The forgotten key point for assuring knowledge consistency in CLIR systems

Jesús Cardeñosa, Carolina Gallardo, Luis Iraola
Facultad de Informática – Universidad Politécnica de Madrid
Campus de Montegancedo, 28660 Madrid – SPAIN
{carde,carolina,luis}@opera.dia.fi.upm.es

Abstract

CLIR is the acronym of a great variety of techniques, systems and technologies that associate information retrieval (normally from texts) in multilingual environments. Many of these systems are based on a double architecture composed by systems in charge of extracting information with a great dependency on the language together with classical machine translation systems. In the early 90's, machine translation systems fell from grace due to the failure of big machine translations projects in Europe, Japan and USA. Due to this reason some approaches, particularly those of linguistic knowledge representation were undeservedly forgotten, and above all the so called "interlinguas". Recently, the re-emergence of these models under the generic name of "ontologies" are supporting most of knowledge representation initiatives, even in an language independent way. However consistency problems are not well solved yet. UNL, initially conceived as a contents representation and multilingual generation system, can also be applied to the CLIR.. This paper aims to show how to create and apply domain specific ontologies using the UNL apparatus, particularly the UNL language as a way of ensuring a consistent representation mechanism.

1 Introduction

Cross-Language Information Retrieval (CLIR) deals with the problem of issuing a query in one language and retrieving relevant information in other languages. It aims to help the user in finding relevant information without being limited by linguistic barriers.

In order to overcome the language barrier, three major approaches exist:

- to translate the query into the documents' languages
- to translate the documents into the query's language
- to translate both into an intermediate representation through the use of domain-specific interlinguas.

1.1 Query translation

Online translation can be applied to the query entered by the user. Online query translation will help the user to formulate his/her query in a language other than his/her own. If the user either has at least some reading skills in the target language, it may be possible for him/her to reformulate, elaborate or narrow down the translation proposed.

Because of its simplicity, query translation via machine-readable bilingual or multilingual dictionaries is a very most common approach (Grefenstette, 1996; Ballesteros and Croft, 1997; Davis and Ogden, 1997). Compared to translating an entire document collection, translating a query by dictionary look-up is far more efficient. However, it is unreliable since short queries do not provide enough context for disambiguation in choosing proper translations of query words, and also because it does not exploit domain-specific semantic constraints and corpus statistics in solving translation ambiguities.

A wide array of resources is used in CLIR (Radwan & Fluhr, 1995; Oard, 1997), ranging from multilingual glossaries or dictionaries to multilingual collections of texts and sophisticated taggers and parsers (e.g., Mulinex and MIETTA projects).

1.2 Document translation

Full document translation can be applied offline to produce translations of an entire document. The translations provide the basis for constructing an index for information retrieval and also offer the user the possibility to access the content in his/her own language. Machine or (large scale) human translation, however, is not always available as a realistic option for every language pair. Typically machine translation systems only translate between language pairs which involve one of the major languages, such as English, German or Spanish, and often English plays a pivotal role.

1.3 Domain-specific ontologies for CLIR

Recent CLIR projects (MuchMore, LIQUID) employ a domain-specific ontology that contains the knowledge of the application domain and serves as an interlingual backbone for a multilingual thesaurus. Relevant terms contained in a query are translated into several languages using the term-to-concept links established in the multilingual thesaurus. Domain knowledge represented in the conceptual layer is exploited for expanding the initial query (see figure 1).

