Is it possible to enrich ontologies with a specialized domain linguistic resource?

María A. Barrios, Luis M. Vilches-Blázquez
Ontology Engineering Group, Fac. Informática, Univ. Politécnica de Madrid

Abstract: Enriching ontologies with linguistic resources is considered an important target in natural language applications. These linguistic resources should contain not only linguistic but knowledge information. However the linguistic resources available, such as WordNet, are built around lexical relations such as synonymy, antonym, hyponymy, etc. and they do not provide enough information for ontology building. On the other hand, ontologies building requires deeper and more accurate knowledge than general vocabulary contains and, consequently, demands specialized domain resources. This paper presents a linguistic resource developed for Spanish, that has been built following some Meaning-Text Theory principles, in order to contain as much possible knowledge related to several specialized domains.

Keywords: Linguistic resource, ontologies enrichment, Meaning-Text Theory.

1. Introduction

Nowadays, scholars claim it is necessary to associate linguistic information with ontologies (Buitelaar *et al.* 2009; Cimiano *et al* 2008). More to the point, more powerful models seem to be necessary. Buitelaar *et al* have created three different models, *LingInfo*, *LexOnto* and *LexInfo*, which associate linguistic information to ontology elements. *LingInfo* (Buitelaar *et al.* 2006) is a RDFS-based lexicon that enables the association of both aspects through the definition of *LingIngfo* objects. *LexOnto* (Cimiano *et al* 2008) specifies the structure of complex linguistic structures. *LexInfo* (Buitelaar *et al.* 2009) was developed as an ontological lexicon using the LMF metamodel. On the other side, Montiel-Ponsoda *et al.* (2008) and Aguado de Cea *et al.* (in press) developed the LIR model which associates structured multilingual information with ontology concepts.

Ontologies are built according to one model of knowledge, usually proposed by experts. The methodology of work (first, extraction of knowledge from experts' reports and then modelization by ontology engineers) guarantees the adequacy of the model. The first component of an ontology[1] is a set of namespace declarations. It implies that the model is based on a list of concepts and its definitions. In order to describe the collection of instance data, ontologies describes the relation between concepts (the properties, such us *flows into*) by

[1] As we work with OWL web ontologies, we will focus on these types of ontologies forward.

domain and range (*river flows into the sea*, where *river* is the domain and *sea* is the range, example taken from *hydrOntology*, Vilches 2009). As domain and range correspond to instances included previously by experts in the list, this system poses difficulties to the ontology engineers: following this example, as *town, village* and *bridge* are not concepts of the *hydrOntology*, the engineer cannot add expressions such as *river pass through the town, river floods the village,* or *river pass under the bridge*, even if they belong to the more general knowledge of river. In summary, engineers need information as complete and accurate as possible, but domain and range structure reduce the knowledge to the previous list of concepts of the domain. Hence, the question that arises is, could the ontology be enriched with this complementary information?

As far as we know, the typical ontology construction process is not flexible enough to be able to include this kind of information. RDF, OWL and even SKOS do only allow for partial linguistic information (Buitelaar, 2009).

This paper briefly describes why we propose the extraction of knowledge from a linguistic resource, called *BADELE.3000*[2] (Barrios and Bernardos, 2007; Barrios 2009; Barrios 2010) that follows the Meaning-Text Theory (MTT) principles (Mel'čuk, 1996 & 1997). Section 1 summarizes the main MTT concepts that have been applied here and its lexicographic tools. Section 2 presents the data base *BADELE.3000*. In the third section, *Hydrontology* is briefly explained. Section 4 includes our proposal of enriching Hydrontology with the knowledge from specialized terms contained in *BADELE.3000*. Finally, some conclusions and future work are also presented in section 5.

2. MTT and its lexicographic tools

As for the lexicographic tools applied to BADELE.3000, we have resorted to three concepts proposed by the Meaning-Text Theory (MTT).

The first one is the *lexical function* (LF) (Mel'čuk, 1996): a LF associates a given lexical expression L (such as *sound*), which is the argument or keyword of F, with a set of lexical expressions – the value of F (such as *loud*, *strong*, *heavy*, *deafening*, etc.) – expressing a specific meaning associated with F (for instance, 'intense' for the examples just mentioned which correspond to the LF known as **Magn**). More than 100 different LFs have been

[2] *BAse de Datos del Español como Lengua Extranjera de los 3000 sustantivos más usados del español de España* (Database of Spanish as a Foreign Language which contains the 3.000 Spanish nouns more used in Spain)

discovered in natural languages. Most of them are universal across languages. For example, the general sense of the verb in the expression *to have a shower* and its equivalents in other languages is called $\mathbf{Oper_1}$, that is, an LF that allows us to translate *to have a shower* automatically into different languages, as shown in (1):

