EcoLexicon and FunGramKB: Applying COREL to Domain-Specific Knowledge

Pilar León Araúz and Arianne Reimerink

Department of Translation and Interpreting (University of Granada), Buensuceso 11, Granada, Spain
{pleon, arianne}@ugr.es

Abstract

EcoLexicon is a multilingual terminological knowledge base (TKB) on the environment. It is currently being converted into a domain-specific ontology, however, ontological properties are modelled according to surface semantics. For this reason, we are integrating our TKB in the form of a "satellite ontology" into FunGramKB, a multipurpose knowledge base specifically designed for natural language understanding. We explain how the dynamism of environmental concepts can benefit from a formal description in meaning postulates and their inclusion in FunGramKB Cognicon scripts. This would lead to the automatic generation of flexible conceptual networks and definitional templates across different contexts.

Introduction

EcoLexicon (http://ecolexicon.ugr.es) is a multilingual terminological knowledge base (TKB) on the environment. It seeks to meet both cognitive and communicative needs of different users, such as translators, technical writers or even environmental experts. Each entry provides a great amount of interrelated information, such as linguistic contexts, images, definitions and conceptual networks (León et al. 2008, Reimerink et al. 2010).

The information is stored in a relational database (RDB), which has been very useful for the quick deployment of the TKB, but this system does not allow for Natural Language Processing purposes. The conceptual information in EcoLexicon is represented according to the relations that exist among the concepts of the domain: hierarchical relations, such as *part_of* and *type_of*, and non-hierarchical dynamic relations, such as *causes*, *result_of*, *has_function*, and *located_at*. This type of surface semantics does not work for complex NLP tasks.

The EcoLexicon RDB is currently being converted into a domain-specific ontology (León et al. 2009), however, ontological properties are still modelled according to surface semantics. For this reason, we are integrating our TKB in the form of a "satellite ontology" into 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). FunGramKB is especially appropriate for our purposes as it has the same top level categories in the ontology (entity, event and quality), and both DBs are based on the extraction of knowledge from language resources. Moreover, all subsystems of FunGramKB are represented through the same formal language, COREL, which is essential for successful inferencing in NLP systems (Periñan-Pascual and Arcas-Túnez 2007a). This formalism provides the kind of deep semantics that EcoLexicon lacks. Apart from the above, EcoLexicon has integrated qualia structure (Pustejovsky 1995) into conceptual relations and definitional templates, whereas previous projects related to FunGramKB have also attempted to use qualia in their formalism (Reimerink et al. 2010, Mairal and Ruiz de Mendoza 2008).

In this paper, we will focus on a first approach to the integration of complex events of the environmental domain into FunGramKB's Cognicon subsystem, which stores procedural knowledge by means of conceptual schemata. In sections 2 and 3, the conceptual structure of both EcoLexicon and FunGramKB are addressed. In section 4, we explain how qualia structure can be applied to the COREL formalism of SEDIMENT and how the complex environmental event WATER TREATMENT can be formalized in COREL and integrated into the Cognicon.

Conceptual structure in EcoLexicon

Broadly speaking, the conceptual structure of EcoLexicon is modelled according to surface semantics based on qualia roles and their combinatorial potential as well as multidimensionality and the dynamic effects of context.

Conceptual relations and definitions

To ensure internal coherence in EcoLexicon, Pustejovsky's qualia structure (Pustejovsky 1995, Pustejovsky et al. 2006) is applied both to the categorization of conceptual relations and the development of definitional templates (León et al. 2008, Reimerink et al. 2010). Pustejovsky describes lexical items according to their qualia structure, which expresses the componential aspect of a word's meaning and is considered the meeting point of both argument and event structure. This is composed of the following roles: (1) formal role: the basic type distinguishing the meaning of a word; (2) constitutive role: the relation between an object and its constituent parts; (3) telic role: the purpose or function of the object, if there is one; (4) agentive role: the factors involved in the object's origins or "coming into being" (Pustejovsky et al. 2006: 3).