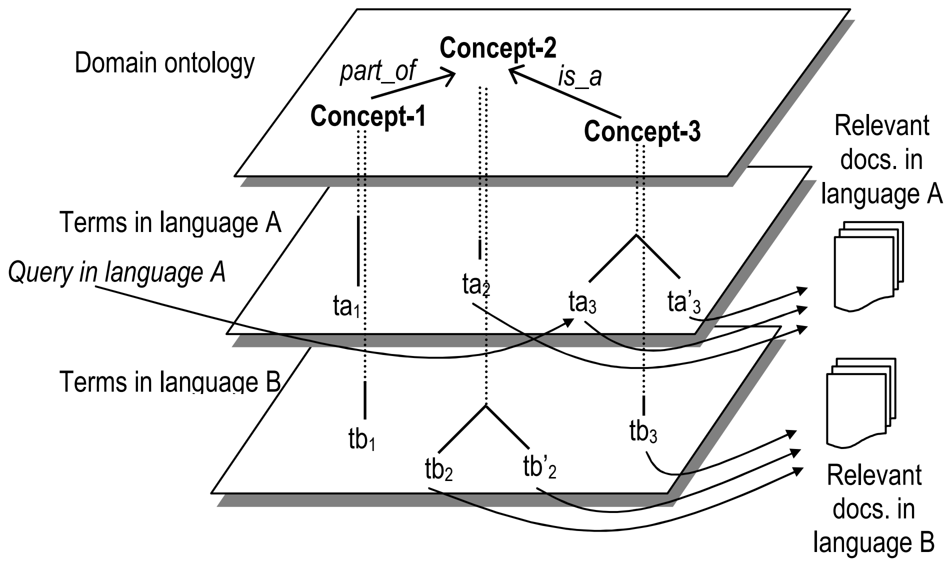

Fig 1: Linking documents and queries through a multilingually mapped ontology

2 Ontologies and Support Languages

Like in many other cases, the definition of an ontology is not completely fixed and agreed on. There are several definitions of ontologies, but for our purpose we will cling to Gruber's one: *an ontology is an explicit specification of a conceptualisation* (Gruber, 1993).

There are two main issues in this definition:

a) Explicit specification
b) Conceptualization

The “explicit specification” of an ontology leads us to the formalization of ontologies and used languages. In this section, we will deal with ontologies support languages as the main way for attaining such explicitness and machine readability.

A conceptualisation is related to the creation of a model of a given domain pointing out the relevant concepts, their relations and functions that made up a complete domain.

In order to support an ontology and inference mechanisms, the question of the language support is crucial. There are two main factors that determine the evolution of ontology languages. These are the *knowledge representation formalism* and *web orientation*.

Regarding the knowledge representation formalism, there appear to be two clear periods that we will refer to as *First Generation Languages* and *Second Generation Languages*. First generation ontology languages are basically frame-based and correspond to the first attempts to build ontologies and establish the ontology engineering discipline (beginning of 90ies). As the most representative frame-based languages are Loom (MacGregor, 1987), Ontolingua (Faquhar et al, 1996) or KIF (Genesereth et al, 1992).

In its beginnings, ontology engineering was highly oriented towards knowledge reuse and share (Neches et al, 1991). All of these languages can be considered as languages for knowledge representation, being KIF (*Knowledge Interchange Format*) the most oriented towards knowledge reuse, since it conforms a sort of “interlingua” of knowledge representation languages.

The common feature of these languages is its frame-based nature. Thus, they are endowed with the usual expressiveness of frames. Basically, they allow for:

- Representing classes and subclasses
- Distinguishing between classes and instances
- Establishing relations between classes.
- Establishing default values.

In a way we could say that these languages are oriented toward a hierarchical conceptualisation of a domain. Needless to say, the Semantic Web wasn’t the main goal in this period. So there is no web integration of these ontologies.

The second generation of ontology languages shows a more logical flavour (although some retain the frame flavour). We are referring to RDF (Lassila et al, 1999), OIL (Horrocks et al, 2000), SHOE (Luke et al, 2000), DAML-OIL (Horrocks et al, 2001) or even XML (Yergeau, et al, 2004). Let’s mention some of the properties of these languages:

- They are based on first order logic (with some possible extensions).
- Use of logic (formal semantics for deduction processes)
- The distinction between class and instance is supported.
- The establishment of taxonomies (class – subclass) is normally supported.
- Representation and inclusion of axioms are supported in some of them.
- Normally no default values are allowed.
- Relations (of different arity) are more or less covered.
- Some of them are oriented towards the Semantic Web (developed by the W3C consortium or either compatible with XML).

These ontology languages show the second parameter: web orientation, they extends the traditional definition of an ontology and try to conceptualise the whole web, that is, the target is no more reuse of knowledge but to achieve the so-called Semantic Web. Thus many of them are based on web languages and technologies (such as XML and RDF developed by the W3C consortium).

It is interesting to see the influence of a standard entity such as W3C as an standardizing body. It is quite obvious the convergence of all these languages towards a unique standard one: OWL (Bechofer et al, 2004).

All these languages seem to have derived in OWL, which is an extension of XML, RDF, DAML and OML. According to the authors, it provides "*greater machine interpretability of Web content than that supported by XML, RDF, and RDF Schema (RDF-S) by providing additional vocabulary along with a formal semantics"*. It was in February, 2004 when it was proposed by W3C to become the standard language for ontology representations in the web.