(1) $\mathbf{Oper_1}$ (ducha) = darse, $\mathbf{Oper_1}$ (shower) = to have, $\mathbf{Oper_1}$ (douche) = prendre, $\mathbf{Oper_1}$ (doccia) = fare

Another LF, $\mathbf{S_i}$ (the name of the participant of the situation) applies to syntagmatic relations such as *school – teacher – student* (cf. (2)):

(2) $\mathbf{S_1}$(school) = student; $\mathbf{S_2}$(school) = teacher

The second concept is the *semantic label* (Polguère, 2003): for our defining purposes, a semantic label is the equivalent to the genus in traditional definitions by genus and differentia. For instance, *whale* could be defined as a 'sea mammal that breathes air through a hole at the top of its head and is hunted for meat and for other purposes, as a source of other materials'. The first part of this definition, 'sea mammal', the genus, is known in MTT approach as semantic label.

The third concept is the *actant* (Mel'čuk, 2004a, 2004b) and its derivate, the actantial structure. Actants correspond to beings or things that participate in the process expressed by a predicate: MTT approach considers that there is a sort of argument structure in all kinds of predicative words, which means that not only do the verbs have actants, but also the adjectives, adverbs and the predicative nouns. The actantial structure reflects the syntactic expression of the actants, as shown in the example of *fleuve* (river) of Dicouèbe[3]:

> (3) *fleuve* (river):
> *qui commence au lieu X, passe par les lieux Z et se termine dans l'étendue d'eau Y* (which starts at the X place, flows through the Z places and finishes at the Y area).

[3] *Dictionnaire en Ligne de Combinatoire du Français*, developed within the MTT framework and accessible via web by http://olst.ling.umontreal.ca/dicouebe/

3. BADELE.3000

BaDELE3000 is a Database for Spanish as a Foreign Language that contains information about the 3,300 most frequently used verbs and 3,600 most frequently used nouns in Spanish[4]. This set of words is based on the statistical study of the Corpus Cumbre[5] by Almela *et al.* (2005) and refined with the information provided by some native speakers.

A systematic development process was followed when building the database (Bernardos and Barrios, 2008), based on general principles and guidelines for the construction of data models and relational databases. First, a knowledge acquisition process (Scott *et al*: 1991) was carried out in order to acquire a comprehensive understanding of the characteristics and requirements concerned. The techniques applied mainly involved the analysis of many publications (Wanner: 1996; Grossmann and Tutin: 2003), dictionaries (*Dicouèbe, DiCE, DECFC*); software systems, such as *CALLEX* (Apresjan *et al*, 2003a, 2003b) and *CALLEX-ESP* (Boguslavsky, Barrios, Diachenko, 2006); and the exchange of information between the linguists and the computer engineers.

The database was built including Lexical Functions and applying Lexical Inheritance Principle (Mel'čuk and Wanner, 1996; Mel'čuk, 1996: 76-78) and the Principle of Domain Inheritance (Barrios, 2010). This principles state that the keywords (nouns) of some noun-verb collocations share some semantic properties, and by applying it to *BADELE.3000* it served to automatically obtain approximately 9,000 Spanish lexical relations (Barrios and Bernardos, 2007). In fact, a total of 20,700 relations are formalized by means of LFs in this database; for this reason we claim that *BADELE.3000* is a linguistic resource that can prove very useful for natural language processing applications and for domain ontology enrichment.

Another interesting asset in the database is its formal typology of the nouns. Based on the traditional semantic labels of the MTT that have been renamed as *hierarchical semantic labels*, our noun typology, unlike other MTT typologies (cf. *Dicouèbe*), shows that lexical units sharing a hierarchical semantic label either belong to the domain of a LF or even share some of its values (for more details see Barrios, 2010).

[4] The database includes the vocabulary necessary for an intermediate-advanced level of language. For that reason, it is not only useful for Spanish students but also for NLP applications requiring vocabulary about general knowledge.
[5] This is a corpus of 20 million words of written and spoken Spanish.

4. hydrOntology: a global ontology of the hydrographical domain

hydrOntology (Vilches-Blázquez *et al.*, 2009) is an ontology in OWL that follows a top-down development approach. Its main goal is to harmonize heterogeneous information sources coming from several cartographic agencies and other international resources. Initially, this ontology was created as a local ontology that established mappings between different data sources (feature catalogues, gazetteers, etc.) of the Spanish National Geographic Institute (IGN-E). Its purpose was to serve as a harmonization framework amongst Spanish cartographic producers. Later, the ontology evolved into a global domain ontology and it attempts to cover most of the concepts of the hydrographical domain.

hydrOntology has been developed according to the ontology design principles proposed by (Gruber, 1995) and (Arpírez *et al.*, 1998). Some of its most important characteristics are that the concept names (classes) are sufficiently explanatory and are correctly written. Thus each class tries to group only one concept and, therefore, classes in brackets and/or with links ("and", "or") are avoided. According to certain naming conventions, each class is written with a capital letter at the beginning of each word, whilst object and data properties are written with lower case letters.