In EcoLexicon, the activation of each qualia role constrains the description of concepts according to their nature. For example, entities can only be related through the constitutive role if the second concept of the proposition is another entity. However, each qualia role can be codified by different relations. For instance, the constitutive role has been split up into six different relations, since not all parts interact in the same way with their wholes. This distinction is based on our domain-specific needs, mainly on ontology reasoning and consistency. For example, if *located_in* were considered as a *part_of* relation, that would cause fallacious transitivity. If a GABION is *part_of* a GROYNE and a GROYNE *part_of* the SEA, an ontology would infer that GABIONS are *part_of* the SEA, which is not a plausible example.

The definitions in EcoLexicon are developed according to qualia-based definitional templates. As shown in Figure 1, the definitional template of the natural entity SEDIMENT includes information regarding the material it is made of, the entities that are made of it, how it is created and how coastal engineers use it. The formal role expresses category membership, and the constitutive role shows that COASTAL LANDFORMS are made of SEDIMENT. The agentive role explains that SEDIMENT results from four processes: WEATHERING, TRANSPORT, DEPOSITION and WASTEWATER TREATMENT. The telic role, however, shows that engineers make specific use of SEDIMENT to nourish BEACHES to protect them from EROSION. These are prototypical features of SEDIMENT from a domain-specific point of view, since a general language definition would rarely include the telic or agentive role.

Sediment		
Formal	*Is_a*	Material
Constitutive	*Made_of*	➢ Clay ➢ Sand ➢ Mud
	Material_of	Coastal landform ➢ Beach ➢ Delta
Telic	*Has_function*	Beach nourishment
Agentive	*Result_of*	Weathering Transport Deposition Water treatment

Figure 1. Definitional template of SEDIMENT

Contextual domains

In knowledge modelling, concepts are very often classified according to very different dimensions (shape, function, colour, etc.). Multidimensionality (Kageura 1997, Rogers 2004) is commonly regarded as a way of enriching traditional static representations, enhancing knowledge acquisition through different points of view in the same conceptual network (León Araúz and Faber 2010). In such a wide domain as the environment, multidimensionality increases the number of possible relations activated by specialized concepts, since the same concept may act as an agent or a patient, as an active process or a result. However, the environmental domain has caused a great deal of information overload, which ends up jeopardizing knowledge acquisition (Figure 2).

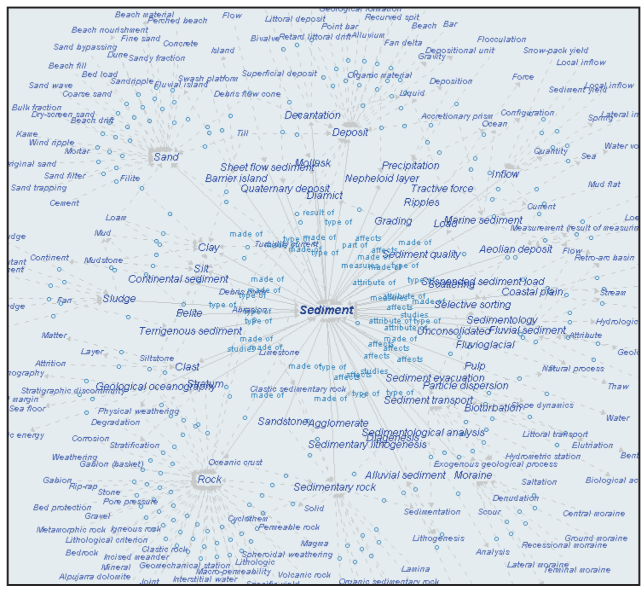

Figure 2. Overloaded conceptual network of SEDIMENT

Actually, multiple dimensions are not always compatible but context-dependent. Their activation generally depends on perspective. For instance, SEDIMENT is a highly context-sensitive concept. It can be regarded as the *result_of* natural (WEATHERING, TRANSPORT, etc.) or artificial processes (DECANTATION), but in the real world, it will rarely, if ever, activate these concepts at the same time.