3 Knowledge Representation vs. Cross-Linguality

Ontologies and knowledge representation are two close concepts. At the end, conceptualisation and formalization of a model or domain are two quite well known issues of Knowledge representation. Ontology engineering does not begin from scratch; many of its theoretical foundations are borrowed form Knowledge Engineering, being formalisms and representation languages no exception.

Historically, semantic nets were the first formalism suitable to represent knowledge, as it extended the expressiveness of pure logical models. The semantic nets were proposed in 1968 by Quillian and he was also who study the knowledge extraction from texts some years later (Quillian, 1968). Even today the degree of conceptual advance in comparison with those years is not high. Possibly the real advance is coming from the capacity of managing great knowledge bases based in an increase of computing power and an increase of the interoperability between heterogeneous systems through standardized formalisms. In those years the main problem was the lack of standardisation of the possible number of relations and also the necessities to expand the amount of information associated to the concepts of a net. This was also the convergence between the conceptual definition of "frame" proposed by Minsky (Minsky, 1974) and the necessities to expand the capacities to encapsulate information in the so called "frame nets" that were the combination of semantic nets with the expansion of the concepts into frames.

Wood in (Brachman et al, 1985) stated two issues that prevent semantic nets from being a good candidate for knowledge representation:

a) Ambiguities in its representation (no specific account of the distinction between class and instance)
b) Lack of a common understanding of the semantic labels, that eventually Wood defines as the "asemanticity" of semantic nets.

For these two reasons, ontology languages turn to frame and logic based formalisms, disregarding the adequacy of semantic nets for the specification of non-hierarchical relations (that is, functions and roles between concepts). Curiously, current ontologies do not fully exploit the most expressive characteristics of semantic nets, resulting in a massive use of relation IS-A. Bearing in mind the features of ontology languages, we could state that there is coverage for vertical relations (class, subclass, instance, plus other) but not for horizontal relations (roles and links between concepts). Horizontal relations enrich the domain representation, as shown in Burg, 1997 and Shamsfard et al, 2004 as attempts to build ontologies from natural language texts. Even if we accept Wood's objection to semantic nets, there is still a wide amount of information that semantic nets offers and ontologies do not exploit, being this the capacity of semantic nets to express horizontal relations, that could be easily integrated into ontology support languages in principle.

Thus relations would not be only limited to a is-a or a-kind-of types, but richer relations will have to be included. A hint of what sort of horizontal relation should be included in domain models is given by natural languages (languages are the main vehicle of expressing knowledge), this is the approach followed in the GUM, following the theoretical positions that Functional Grammar established (Bateman et al, 1990), or as we will see later in the Universal Networking Language (UNL).

A major problem for knowledge based approaches is the creation of the necessary resources: in addition to a multilingual thesaurus such as MeSH (Medical Subject Index), SNOMED (Systematized Nomenclature of Medicine) and ICD (International Classification of Diseases) for the medical domain, these systems require a domain-specific ontology. In order to extract relevant knowledge from technical documentation containing the domain knowledge, several person-years of highly qualified work are required (Gonzalo et al. 1998).

By knowledge bases in our context we understand the set of concepts belonging to a specific domain and the relations between these concepts that also belong to this domain. But when we turn to ontologies, the richness of a domain becomes relegated to a mere enumeration of concepts and a taxonomic organization of them. That is, there is danger of identifying ontologies as mere thesauri.

4 Some Advances: new approaches

Our group is the Spanish Language Centre (www.unl.fi.upm.es) of the UNL Programme of the United Nations (www.undl.org). UNL is basically an artificial language for knowledge representation designed for representing contents written in any language and for generating such contents in any natural language. Borrowing the term from the Machine Translation literature, UNL is an interlingua since it plays the role of an intermediate representation of the text meaning in a language independent way. The next section will depict UNL in more detail.

4.1 UNL as an interlingua

Formally speaking, UNL follows the schema of semantic nets (that is, UNL expresses binary relations between concepts, labelled by a number of semantic tags). The specifications of the language (UNL Center, 2003) formally define the set of relations, concepts and the so-called attributes. Let’s have a look at them in more detail.

