCESSES (e.g. constructions), which can generate or prevent EFFECTS normally caused by natural processes.

For instance, a TSUNAMI, as a large high-velocity wave, can initiate a FLOOD-EVENT, which affects a patient (LANDFORM or LAND AREA) and produces a certain result (EROSION, MODIFIED LANDFORM. etc.). Alternatively, within the context of other processes or events, it can be regarded as the result of the displacement of the sea floor (i.e. sudden faulting, landsliding, or volcanic activity). This event provides a template or frame applicable to all levels of information structuring. Currently, the event roles provide the category templates for domain concepts. The resulting general frame facilitates and enhances knowledge acquisition since the information in term entries is internally as well as externally coherent (Faber et al. 2007).

The user interface offers various types of information, such as multilingual term correspondences in various languages, graphical files with images of the concept, domain membership, and a ThinkMap representation showing links to related concepts (Figure 2).

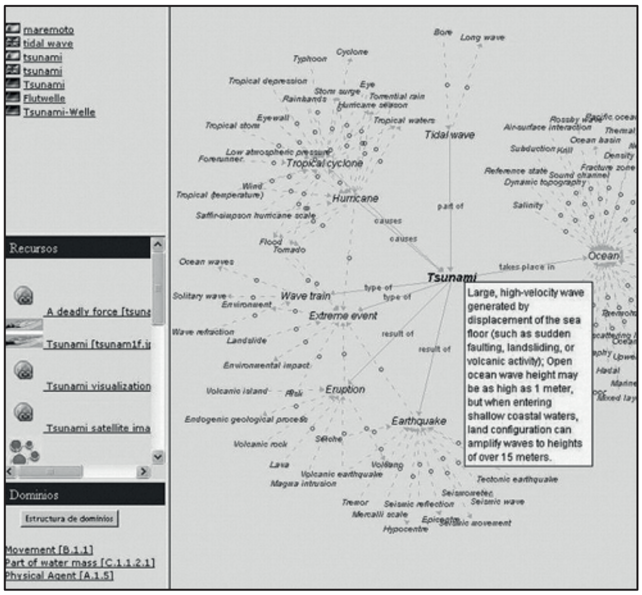

Figure 2. EcoLexicon representation of TSUNAMI

This type of visualization was selected because a semantic network was considered to be the most appropriate representation method for capturing and encapsulating large amounts of semantic information in an intelligent environment (Peters and Shrobe 2003). Each concept is linked to other concepts by a closed inventory of semantic relations. Apart from the conceptual representation and the definition that appears when the cursor is placed on the concept, information is also provided regarding the following: (i) the terms in different languages; (ii) graphical resources; (iii) conceptual role in the general event structure (see Figure 2).

However, for the specialized environmental knowledge in EcoLexicon to be used in more complex tasks, its stored data is now in the process of being integrated into the ontological system of FunGramKB, a multipurpose knowledge base that has been specifically designed for natural language understanding with modules for lexical, grammatical, and conceptual knowledge (Periñan-Pascual and Arcas-Túnez 2010). Its deep semantic representation of concepts has greater expressive power, and also allows the statement of co-reference between internal conceptual units, which is impossible to fully describe via surface-semantic representations, based on semantic relations.

3. FunGramKB

The translation of conceptual information from EcoLexicon to FunGramKB is feasible because both resources are based on the extraction of concepts from natural language resources. FunGramKB is an online environment for the semiautomatic construction of a multipurpose lexico-conceptual knowledge base for NLP systems (Figure 3).