Any specialized domain reflects different situations in which certain conceptual dimensions become more or less salient. Background situations constrain conceptual processing in many tasks (e.g., recall, recognition, categorization, lexical decision, etc.) across many areas of cognition (Yeh and Barsalou 2006). However, concepts also have their own situated nature. In this respect, Barsalou (2009:

1283) states that a concept itself produces a wide variety of situated conceptualizations that support goal achievement in specific contexts. The environmental field has thus been divided into different contextual domains according to corpus information and expert collaboration: HYDROLOGY, GEOLOGY, METEOROLOGY, BIOLOGY, CHEMISTRY, ENGINEERING, WATER TREATMENT, COASTAL PROCESSES and NAVIGATION.

These domains have been allocated in a similar way as the General European Multilingual Environmental Thesaurus, whose structure is based on themes and descriptors, reflecting a systematic, category or discipline-oriented perspective (GEMET 2004). They can also be related to the notion of Guha's (1991) micro-theories. In this way, our contextual domains provide the clues to simplify the background situations in which concepts can occur in reality (León and Reimerink 2010). Domain-based constraints are neither applied to individual concepts nor to individual relations. They are instead applied to conceptual propositions (León Araúz et al. 2009). For instance, SEDIMENT is linked to DECANTATION through a *result_of* relation, but this proposition is irrelevant if users only want to know how SEDIMENT is naturally created or how it serves as BEACH FILL in ARTIFICIAL NOURISHMENT projects. Consequently, the proposition SEDIMENT *result_of* DECANTATION will only appear in a WATER TREATMENT context. As a result, when constraints are applied, SEDIMENT only shows relevant dimensions for each contextual domain (Figure 3).

The concept itself does not change, but rather behaves very differently depending on its relations with other concepts in each contextual domain. The recontextualization of conceptual networks disambiguates the situation in which concepts may occur, since the specification of context narrows down the multidimensional relational power of concepts. In the WATER TREATMENT domain, SLUDGE, and not ROCKS, is the main SEDIMENT subtype and no geological processes are involved, but different artificial events, such as DECANTATION, SCATTERING or PRECIPITATION.

FunGramKB

FunGramKB is structured in three information levels: the lexical level, the grammatical level and the cognitive level. In the lexical level, the lexicon and the morphicon store morphosyntactic, pragmatic and collocational information. The grammatical level, the Grammaticon, stores constructional schemata. The conceptual level consists of three subsystems. 1) the Ontology, consisting of a general-purpose module (core ontology) and several domain-specific modules (satellite ontologies) that store semantic information 2) the Cognicon, which stores procedural knowledge, which takes into account temporal continuity through conceptual schemata, and 3) the Onomasticon, which stores episodic knowledge about instances of entities (Periñan-Pascual and Arcas-Túnez 2007a, 2010).

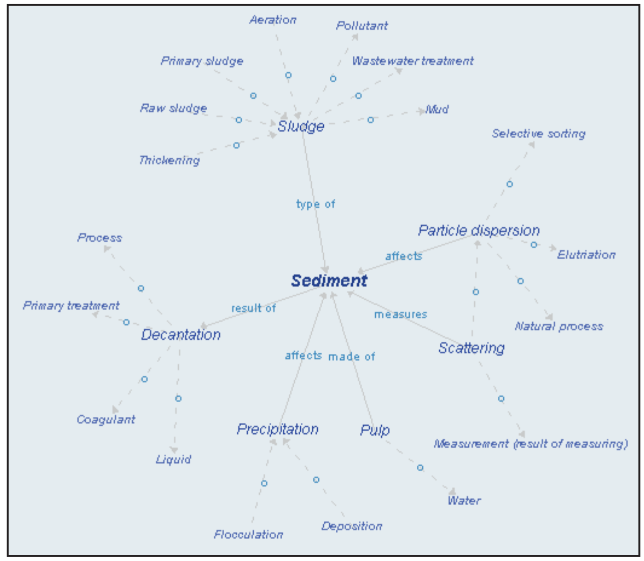

Figure 3. SEDIMENT in the WATER TREATMENT domain

Conceptual level of FunGramKB

The subdivision of the conceptual level is based on the combination of protypicality and temporality, which results in a typology of four different conceptual schemata:

1) proto-microstructures (meaning postulates): prototypical knowledge, which does not take into account the time factor, such as the conceptual representation of SONG;
2) proto-macrostructures (scripts): prototypical knowledge that implies the passage of time, such as the description of EATING AT A RESTAURANT;
3) bio-microstructures (snapshots): instances of entities, such as the Eiffel Tower;
4) bio-macrostructures (stories): for example, the construction of the Eiffel Tower.