Universal words. They conform the vocabulary of the language, i.e., they can be considered the lexical items of UNL. To be able to express any concept occurring in a natural language, the UNL proposes the use of English words modified by a series of semantic restrictions that eliminate the innate ambiguity of the vocabulary in natural languages. If there isn’t any English word suitable to express the concept, the UNL allows the use of words from other languages, if the semantic restrictions describe the meaning of the base word with precision. In this way, the language gets an expressive richness from the natural languages but without their ambiguity. Take, for example, the English word “construction” meaning “the action of constructing” and the “final product”. Thus, the word “construction” will be paired with two different universal words:

construction$_1$ $\rightarrow$ construction(icl>action)
construction$_2$ $\rightarrow$construction(icl>concrete thing)

where “icl” is the abbreviation for “included”. The set of UWs is included in the UNL dictionary.

Relations. These are a group of 41 relations that define the semantic relations among concepts. They include argumentative (agent, object, goal), circumstantial (purpose, time, place), logic

(conjunction, and disjunction) relations, etc. For example, in a sentence like “The boy eats potatoes in the kitchen”, there is a main predicate (“eats”) and three arguments, two of them are instances of argumentative relations (“boy” is the *agent* of the predicate “*eats”,* whereas “potatoes” is the *object*) and one circumstantial relation (“kitchen” is the *physical place* where the action described in the sentence takes place). The specifications provides a definition in natural language of the intended meaning of semantic relations and establishes the contexts where relations may apply, like the nature of the origin and final concept of the relation.
For example, an agent relation can link an action (as opposed to an event or process) and an volitional agent (as opposed to a property or a substance). This characterization of concepts implies a top “ontology” or “taxonomy” similar to the Wordnet (Fellbaum, 1998), whose main purpose is validating the correct application of conceptual relations.

Attributes. They express several types of semantic information that usually modifies the predication described by the net of uws linked through the relations. This information includes time and aspect of the event, polarity and modality of the predication, type of reference of the entities described by the UWs, number and/or gender, etc. In the previous sentence, attributes are needed to express plurality in the object (“potatoes”), definite reference in the both the agent (“boy”) and the place (“kitchen”) and finally and special attribute denoting which UW is the head of the whole expression (the *entry* node).

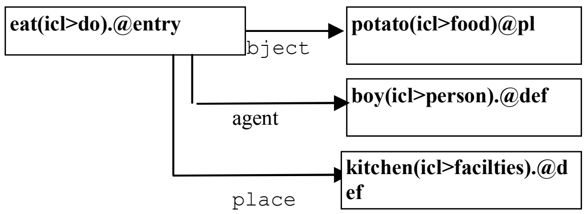

Fig 2: Representation of a UNL expression.

The textual representation in this UNL graph is the following:

agt(eat(icl>do).*@entry*, boy(icl>person).*@def*)
obj(eat(icl>do).*@entry*, potato(icl>food).*@pl*)
plc(eat(icl>do).*@entry*, kitchen(icl>facilities).*@def*)

One of the main objections done to GUM (impossibility of representing information about hearer and speaker) is resolved in UNL by means of attributes: the subset of the language UW + Conceptual Relations defines the propositional part of a given text, the addition of attributes adds contextual meaning like epistemic and deontic modality, speaker’s intention, speaker’s attitudes, informative structure etc.

4.2 UNL as language for knowledge representation

UNL is mainly used as a support language for multilingual generation of contents coming from different languages. However, its design allows for non language centred applications, that is, UNL could serve as a support for knowledge representation in generic domains. When there is a need to construct domain-independent ontologies, researches turn back to natural language (such as Wordnet, GUM or even CyC[1]) to explore the “semantic atoms” that knowledge expressed in natural languages is composed of. UNL follows this philosophy, since it provides an interlingual analysis of natural language semantics. The reasons why UNL could be backed as a firm knowledge representation language are:

[1] http://www.cyc.com

a) The set of necessary relations existing between concepts is already standardized and well defined (overcoming the objection posed by Woods about the asemanticity of semantic networks).
b) It is the product of intensive research on the thematic roles existing in natural languages by a number of experts in the area of MT and IA, guaranteeing wide coverage of all contents expressed in any natural language.
c) Similarly, the set of necessary attributes that modify concepts and relations is fixed and well-defined, guaranteeing a precise definition of contextual information.
d) UNL syntax and semantics are formally defined.

But to really serve as a language for knowledge representation, it must support deduction mechanisms and must specify how a knowledge base could be build up in the UNL language. We will explore this idea by looking closer at the UWs part of the UNL system and how to link them in knowledge base.