Regarding methodological issues, the approach adopted is METHONTOLOGY, a widely used ontology building methodology. This methodology emphasises the reuse of existing domain and upper-level ontologies and proposes using, for formalisation purposes, a set of intermediate representations that can be later transformed automatically into different formal languages. A detailed description of the methodology of this ontology building can be found in (Gómez-Pérez et al., 2003).

In order to develop this ontology, following a top-down approach, different knowledge models (feature catalogues of the IGN-E, the Water Framework European Directive, the Alexandria Digital Library, the UNESCO Thesaurus, Getty Thesaurus, GeoNames, FACC codes, EuroGlobalMap, EuroRegionalMap, EuroGeonames, several Spanish Gazetteers and many others) have been consulted; additionally, some integration issues related to geographic information and several structuring criteria (Vilches-Blázquez *et al.*, 2007) have been considered. The aim was to cover most of the existing GI sources and build an exhaustive global domain ontology. For this reason, the ontology contains one hundred and fifty (150) relevant concepts related to hydrography (e.g. river, reservoir, lake, channel, and others), 34 object properties, 66 data properties and 256 axioms.

A snapshot of the class hierarchy of hydrOntology in the Protégé ontology editor can be seen in Figure 1. The concept Río (River) has been chosen for illustration. It has nine annotations related to it: three provenance annotations, two comment annotations, three label annotations, and one source annotation. As previously reported, the provenance annotation gives information about the linguistic resources (glossaries, thesauri, dictionaries, etc.) from which labels have been obtained. Since there are no mechanisms for relating the label (e.g. River) with its source of provenance (e.g. Water Framework Directive), the authors have decided to include the label in the provenance text for the sake of clarity (e.g.; "River – Water Framework Directive. European Union"@en).

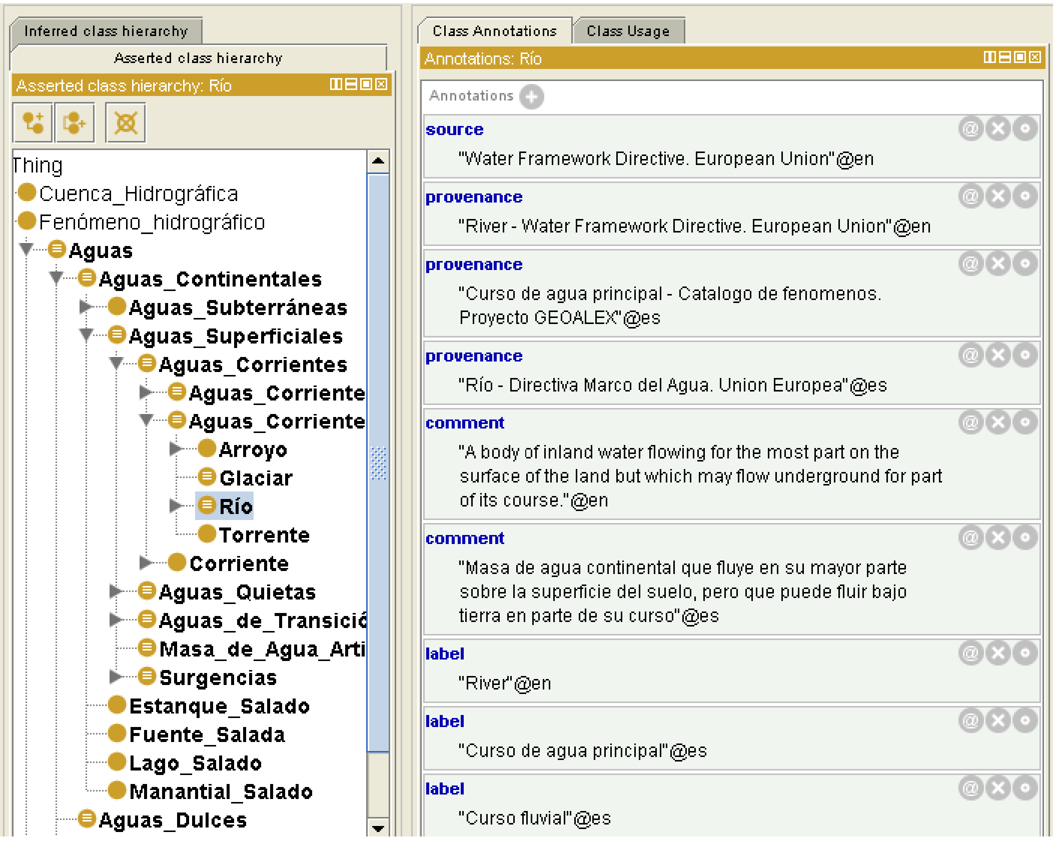

Figure 1: Snapshot of hydrOntology and the linguistic information associated to the Río ontology concept