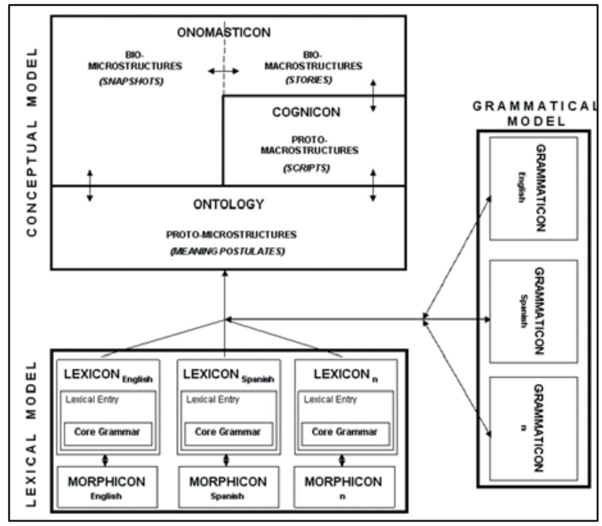

Figure 3. Architecture of FunGramKB

As observed in Figure 3, FunGramKB has a lexical level and a grammatical level, which are language-specific, and a conceptual level, which is not. This paper focuses on the integration of EcoLexicon and FunGramKB at the conceptual level. The conceptual level in FunGramKB is composed of the following: (i) an Ontology of concepts defined with meaning postulates; (ii) a Cognicon with procedural knowledge stored as scripts; (iii) an Onomasticon with information about instances of entities and events in the form of stories and snapshots.

3.1 The FunGramKB Ontology

The FunGramKB ontology is a concept taxonomy, derived from linguistic concepts, in which interlinguistic differences in syntactic constructions do not involve conceptual differences. It is general-purpose, and not domain-specific. However, since expert knowledge stems from general knowledge, it can be extended to include specialized knowledge by establishing links to satellite domain-specific ontologies, as shown in Figure 4.

The concepts of FunGramKB belong to three levels. The upper level is composed of 42 metaconcepts, marked with the symbol #. They constitute the upper level in the taxonomy as a result of the analysis of the most relevant linguistic ontologies, such as DOLCE (Gangemi et al. 2002), SIMPLE (Lenci 2000), SUMO (Niles and Pease 2001), etc. These metaconcepts are distributed in three subontologies: #ENTITY, #EVENT, and #QUALITY.

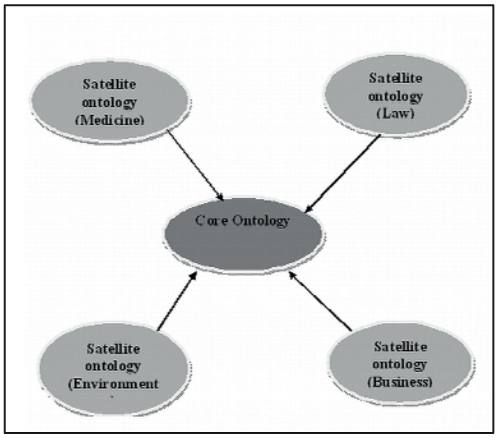

Figure 4. Domain-specific satellite ontologies

Concepts at the middle level are marked by + (e.g. +BOOK_00). These concepts are used in the meaning postulates that define basic and terminal concepts, and also encode the selection restrictions in thematic frames. The third level is composed of terminal concepts, marked by $ (e.g. $METEORITE_00). The difference between basic and terminal concepts is that basic concepts are used to define other concepts in meaning postulates, whereas terminal concepts are not. Evidently, in the satellite ontologies for specialized knowledge, terminal concepts in FunGramKB will have to be extended.

The ontology is grounded on a spiral model, where conceptual promotion and demotion can occur between the basic and terminal levels, as shown in Figure 5 (Periñan-Pascual and Arcos-Tuñez 2010).

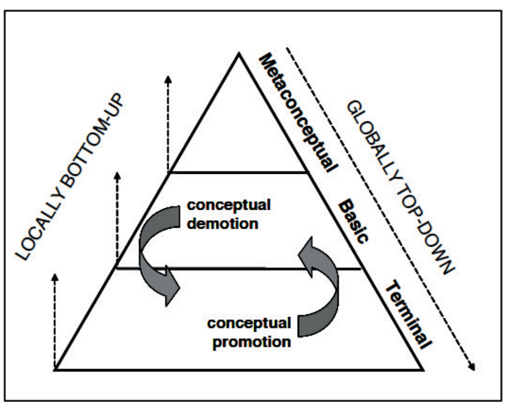

Figure 5. Conceptual promotion and demotion

Terminal concepts can thus become basic concepts when the inclusion of a new language or in this case, a more specialized conceptual content, demands a more specific world model. Promotion and demotion are organizational principles resulting from the factorization of concept definitions. Inversely, basic concepts can be demoted to terminal concepts in the case that they are not used to describe other concepts. However, the metaconceptual level always remains stable.