FunGramKB is therefore in consonance with studies on long-term memory in the human brain, which classify memory into semantic memory, procedural memory and episodic memory. According to Tulving (1985), long-term memory components do not work in an isolated way but they interact with each other in order to facilitate information storage and retrieval. Therefore, the FunGramKB researchers insist that it is essential for successful inferencing in an NLP system that all these knowledge schemata are represented through the same formal language, so that information sharing can take place effectively among all conceptual modules (Periñan-Pascual and Arcas-Túnez 2010: 2668). This language is called COREL, COnceptual REpresentation Language, and it describes each concept with a set of one or more logically connected predications which carry the generic features of the concept (Mairal and Periñan-Pascual 2009). This representational language consists of: (1) a collection of conceptual units that allow for hierarchical inheritance enhancing definitional structures and thus increasing semantic expressiveness and (2) an annotation syntax for interlinguistic representations.

FunGramKB classifies conceptual units in three different types:

1) Metaconcepts (e.g. #abstract, #emotion), which belong to the top level and and are distributed in three subontologies: #entity, #event, and #quality;
2) Basic concepts (e.g. +hand_00, +forget_00), which are used in the meaning postulates that define basic concepts and terminal concepts, and also encode the selection restrictions in thematic frames;
3) Terminal concepts (e.g. $skyscraper_00, $varnish_00), which are not used to define other concepts in meaning postulates.

Metaconcepts are always at the top level of the ontology. Basic concepts and terminal concepts, however, can be promoted or demoted according to specific needs. A satellite ontology on a specialized domain such as the environment will have to promote certain terminal concepts of the Core Ontology to basic concepts, and the number of terminal concepts will be increased.

The Cognicon

According to Periñán-Pascual and Arcas-Túnez (2007b), meaning postulates are not sufficient to describe commonsense knowledge, but they contribute actively to build cognitive macrostructures in the Cognicon. FunGramKB thus integrates semantic knowledge from the ontology with procedural knowledge from the Cognicon.

In the Cognicon, proto-macrostructures, called scripts, are structured into one or more predications within a linear temporal framework, based on Allen's interval temporal model (1983).

In (1) the first nine predications of the proto-macrostructure @EATING_AT_RESTAURANTS the COREL formalism is illustrated (Periñán-Pascual and Arcas-Túnez 2010: 2669). The first meaning postulate describes that a CUSTOMER (x_1, Agent) enters the RESTAURANT (x_4, Location) in event 1, because s/he is HUNGRY (x_5, Attribute), which is event 2. Event 5 describes how the WAITER (x_6) brings the MENU and/or the WINE LIST (x_{10}, Theme) to the TABLE (x_9). In FunGramKB, each predication in a meaning postulate is preceded by a reasoning operator to state if the predication is strict (+) or defeasable (*). The first event is preceded by +, which means that the meaning postulate is obligatory for the script. Event 5, on the other hand, is optional (represented by *) because the waiter does not always bring the menu to the table. Concepts such as CUSTOMER or WAITER are considered basic concepts in the FunGramKB ontology and are therefore used in the meaning postulates of other concepts, which is shown by the + symbol, whereas MENU is a terminal concept, marked by $.