Two comments are included, one in Spanish, and one in English, though no relation to any of the labels is given. Finally, three label annotations are given: two in Spanish (in addition to the one given in the URI, i.e., Río) and one in English. The two additional labels are *Curso de agua principal* (Main Watercourse), and *Curso fluvial* (Watercourse).

5. Comparing the knowledge contained in hydrOntology and BADELE.3000

In (Barrios, Aguado and Ramos, 2009) we presented how we have enriched *BADELE.3000*, a general purpose linguistic resource, with geographic domain terms in order to convert it into a highly exploitable flexible lexicographic resource. However, in this paper we have focused on how to improve *hydrOntology* with relevant linguistic information taken from *BADELE*.3000, in order to increase possible inferences of the domain knowledge. To this end, we first explain the methodology followed when adding new hydrographical terms to the database, and then we compare the domain knowledge extracted from the ontology to the results obtained from the database. Finally, we draw some conclusions about how the results obtained from both knowledge resources can be applied in the future construction of related domain ontologies.

First of all, we extracted 37 concepts (terms) from *hydrOntology* and we added the labels representing the classes (nouns or nominal groups) to *BADELE.3000*. Then, we analyzed them in the light of the MTT framework, meaning that we developed 37 lexicographic entries enriched with semantic labels, definitions, combinatory and Lexical Functions. Note that information taken from the ontology led to the hierarchical classification of the nouns or nominal groups. This implied a new hierarchy of semantic labels in the database, which constituted a valuable aid when defining the 37 new terms, although some inconsistencies were detected. These inconsistencies are due to the fact that very often domain experts use different sources for definitions: for instance, *conducto* (channel) is defined as “tube that is used to carry liquids”, and *tubería* (pipe), which is a subclass of *conducto*, is defined as “channel that is used to carry liquids or gas”. Obviously, the first definition which is wider should subsume the second one and then *conducto* should be defined also as “used to carry liquids or gas” to include the more specific one.

As a second step, we compared the properties contained in the ontology and the ones in the database that cater for the knowledge attached to the collocations. Properties contained in the ontology are much less than the ones contained in the linguistic resource, owing to the lack of flexibility of domain-range structure, which demands instances of the ontology as we have exposed. For instance, as Figure 2 shows, *conducto* (*channel*) has two properties in *hydrOntology* (rounded in green): *desaguar* (to drain away) (*el conducto desaguaen el mar;* the channel drains away into the sea) and *conducir* (to channel) (*el conducto conduce aguas superficiales;* it channels superficial water). Three classes of the ontology are *conducto* (channel), the domain (in blue); and *agua* (water) and *aguas superficiales* (superficial water), the range (rounded in red):