The inclusion of a domain-specific ontology entails both conceptual demotion and promotion. Evidently certain basic concepts in FunGramKB, such as +BALL_00, (used to define the terminal concept, $FOOTBALL_00) are demoted to terminal concepts or excluded altogether since they are not relevant. In contrast, $METEORITE_00, which is a terminal concept in FunGramKB, is promoted to a basic concept in EcoLexicon since it defines more specific meteorite types, such as CHRONDITIC_METEORITE and ACHRONDITIC_METEORITE. These meteorite types are also basic concepts since they are used to define more specific subtypes. When a larger inventory of concept profiles are available, an algorithm will be designed that will make the process automatic.

3.2 Meaning Postulates

In FunGramKB each basic and terminal concept is related to a thematic frame and is described in terms of meaning postulates, in other words, a set of one or more logically connected predications (e_1, e_2... e_n) (Mairal and Periñan-Pascual 2009). For example, the codification in terms of meaning postulates of the basic concept of $METEORITE (natural object of stone or metal that falls from space to the earth) is the following:

(1) $METEORITE_00+(e1: +BE_00 (x1: $METEORITE_00) Theme (x2: +NATURAL_ OBJECT_00) Referent) *(e2: +BE_01 (x1)Theme (x3: +STONE_00 +METAL_00) Attribute) *(e3: +DESCEND_00 (x4)Agent (x1)Theme (x5: +SPACE_01)Location (x6) Origin (x7: +WORLD_00)Goal)

Regarding processes, basic concepts in FunGramKB can be used in specialized knowledge frames. For instance, the representation of +LEAVE_00 (2) in the dimension of #MOTION is the following:

(2) + (e1: +MOVE_00 (x1) Agent (x2) Theme (x3) Location (x4) Origin (x5) Goal (f1: (e2: +BE_02 (x2) Theme (x4) Location (f2: +IN_00) Position)) Condition (f3: (e3: +BE_02 (x5) Theme (x4) Location (f4: +OUT_00) Position)) Condition)

This representation is composed of three events (e1, e2, and e3) (Mairal and Periñan 2009). In the first event, an Agent (x1) causes another entity (x3) with the role of Theme) to move from an Origin (x4) to a Goal (x5), provided two other events occur. The second event states that the Theme should be located at the Origin, whereas the third event states that the Goal (x5) should be located outside the Origin.

In environmental knowledge, this basic predication can be applied to processes such as SEDIMENT_TRANSPORT in which an ocean current causes sand (sediment) to move from one point to another, provided that the sand is at a certain location and that the place where the sand is carried is not at the original location. As can be seen, the basic difference in this case between general and specialized knowledge is the fact that the semantic arguments are more specific and are entities belonging to the subject domain. The process itself remains the same.

4. Knowledge Integration

EcoLexicon and FunGramKB, have many similarities. Both share the same upper conceptual level. They also use language as a mirror of conceptual structure, and include predicate-argument structures in the description of linguistic concepts. Both give a central role to the meaning definition of concepts though there is a difference in the semantic depth of their representations.

In EcoLexicon, conceptual descriptions are derived from surface semantics. The definition of concepts is a short text, also encoded as a set of propositions that reflect conceptual relations. A semantic network based on these propositions visually represents concept meaning.

In contrast, FunGramKB has a deep semantic representation that better exploits the cognitive content of lexical units, particularly when the definition requires the use of defining concepts, which do not directly modify the genus of the definition, but rather qualify neighboring concepts in the meaning postulate (Periñan-Pascual and Arcas-Tuñez 2007). Unlike surface semantics, deep semantics is able to represent phenomena such as aspectuality, temporality, and modality, and is better suited for reasoning and inference making.

In EcoLexicon, the exclusion of this type of meaning has posed a problem in the description of concepts, especially in the case of natural and artificial environmental processes, which should include aspectuality and temporality. For this reason, the conversion of the surface semantic representations in EcoLexicon to deep semantic representations helps to reduce the excessive information generated for general concepts (e.g. WATER, SEDIMENT, EARTH, and SEA). It also makes conceptual descriptions more fine-grained.

For example, WEATHERING is a process involving the decomposition or breakdown of rocks and minerals at and just below the Earth's surface, caused by the action of atmospheric agents, water, and living things. This concept encodes a process, which is initiated by natural forces, occurs in time and space, and affects natural entities. It is represented in EcoLexicon as shown in Figure 6.