(1)
@EATING_AT_RESTAURANTS
*(e_1: +ENTER_00 (x_1: +CUSTOMER_00)$_{Agent}$ (x_1)$_{Theme}$ (x_2)$_{Location}$ (x_3)$_{Origin}$ (x_4: +RESTAURANT_00)$_{Goal}$ (f_1: (e_2: +BE_01 (x_1)$_{Theme}$ (x_5: +HUNGRY_00)$_{Attribute}$))$_{Reason}$)
*(e_3: $ACCOMPANY_00 ($x_6$: +WAITER_00)$_{Agent}$ (x_6)$_{Theme}$ (x_7)$_{Location}$ (x_8)$_{Origin}$ (x_9: +TABLE_00)$_{Goal}$)
*(e_4: +SIT_00 (x_1)$_{Theme}$ (x_9)$_{Location}$)
*(e_5: +TAKE_01 (x_6)$_{Agent}$ (x_{10}: $MENU_00 | WINE_LIST_00)$_{Theme}$ (x_{11})$_{Location}$ x_{12})$_{Origin}$ (x_9)$_{Goal}$)
*(e_6: +REQUEST_01 (x_1)$_{Theme}$ (x_{13}: +FOOD_00 +BEVERAGE_00)$_{Referent}$ (x_6)$_{Goal}$)
+(e_7: +SAY_00 (x_6)$_{Theme}$ (x_{14}: (e_8: +COOK_00 (x_{15}: $COOK_D_00)$_{Theme}$($x_{16}$:+FOOD_00)$_{Referent}$))$_{Referent}$ (x_{15})$_{Goal}$)
*(e_9: +TAKE_01 (x_6)$_{Agent}$ (x_{17}: +BEVERAGE_00)$_{Theme}$ (x_{18})$_{Location}$ (x_{19}: $BAR_00)$_{Origin}$ (x_9)$_{Goal}$)

Applying COREL to EcoLexicon

The earlier definitional template of SEDIMENT (figure 1) can be formalized in the following meaning postulate (2):

(2)
+SEDIMENT
+[(e_1: +BE_00 (x_1: +SEDIMENT_00)$_{Theme}$ (x_2: +MATERIAL_00) $_{Referent}$ (f_1: e_2: +BE_01 (x_1)$_{Theme}$ (x_3: $WEATHERED_00, $TRANSPORTED_00, $DEPOSITED_00)$_{Attribute}$))$_{Result}$) Q_A] Q_F
+[(e_3: +BE_02 (x_1) $_{Theme}$ (x_4: +CLAY_00 | +SAND_00 | +MUD_00) $_{Attribute}$] Q_C
*[(e_4: +CONTAIN_00 (x_1)$_{Theme}$ (x_5: +COASTAL LANDFORM)$_{Location}$] Q_C
*[(e_5: $DREDGE_00 ($x_6$: $DREDGER_00 ^ +HUMAN_00)$_{Agent}$ (x_1)$_{Theme}$ (x_5: +COASTAL LANDFORM_00)$_{Origin}$ (x_7: +COASTAL LANDFORM_01)$_{Goal}$) (f_2: (e_6: $NOURISH_00 ($x_1$)$_{Theme}$ (x_7)$_{Referent}$))$_{Purpose}$]Q_T
*[(e_7: +REMOVE_00 (x_8: +GRAVITY)$_{Agent}$ (x_1)$_{Theme}$ (x_9: $SEDIMENTATION TANK)$_{Location}$ (x_{10}: $WASTEWATER)$_{Origin}$ (x_{11}:$FILTER_00)$_{Goal}$ (f_3: e_8: BE_01 (x_1)$_{Theme}$ (x_{12}: $DECANTED_00)$_{Attribute}$))$_{Result}$)]$Q_A$

The first two predications are preceded by + because they are prototypical in our specialized domain (a SEDIMENT (x_1) is a *type_of* MATERIAL (x_2) that can be *made_of* CLAY, SAND or MUD (x_4) or a combination). They correspond to the formal and constitutive roles, although the formal role includes an agentive one as a satellite, since materials that are not the result of WEATHERING, TRANSPORT and DEPOSITION (x_3) cannot be regarded as SEDIMENTS. In the earlier definitional template, the agentive role appeared as a separate module from the formal one. However, as opposed to surface semantics, this representation allows us to combine different relations in the form of conceptual chains, where the relations *is_a* and *result_of* are interdependent.

The following predication is preceded by *, as it is not a prototypical one, but the proposition COASTAL LANDFORM (x_5) *made_of* SEDIMENT is especially prevalent in our domain. However, the telic and agentive roles are also preceded by * but are only related to specific contextual domains and scripts. In the COASTAL PROCESSES domain, SEDIMENTS are DREDGED from a COASTAL LANDFORM to another in order to NOURISH (e_6) it, whereas in the WATER TREATMENT domain, SEDIMENTS are DECANTED (x_{12}) from WASTEWATER (x_{10}) in a SEDIMENTATION TANK (x_9).