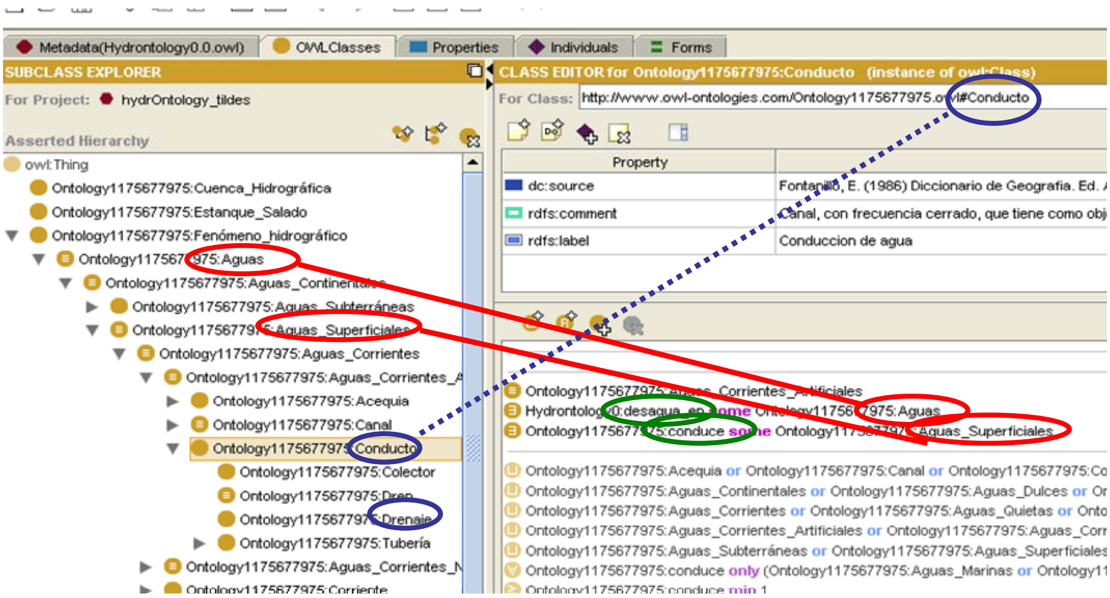

Figure 2: Two properties attached to *channel* verbal collocations in hydrOntology

In contrast, *BADELE.3000* contains 43 properties attached to the lexical combinatory of the noun *conducto* (channel). Figure 3 shows most of them. As we can see, the collocations, such as *acondicionar un conducto* (to retrofit or fit a channel), or *abrirse un conducto* (to open a channel), are translated into the formal language of Lexical Functions (LFs). Moreover, LFs can be converted into natural language, for instance, first, we have **PreparFact$_0$** that means 'someone prepares something (the channel, in this case) in order to do what is supposed to be done with it', in other words, 'to fit a device (the channel) for use in good conditions' and in this way it corresponds to *retrofit the channel*; second, we also have **IncepFact$_0$** that means 'something (the channel, in this case) starts to do what is supposed to be done with it', and thus it corresponds to *the channel is opened*, etc.

On the other hand, we have also some more LFs: **FinFact$_0$** that means 'something (a channel) finishes doing what it is supposed to do', and covers the collocation *cerrarse el conducto* (the channel closes); **AntiBonFact$_0$** means 'something (the channel) does what it is not supposed to do', and applies to *taponarse/ atascarse el conducto* (a channel gets blocked/ becomes obstructed); **CausIncepFact$_0$** means 'someone causes something (the channel) to start doing what it is supposed to do', and includes *poner/ instalar un conducto* (to put/ install a channel); **IntentarCausBonFact$_0$** means 'someone tries that something (a channel) does well what it is supposed to do', and covers *desinfectar/ inspeccionar/ limpiar un conducto* (to disinfect/ inspect/ clean a channel); **CausDenuevoFact**$_0$ means ‘someone causes that something (a channel) does one more time what it is supposed to do’, and includes *taponarse/ atascarse el conducto* (a channel gets blocked/ becomes obstructed); **LiquFact**$_0$ means ‘someone causes something (a channel) to finish doing what is supposed to be done’, and corresponds to *cerrar un conducto* (to close a channel); **LiquAntiBonFact**$_0$ means ‘someone causes something (a channel) to finish doing what is supposed to be done in a wrong way’, and includes *cambiar/ sustituir un conducto* (to replace/ change a channel); **LiquFunc**$_0$ means ‘someone causes that something (the channel) ceases to exist, and corresponds to *destruir un conducto* (to destroy a channel); **CausAntiBonFact**$_0$ means ‘someone causes that something (a channel) does what is supposed to be done in a wrong way’, and covers *dañarse un conducto* (to damage a channel); and **CausDegrad** means ‘someone causes something (a channel) to break down/ go wrong’, and includes *perforar un conducto* (to drill a channel).

id-FL	id-L	id-UL
PreparFact0	conducto	acondicionar
IncepFact0	conducto	abrirse
FinFact0	conducto	cerrarse
AntiBonFact0	conducto	taponarse
AntiBonFact0	conducto	atascarse
CausFact0	conducto	abrir
AntiBonFact0	conducto	obstruirse
CausIncepFact0	conducto	poner
IntentarCausBonFact0	conducto	desinfectar
PreparFact0	conducto	canalizar
CausDenuevoFact0	conducto	reparar
LiquFact0	conducto	cerrar
PreparFact0	conducto	desviar
IntentarCausBonFact0	conducto	inspeccionar
LiquAntiBonFact0	conducto	cambiar
LiquAntiBonFact0	conducto	sustituir
IntentarCausBonFact0	conducto	limpiar
LiquFunc0	conducto	destruir
PreparFact0	conducto	manipular
CausAntiBonFact0	conducto	dañar
CausDegrad	conducto	perforar

Figure 3: Twenty one of the forty three properties attached to *channel* verbal collocations in *BADELE.3000*