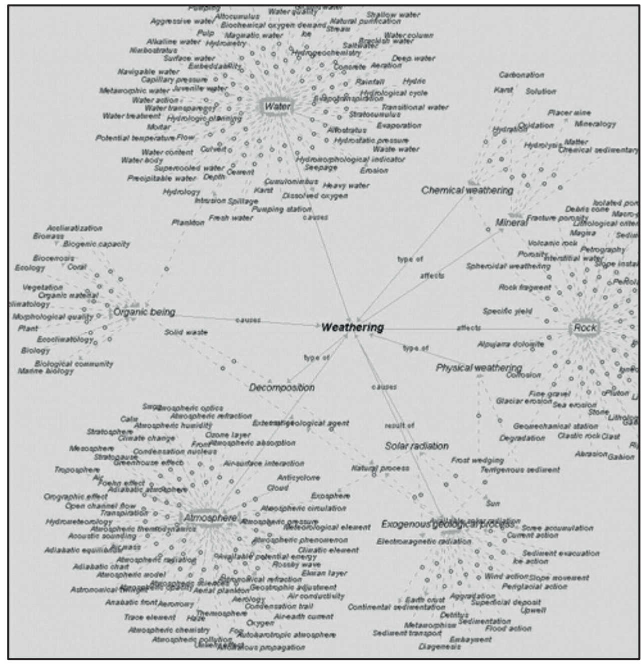

Figure 6. EcoLexicon representation of WEATHERING

This semantic network is based on the following propositions: (i) Weathering TYPE_OF Decomposition; (ii) Weathering AFFECTS Rock; (iii) Weathering AFFECTS Mineral; (iv) Atmosphere CAUSES Weathering; (v) Water CAUSES Weathering; (vi) Solar radiation CAUSES Weathering; (vii) Organic being CAUSES Weathering; (viii) Mechanical weathering TYPE_OF Weathering; (ix) Chemical weathering TYPE_OF Weathering. Nevertheless, this representation does not allow the codification of important information regarding typical temporality or the subprocesses by which water, solar radiation, and/or living organisms cause weathering. A deep semantic representation makes it possible to use basic concepts to encode the events that take place as part of the process, and establish their temporal order. The basic act of decomposition is thus codified as follows in terms of two events.

(3) +(e1: +CHANGE_00 (x1: WATER ^ SOLAR RADIATION ^ ORGANIC BEING) Theme (x2: ROCK ^ MINERAL) Referent (f1: (e2: +MOVE_00 (x2) Theme (x3: +APART_00) Attribute))Result)

In this formulation, there is a first event (e1) in which an entity (x1) that is water, solar radiation or an organic being (Theme) effects a change in a rock or mineral (Referent). In the second event (e2), which codifies the result, the rock or mineral moves apart or separates. The result may involve a chemical change (as in the case of chemical weathering) or not (as in physical weathering). FunGramKB also makes it possible to codify scripts for mechanical and chemical weathering processes and reactions, such as freeze-thaw, salt wedging, oxidation, hydrolysis, and acidification.

5. Conclusions

Currently, most lexical ontologies adopt a relational approach to meaning representation since conceptual meaning is difficult to express in its entirety. However, a relational approach generates a very partial representation, and excludes crucial information regarding the aspectuality and temporality of concepts. In contrast, a deep semantic representation has greater expressive power, despite the fact that it is more laborious and complicated to encode. The advantages of this trade-off have been reflected and exemplified in the description of specialized environmental concepts have been formalized and linked to the general knowledge in FunGramKB.

Acknowledgements

This research was funded by the Spanish Ministry of Science and Innovation (FFI2008-06080-C03-01 and FFI2008-05035-C02-01).

References

Barsalou, L. W. 2008. Grounded Cognition. *Annual Review of Psychology* 59: 617–645.

Buitelaar, P.; Cimiano, P.; Haase, P.; and Sintek, M. 2009. Towards Linguistically Grounded Ontologies. In *Proceedings of the 6th European Semantic Web Conference.* 111-125. Berlin: Springer.

Eriksson, H. 2007. The Semantic Document Approach to Combining Documents and Ontologies. *International Journal of Human-Computer Studies* 65: 624–639.

Evans, V.; Bergen, B. K; and Zinken, J. 2007. The Cognitive Linguistics Enterprise: an Overview. In *The Cognitive Linguistics Reader.* 2-36. London: Equinox.

Faber, P.; León, P.; Prieto, J; and Reimerink A. 2007. Linking Images and Words: the Description of Specialized Concepts. International Journal of Lexicography 20: 39-65.

Faber, P., C. Márquez and M. Vega. 2005. Framing Terminology: A Process-Oriented Approach. *META* 50(4). CD-Rom.