In the following meaning postulates (3), SEDIMENT (x_{15}) is included and properly contextualized in the script @TREATING_WASTEWATER.

(3)
@ TREATING_WASTEWATER
+(e_1: +ENTER_00 (x_1: $WASTEWATER_00)$_{Agent}$ (x_1)Theme (x_2)$_{Location}$ (x_3: $SEWER_00)$_{Origin}$ (x_4: $WASTEWATER TREATMENT PLANT_00)$_{Goal}$ (f_1: (e_2: +BE_01(x_1)$_{Theme}$ (x_5: $GREY_00 ^ $BLACK_00)$_{Attribute}$))$_{Reason}$)
*@SCREENING
+(e_3: +REMOVE_00 (x_6: $BAR SCREEN_00)$_{Agent}$ (x_7: +SOLID_00)$_{Theme}$ (x_8)$_{Location}$ (x_1: $WASTEWATER_00)$_{Origin}$ (x_9: LANDFILL_00)$_{Goal}$ (f_2: (e_4: +BE_00 (x_7: +SOLID_00)$_{Theme}$ (x_{10}: +BIG_01)$_{Attribute}$))
*@GRIT_REMOVAL
+(e_5: +REMOVE_00 (x_{11}: +GRAVITY_00)$_{Agent}$ (x_{12}: GRIT_00 | STONE_00 | GLASS_00)$_{Theme}$ (x_{13}: $GRIT CHAMBER_00)$_{Location}$ (x_1: WASTEWATER_00)$_{Origin}$ (x_{14})$_{Goal}$
*@DECANTATION
(e_6: +REMOVE_00 (x_{11}: +GRAVITY_00)$_{Agent}$ (x_{15}: **$SEDIMENT_00**)$_{Theme}$ (x_{16}: $SEDIMENTATION TANK)$_{Location}$ (x_1: $WASTEWATER)$_{Origin}$ (x_{17}: $FILTER_00)$_{Goal}$
*@FAT_AND_GREASE REMOVAL
+(e_7: +REMOVE_00 (x_{18}: SKIMMER_00 | $AIR BLOWER_00)$_{Agent}$ (x_{19}: FAT_00 | GREASE_00)$_{Theme}$ (x_{16}: + SEDIMENTATION TANK_00) $_{Location}$ (x_1: WASTEWATER_00)$_{Origin}$ (x_{20})$_{Goal}$
*@FILTERING
+(e_8: +REMOVE_00 (x_{21}: +SAND_00 | +CARBON_00)$_{Agent}$ (x_{22}: $PARTICULATE MATTER)$_{Theme}$ (x23)$_{Location}$ (x_1: $WASTEWATER_00)$_{Origin}$ (x_{24}: $FILTER MEDIUM_00)$_{Goal}$ (f_3: (e_9: +BE_00 (x_1: $WASTEWATER_00)$_{Theme}$ (x_{25}: $FILTERED_00)$_{Attribute}$))$_{Result}$)
*@LAGOONING
+(e_{10}: +REMOVE_00 (x_{26}: +ORGANISM_00)$_{Agent}$ (x_{22}: $PARTICULATE MATTER_00)$_{Theme}$ (x_{27}: $LAGOON_00)$_{Location}$ (x_1: $WASTEWATER_00)$_{Origin}$ (x_{28})$_{Goal}$ (f_4: (e_{11}: +BE_00 (x_1: $WASTEWATER_00)$_{Theme}$ (x_{29}: $LAGOONED_00)$_{Attribute}$))$_{Result}$)
*@NUTRIENT_REMOVAL
+e_{12}: +REMOVE_00 (x_{30}: $BACTERIA_00)$_{Agent}$ (x_{31}: $PHOSPHORUS_00 | NITROGEN_00)$_{Theme}$ (x_{32})$_{Location}$ (x_1: $WASTEWATER)$_{Origin}$
*@DESINFECTION
+e_{13}: +REMOVE_00 (x_{33}: $CHLORINE_00 | $OZONE_00)$_{Agent}$ (x_{34}: $MICROORGANISM_00)$_{Theme}$ (x_{35})$_{Location}$ (x_1: $WASTEWATER_00)$_{Origin}$ (x_{36})$_{Goal}$ (f_5: (e_{14}:+BE_00 (x_1: $WASTEWATER_00)$_{Theme}$ (x_{37}: $DESINFECTED_00)$_{Attribute}$))$_{Result}$)
+(e_{15}: +BECOME_00 (x_1: $WASTEWATER_00) $_{Theme}$ (x_{38}: +EFFLUENT_00) $_{Referent}$ (f_6: (e_{16}: @SCREENING | @GRIT_REMOVAL | @DECANTATION | @FAT_AND_GREASE REMOVAL | @FILTERING | @LAGOONING | @NUTRIENT_REMOVAL | @DESINFECTION))$_{Condition}$
+(e_{17}: +DISCHARGE_00 (x_{39})$_{Agent}$ (x_{40}: +EFFLUENT_00)$_{Theme}$ (x_4: $WASTEWATER TREATMENT PLANT_00)$_{Origin}$ (x_{41}: +SEA_00 | +RIVER_00 | +LAKE_00 | +GROUND_00)$_{Goal}$