As Figure 3 shows, most of the collocations in *BADELE.3000* are verbal and can match properties in the ontology, but we also find adjectival collocations. As Figure 4 shows, four of them correspond to *qualifying adjectives,* which are formalized by non standard LF **Cal** (see next paragraph), such as *estrecho* (narrow), *grueso* (thick), *fuerte* (strong), *invisible* (invisible); two of them correspond to *relational adjectives*, which are formalized by non standard LF **Rel**: *subterráneo* (underground), *transmisor* (transmitter); and several prepositional groups functioning as adjectives, such as *de aire* (of air), *de gas* (of gas), *de agua* (of water)[6]:

id-FL	id-L	id-UL
ProxLabreal12	conducto	usar (como)
CuasiFact0	conducto	servir (de)
ProxLabreal12	conducto	utilizar (como)
Rel	conducto	de combustible
Cal	conducto	invisible
Rel	conducto	de aire
Rel	conducto	de gas
Rel	conducto	de evacuación
Cal	conducto	grueso
Cal	conducto	fuerte
IntentarCausBonFact0	conducto	purgar
Rel	conducto	de agua
Rel	conducto	de oxígeno
Rel	conducto	de entrada
Real1	conducto	desplazarse (por)
Cal	conducto	estrecho (adj)
Real1	conducto	discurrir (por)
Real1	conducto	discurrir (por algún lugar)
Rel	conducto	de salida
Rel	conducto	subalterno
Rel	conducto	de ventilación

Registro: 5190 de 20274

Figure 4: Properties attached to *channel* adjectival and verbal collocations in *BADELE.3000*

There are 65 LFs called standard LFs, because they are universal and proved useful in several languages. However, MTT allows to create and work with other LFs before being accepted as standard LFs. For instance, **Rel** and **Cal** have been created in our database as non standard adjectival LFs, in order to distinguish adjectival collocations corresponding to relational adjectives (such as *underground channel*) from the ones corresponding to qualifying adjectives (such as *strong channel*).

We have also created non standard LFs, which do not correspond necessarily to collocations, but to lexical combinations attached to different types of every one of the concepts. This working practice provides for database entries richer in knowledge than collocations does.

One particularly interesting entry is *drenaje*, drain. hydrOntology contains only one property of drain, *desagua en* (drain away), and six added values (*superficial*, superficial; *mixto*, mixed; *subterráneo*, underground; *cerrado,* closed; *desconocido,* unknown; *sin catalogar* uncatalogued). Whereas in the database several properties are reflected via LFs and

[6] As we explain in the next paragraph, there are standard LFs and non standard LFs. Among others, in our database we have created two non standard adjectival LFs, **Rel** and **Cal**, in order to differentiate adjectival collocations corresponding to relational adjectives from the ones corresponding to qualifying adjectives.

lexical combinations, such as *desaguar (la tierra) con un drenaje* (to drain (the land) with a drain), *impermeabilizar (los cimientos) con un drenaje* (make ground impervious with a drain), *efecto de drenaje* (drain effect), *pozo de drenaje* (drain well), *paquete de drenaje* (drain pack), *técnicas de drenaje* (drain technique), *sistema de drenaje* (drain system), *dispositivo de drenaje* (drain device), *estructura de drenaje* (drain structure), *conducto de drenaje* (drain channel).

The resort to non standard LFs made it possible to arrive to another set of 38 properties, most of them corresponding to different types of draining. In addition to **Rel** and **Cal**, we used the non estándar LFs **Finalidad** (Finality), and **Localización** (Localisation), which respectively covers combinatory attached to the finality and localisation of the concept.

The finality or objective of the *drain* corresponds to expressions such as *para el tratamiento de los deslizamientos del terreno* (for landslide treatment), *para aguas de depuradora* (for water filter systems or for sewage treatment plant), *para aguas de lluvia* (for rain water), *para recuperación del suelo* (for terrain recovery), *para bajar el nivel del agua* (for lowering the bar of water), *para hacer un terrario* (for making a terrarium), *para zonas de baja permeabilidad* (for low permeability areas), *para aguas pluviales* (for storm water), *para aguas sucias* (for dirty water), *para evitar el encharcamiento* (for avoiding stagnant water); *para el control de exceso hídrico* (for water excess control), *para reforzar terrenos flojos* (for fortifying weak terrains).

The localisation of the drain is expressed by *para aeropuertos* (for airports), *para caminos rurales* (for country roads), *para muros de contención* (for embankment), *para césped artificial* (for artificial grass), *para cimentación* (for tunnel's foundation), *para jardinería* (for gardening), *para pistas deportivas* (for playground), *para acondicionador de aire* (for air-conditioning product), *para árboles* (for trees).