Faber, P.; Montero, S.; Castro, M. R.; Senso, J.; Prieto, J. A.; León, P.; Márquez, C.; and Vega, M. 2006. Process-oriented Terminology Management in the Domain of Coastal Engineering. *Terminology* 12(2): 189–213.

Fillmore, C. 1985. Frames and the Semantics of Understanding. Quaderni di Semantica 6: 222-254.

Gahegan, M.; Agrawal, R.; Jaiswal, A.; Luo, J.; and Soon, K. H. 2008. A Platform for Visualizing and Experimenting with Measures of Semantic Similarity in Ontologies and Concept Maps. *Transactions in GIS* 12(6): 713–732.

Garcia Castro, A.; Rocca-Serra1, P.; Stevens, R.; Taylor, C.; Nashar, K.; Ragan, M.; and Sansone, S. 2006. The Use of Concept Maps during Knowledge Elicitation in Ontology Development Processes: the Nutrigenomics Use Case. *BMC Bioinformatics* 7:267. doi:10.1186/1471-2105-7-267.

Gangemi, A.; Guarino, N.; Masolo, C.; Oltramari, A.; and Schneider, L. 2002. Sweetening Ontologies with DOLCE. In *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management*, London: Springer.

Hsieh. S.; Lin, H. T.; Chi, N. W.; Chou, K. W.; and Lin, K. Y.; 2010. Enabling the Development of Base Domain Ontology through Extraction of Knowledge from Engineering Domain-Handbooks. *Advanced Engineering Informatics*. doi:10.1016/j.aei.2010.08.004.

Lenci, A. 2000. Building an Ontology for the Lexicon: Semantic Types and Word Meaning. In *Ontology-Based Interpretation of Noun Phrases.* 103-120. Kolding: University of Southern Denmark Press.

Leon, P.; Magaña, P.; and Faber, P. 2009. Building the SISE: an Environmental Ontology. In Proceedings of the EU Conference Towards eEnvironment. Available at: http:/www.e-envi2009.org/proceedings/

Leon, P. and Magaña, P. 2010. EcoLexicon: Contextualizing an Environmental Ontology. In Proceedings of the Terminology and Knowledge Engineering Conference (TKE), Dublin, Ireland. CD-Rom.

Mairal R. and Periñan-Pascual, C. 2009. The Anatomy of the Lexicon within the Framework of an NLP Knowledge Database. *Revista Española de Lingüística Aplicada* 22: 217-244

Montiel-Ponsoda, E.; Aguado de Cea, G.; Gómez-Pérez, A.; and Peters, W. 2010. Enriching Ontologies with Multilingual Information. *Natural Language Engineering.* doi: 10.1017/ S1351324910000082

Niles, I. and Pease, A. 2001. Towards a Standard Upper Ontology. In *Proceedings of the International conference on Formal Ontology in Information Systems*. 2-9. New York: ACM.

Periñán-Pascual, C. and Arcas-Túnez, F. 2007. Deep Semantics in an NLP Knowledge Base. In *Proceedings of the 12th Conference of the Spanish Association for Artificial Intelligence*. 279-288. Salamanca: University of Salamanca.

Periñan-Pascual, C. and Arcas-Túnez, F. 2010. The Architecture of FunGram KB. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation.* 2667-2674. Valetta, European Language Resources Association.

Peters, S. and Shrobe, H. 2003. Using Semantic Networks for Knowledge Representation in an Intelligent Environment. In Proceedings of the First IEEE International Conference on Pervasive Computing and Communications. 323-337. Washington D. C., IEEE Computer Society.

Raskin, R. and Pan, M.J., 2005. Knowledge Representation in the Semantic Web for Earth and Environmental Terminology (SWEET). In *Geosciences, Computers & Geosciences* 31: 1119–1125.

Samwald, M.; Chen, H.; Ruttenberg, A.; Lim, E.; Marenco, L.; Miller, P.; Shepherd, G.; andCheung, K. H.2010. Semantic SenseLab: Implementing the Vision of the Semantic Web in Neuroscience. Artificial Intelligence in Medicine 48: 21–28.

Sowa, J. F. 2000. *Knowledge Representation: Logical, Philosophical and Computational Foundatio*ns. Pacific Grove, CA: Brooks/Cole.

Talmy, L. 2000. *Toward a Cognitive Semantics*, Cambridge, MA: MIT Press.