This script is composed of several sub-scripts, such as @FAT_AND_GREASE_REMOVAL, @SCREENING, @GRIT_REMOVAL, @DECANTATION, etc. Most of them are represented through the predicate REMOVE, which requires different thematic roles, such as agent, theme, location, origin and goal. In the first predication, WASTEWATER (x_1) enters the TREATMENT PLANT (x_4) Then, the sub-scripts codify all the processes experienced by WASTEWATER, where different agents (BAR SCREENS, GRAVITY, BACTERIA, OZONE, etc.) remove different materials (GRIT, GREASE, SEDIMENT, MICROORGANISMS, etc.) from WASTEWATER. Once all these processes are completed, the concept WASTEWATER becomes EFFLUENT (e_{16}), which is discharged from the TREATMENT PLANT (x_4) to the ENVIRONMENT (x_{41}). All main predications are preceded by + because the process is representative enough in the contextual domain of WATER TREATMENT. However, all sub-processes are preceded by * because not all of them are always applied to WASTEWATER.

In FunGramKB, nuclear features for the elaboration of meaning postulates are established according to the notions of typicality and majority (Periñán-Pascual and Mairal Usón 2010). For example, a bird should be described as a vertebrate (+) that has many feathers (*) and flies (*). The first feature is always shared by all of its exemplars whereas the last two features are only shared by most of them. However, another feature such as "birds sing" is not included in the meaning postulate because it is not representative enough. In EcoLexicon, the level of representativeness is identified according to corpus data and varies across context, since domain-specific events should be described in a more detailed and dynamic way. Therefore, since dynamism is mostly encoded in non-hierarchical knowledge, the telic and agentive roles of environmental concepts are the conceptual underpinnings for the development of different scripts associated to each contextual domain.

Conclusions and future research

Since context, knowledge and reasoning are strongly related, the main aims of artificial intelligence with regards to the formalization of context seem obvious: performing automatic inferences and reasoning, identifying relational constraints for context-aware applications, improving automatic information retrieval, resolving ambiguities in natural language processing, etc.

In EcoLexicon, automatic reasoning and knowledge extraction can be improved through the application of the MicroKnowing and the Presupposition Builder modules, implemented in FunGramKB (Periñán-Pascual and Mairal Usón 2009) as a reasoning process for the construction of extended meaning postulates and the inference of background knowledge. Furthermore, FunGramKB researchers state that in contrast with the semantic knowledge repository of the Ontology, the possibility of calling a whole script within another script gives us the chance to introduce culturally-biased knowledge in the Cognicon, since every script is assigned a geographical feature determining the continent, country, etc where that knowledge is typically true. In this sense, the application of scripts to EcoLexicon would allow us to establish different levels of correspondence among our contextual domains and meaning postulates. Since contextual variations in our domain do not depend on countries or cultures, that geographical feature would be substituted by a particular contextual domain.