Coming back to **Rel** and **Cal** non standard LFs, we have also found collocations in *drain* entry corresponding to these two types of relations: **Rel** covers *drenaje de nivel freático* (groundwater table drain), *drenaje de agujeros* (hole draining), *drenaje semanal* (week draining), *drenaje lineal* (linear draining); and **Cal** covers *drenaje imperfecto* (imperfect drain), *drenaje pobre* (poor drain), *drenaje eficaz* (effective drain), *drenaje seguro* (safe drain), *drenaje apropiado* (appropriate drain), *drenaje adecuado* (adequate drain), *drenaje activo* (active drain), *drenaje moderno* (modern drain), *drenaje amplio* (large drain).

In summary, if hydrOntology contains one property and six added values, the linguistic resource has 48 properties related to drain.

6. Conclusions

Noting the data we have obtained, the knowledge included in hydrOntology could increase more than 95% by retrieving BaDELE.3000 readable knowledge. However, this way of working implies two problems: first of all, enriching the ontology would imply adding this information manually, and, as we mentioned at the beginning, ontology editors are not prepared to incorporate this kind of information; secondly, the linguistic resource currently contains general vocabulary, geography terms and hydrology terms. As section 4 shows, every entry of the database is developed as deeply as possible, which implies human resources working for months on this database. However, no doubt these efforts would produce very interesting results in speeding up the conceptualization phase of new ontologies.

Consequently, there are two possible positions when considering our future work: we could use the linguistic resource only as a source of information, in the same way we use the experts reports or WordNet resource; or we could try to automatically extract the information needed from this linguistic resource, and follow LexInfo example, which would imply to develop new tools.

7. Acknowledgements

This work has been partially funded by the National Project “GeoBuddies: Anotación semántica colaborativa con dispositivos móviles en el Camino de Santiago” (TSI 2007-65677 C02), the bilateral collaboration UPM-IGN 2010, and the European Project Monnet (FP7-248458).

8. References

Almela, R., Almela, M., Cantos, P., Sánchez, A., Sarmiento, R. (2005). Frecuencias del español. Diccionario y estudios léxicos y morfológicos. Madrid: Universitas.

Apresjan et al. (2003a). Ju. D. Apresjan, I. M. Boguslavsky, L. L. Iomdin, A. V. Lazurskij, L. G. Mitjushin, V. Z. Sannikov, L. L. Cinman . ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the Meaning ⇔Text Theory. Proceedings of the 1st International Conference on Meaning-Text Theory. Paris : Ecole Normale Supérieure, 279-288. [http://meaningtext.ilpga.fr/ mtt2003/proceedings/27.Apresian.Boguslavsky.Iomdin.Lazursky.Sannikov.Sizov.Tsinman.pdf]

Apresjan Jury., Igor Boguslavsky, Leonid Iomdin, Leonid Tsinman (2003b). Lexical Functions as a Tool of ETAP-3. *Proceedings of the 1st International Conference on Meaning-Text Theory*: Paris, Ecole Normale Supérieure. [http://meaningtext.ilpga.fr/mtt2003/proceedings.html]

Arpírez JC, Gómez-Pérez A, Lozano A, Pinto HS (1998). "(ONTO)2Agent: An ontology-based WWW broker to select ontologies". In: Gómez-Pérez A, Benjamins RV (eds) ECAI'98 Workshop on Applications of Ontologies and Problem-Solving Methods: Brighton, 16-24.

Aguado de Cea, Guadalupe, Asunción Gómez-Pérez, Elena Montiel-Ponsoda, Luis Vilches-Blázquez (2010). "Rivière or Fleuve? Modelling Multilinguality in the Hydrographical Domain". In Proceedings of the 1st Workshop on the Multilingual Semantic Web, Raleigh, North Carolina, USA.

Barrios Rodríguez, María Auxiliadora (2010). *El dominio de las funciones léxicas en el marco de la Teoría Sentido-Texto*. En *Estudios de lingüística del español (EliEs), vol. 30* [http://elies.rediris.es/elies30/]

Barrios Rodríguez, María Auxiliadora (2009). "Domain, Domain Features of Lexical Functions and Generation of Values by Analogy according to the MTT Approach". In Beck, David; Gerdes, Kim; Milićević, Jasmina; and Polgère, Alain (eds.) pp. 45-54. [http://meaningtext.ilpga.fr/mtt2009/ProceedingsMTT09.pdf]

Barrios María A., Aguado de Cea, Guadalupe., Ramos Jose Ángel (2009) Enriching a lexicographic tool with domain definitions: Problems and solutions. In Sierra *et al* (ed.) *Proceedings of the 1st International Workshop on Definition Extraction*. Incoma. Shoumen (Bulgaria), 14-20.

Barrios, María A. and Socorro Bernardos (2007) *BaDELE.3000*: An implementation of the lexical inheritance principle. In GERDES *et al*, (eds.) Meaning-Text Theory 2007. *Proceedings of the 3nd International Conference on Meaning-Text Theory*. Wiener Slawistischer Almanach. Sonderband, 69. 97-106. [http://meaningtext.net/mtt2007/ proceedings/]

Bernardos M.S. & Barrios M.A. (2008). "Data model for a lexical resource based on lexical functions". *Research in Computing Science*, vol. 27.

Boguslavsky, Igor, Maria A. Barrios, Pavel Diachenko (2006). *CALLEX-ESP*: a software system for learning Spanish lexicon and collocations. *Current Developments in Technology-Assisted Education. Formatex*. Sevilla, 22-26. [http://www.formatex.org/micte2006/pdf/22-26.pdf]

P. Buitelaar, M. Sintek, and M. Kiesel (2006). “A lexicon model for multilingual/multimedia ontologies”. In *Proceedings of the 3rd European Semantic Web Conference (ESWC06).*

P. Buitelaar, P. Cimiano, P. Haase, and M. Sintek (2009). “Towards linguistically grounded ontologies”. In *Proceedings of the European Semantic Web Conference (ESWC09),* 111-125.

P.Cimiano, P. Haase, M. Herold, M. Mantel, and P. Buitelaar (2007). “Lexonto: A model for ontology lexicons for ontology-based NLP”. In *Proceedings of the OntoLex workshop at ISWC07*.

Montiel-Ponsoda, Elena, Guadalupe Aguago de Cea, Asunción Gómez-Pérez, W. Peters (2008). “Modelling multilinguality in ontologies”. In *Proceedings of the 22nd International Conference on Computational Linguistics, Coling 2008*. Companion: Manchester.

DiCE: Diccionario de Colocaciones del Español [http://www.dicesp.com/paginas]

Dicouèbe. Dictionnaire en ligne de combinatoire du Français. [http://olst.ling.umontreal.ca/dicouebe/]

Gómez-Pérez, Asunción, M. Fernández-López, Oscar Corcho, (2003). *Ontological Engineering*. Springer-Verlag: London.

Grossmann, Francis et Agnes Tutin (eds.) (2003). *Les collocations analyse et traitement*. De Werelt: Amsterdam.

Gruber T.R. (1995). “Toward principles for the design of ontologies used for knowledge sharing”. *International Journal of Human-Computer Studies*, 1995, v.43 n.5-6.

Mel’čuk, Igor (2004a/b). Actants in semantics and syntax I: Actants in semantics; Actants in semantics and syntax II: Actants in syntax. *Linguistics*, 42:1, 1-66; 42:2, 247-291.

Mel’čuk, Igor (1996) “Lexical functions: A tool for the description of lexical relations in a lexicon”. In Leo Wanner (ed.), *Lexical functions in lexicography and natural language processing*. Amsterdam/ Philadelphia: John Benjamin, 37-102.

Mel’čuk, Igor, Leo Wanner (1996). “Lexical functions and lexical inheritance for emotion lexemes in German”. In Leo Wanner (ed.), *Lexical functions in lexicography and natural language processing*. Amsterdam/ Philadelphia: John Benjamin, 209-278.

Polguère Alain (2003) “Étiquetage sémantique des lexies dans la base de données DiCo”. *TAL*. Vol 44, nº 2/2003, 39-68.

Scott, A.C., Clayton, J.E. & Gibson, E.L. 1991. *A Practical guide to Knowledge Acquisition.* Reading: Adisson-Wesley.

Vilches-Blázquez, L. M., Ramos, J. A., López-Pellicer, F. J., Corcho, O., Nogueras-Iso, J. (2009). "An approach to comparing different ontologies in the context of hydrographical information". In Popovich et al., (eds.), *IF&GIS'09*. LNG&C Springer: St. Petersburg, 193-207.

Vilches-Blázquez L.M., Bernabé-Poveda M.A., Suárez-Figueroa M.C., Gómez-Pérez A., Rodríguez-Pascual A.F. "Towntology & hydrOntology: Relationship between Urban and Hydrographic Features in the Geographic Information Domain". In Ontologies for Urban Development. Studies in Computational Intelligence, Springer. 2007, vol.61, pp73–84

WANNER, Leo (ed.) (1996). *Lexical functions in lexicography and natural language processing*. Amsterdam/ Philadelphia: John Benjamin.