On the one hand, it would enhance the creation of flexible definitions upon context. If users wanted to access a prototypical definition of +SEDIMENT only formal and constitutive roles would arise. However, mapping the script @TREATING_ WATER into the meaning postulate of +SEDIMENT would allow us to convert the * predication from (2) into a + predication, highlighting that predication as a typical one in the definition.

On the other hand, the association of scripts with contextual domains would allow us to improve knowledge extraction and enhance user queries. If the different arguments of contextual scripts could be matched with our corpus, different texts could be quickly classified according to contextual domains. Furthermore, if user queries could be matched with the arguments in our scripts, conceptual networks would appear recontextualized accordingly.

Acknowledgements

This research has been carried out in projects FFI2008-06080-C03-01/FILO and FFI2008-05035-C02-01 funded by the Spanish Ministry of Science and Innovation.

References

Allen, J. 1983. Maintaining Knowledge about Temporal Intervals. In Communications of the ACM 26(11): 832-843.

Barsalou, L.W. 2009. Simulation, Situated Conceptualization and Prediction. Philosophical Transactions of the Royal Society of London, *Biological Sciences* 364: 1281-1289.

GEMET. 2004. About GEMET. General Multilingual Environmental Thesaurus.

Guha R. V. 1991. *Contexts: A Formalization and Some Applications*. Stanford PhD Thesis.

Kageura, K. 1997. Multifaceted/Multidimensional Concept Systems. In Wright, S.E. and Budin, G. eds. *Handbook of Terminology Management*: *Basic Aspects of Terminology Management*, 119-32. Amsterdam/Philadelphia: John Benjamins.

León Araúz, P. and Faber, P. 2010. Natural and Contextual Constraints for Domain-Specific Relations. In *Proceedings of Semantic relations. Theory and Applications*. Malta.

León Araúz, P. and 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.

Leon, P., Magaña, P. and Faber, P. 2009. Building the SISE: an Environmental Ontology. In *Proceedings of the EU Conference, Towards eEnvironment*. Prague, Czech Republic.

León Araúz, P., Reimerink, A, Faber, P. 2008. PuertoTerm and MarcoCosta: a Frame-Based Knowledge Base for the Environmental Domain. In *Proceedings of the XVIII FIT World Congress*. Shanghai.

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

Mairal Usón, R and Ruiz de Mendoza Ibañez, F. J. 2008. New Challenges for Lexical Representation within the Lexical-Constructional Model (LCM). *Revista Canaria de Estudios Ingleses* 57: 137-158.

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

Periñán-Pascual, C. and Arcas-Túnez, F. 2007b. Cognitive Modules of an NLP Knowledge Base for Language Understanding, 197-204. *Sociedad Española para el Procesamiento del Lenguaje Natural*.

Periñán-Pascual, C. and Arcas-Túnez, F. 2010. The Architecture of FunGram KB. In *Proceedings of the 7th LREC Conference*. Malta.

Periñán-Pascual, C. and Mairal Usón, R. 2010. La Gramática de COREL: un Lenguaje de Representación Conceptual. *Onomázein* 21:11-45.

Pustejovsky, J 1995. *The Generative Lexicon*. Cambridge, Mass: MIT Press.

Pustejovsky, J; Havasi, C; Littman, J; Rumshisky, A; Verhagen, M. 2006. Towards a Generative Lexical Resource: The Brandeis Semantic Ontology. In *Proceedings of LREC 2006*, Genoa, Italy.

Reimerink, A.; León, P.; Faber, P. 2010. A Qualia-based Description of Specialized Knowledge Units in the Lexical-Constructional Model. *Terminàlia* 1: 17-25. doi:10.2436/20.2503.01.9.

Rogers, M. 2004. Multidimensionality in Concepts Systems: a Bilingual Textual Perspective. *Terminology* 10(2): 215-240.

Tulving, E. 1985. How Many Memory Systems Are There? *American Psychologist* 40: 385-398.

Yeh, W. and Barsalou, L. W. 2006. The Situated Nature of Concepts. *American Journal of Psychology* 119: 349-384.