4.2.1 The UNL dictionary and its companion KB

The UW dictionary is a repository of UWs and as such does not organise its contents in any way. It is just a (big) set of UWs, each element having no relation with any other. The necessity of establishing certain relations between UWs arises when considering several desirable features of the UNL system:

- Setting the combinatory possibilities of each UW with respect to any other UW regarding the conceptual relations that may link them and the attributes they may accept.
- Enabling a "fall-back" generation mechanism for those UWs that are not linked with HWs in a given language at a given time. Those UWs would be replaced with semantically close, linked UWs so allowing generation to continue.

If word sense disambiguation were the *only* reason for introducing semantic restrictions into UNL, any of the previous approaches could be adopted. However, semantic restrictions have been also used for a different though related purpose: providing a semantic structure to the otherwise "flat" UW dictionary. However, and in order to support these features, the devised solution consists in creating a *network* with the set of UWs as nodes and *semantic relations* as arcs. In such a network, we use the same information both for disambiguating and for building the KB. The semantic restrictions attached to the UWs for disambiguation purposes *also* express knowledge stored in the KB and conversely; the semantic knowledge serves for disambiguation. Such network is called the UNL KB.

From an *extensional* point of view, the UNL KB can be viewed as a finite set of tuples of the form:

$$\langle \text{semantic relation}, uw_1, uw_2 \rangle$$

which can be graphically displayed as:

$$uw_1 \text{ —semantic relation} \rightarrow uw_2$$

The following are examples of tuple, being "icl" and "agt" abbreviations for "included" and "agent" respectively:

helicopter —icl→ concrete thing
ameliorate —icl→ do
do —agt→ thing

Given the huge amount of tuples that it may contain, the UNL KB is best viewed from an *intensional* point of view as a first order logical theory composed of a finite set of axioms and inference rules[2].

Most of the axioms state plain semantic relations among UWs, now viewed as atomic formulas:

$$\text{relation}(uw_1, uw_2)$$

Examples:

icl(helicopter, concrete thing)
icl(ameliorate, do)
agt(do, thing)

Besides atomic formulas, the theory contains complex formulas, like the one stating the transitivity of the "icl" relation:

$$\forall w_1 \forall w_2 \forall w_3 (\ \text{icl}(w_1, w_2) \wedge \text{icl}(w_2, w_3) \rightarrow \text{icl}(w_1, w_3)\ )$$

As for the inference rules, a subset of the standard rules present in first order theories may suffice for defining the relation of syntactic consequence among formulas. The UNL KB is then formally defined as the closure of the set of axioms under the consequence relation.

We can now turn to the tasks the UNL KB is intended to be used for, and get a clearer picture of its concrete contents according to those tasks. The first task we have mentioned is setting the combinatory possibilities of every UW with respect to the rest of UWs and to the set of conceptual relations (and attributes) included in UNL. For any two UWs w_1, w_2 and any conceptual relation *r*, the UNL KB should be able to determine whether linking w_1, w_2 with *r* is allowed (makes sense in principle) or if it is against the intended use of w_1, w_2 and *r*. If we view the KB as a theory, the question is then if the formula $r(w_1, w_2)$ is a consequence (a theorem) of the set of axioms that form the KB or it is not. The axioms needed for answering such questions are mostly derived from the intended usage of the UNL conceptual relations and the broad semantic classes each UW belongs to.

4.2.2 Example

The *instrument relation* ("ins") holds between an event and the concrete thing involved as instrument used for completing the action. In the UNL specifications this is expressed very much like one of our previous formulas:

ins(do, concrete thing)

That is, there is an "ins" arc between UWs "do" and "concrete thing":

do —ins→ concrete thing

On the other hand, the method relation ("met") holds between an event and the mean or method applied for doing the action. This is expressed in the specifications with the formula:

met(do, abstract thing)

[2] This idea is fully developed in the document "The UNL Knowledge Base, a formal description", Luis Iraola. Internal Report, Spanish Language Center. January 1999.

Graphically:

do —met→ abstract thing

The differences between "ins" and "met" impose a semantic difference between concrete and abstract things. In order to set the combinatory possibilities of nominal concepts as destination of these relations, nominal UWs must be included under the "concrete thing" or "abstract thing" respectively. Verbal concepts included under "do" qualify as origin of both relations. These inclusions plus the axioms governing "ins" and "met" are all that is needed in the KB for setting the combinatory possibilities regarding "ins" and "met".

5 Considerations for building the UNL Ontology

The UNL ontology has been developed with several considerations in mind:

- Linguistic relevance. The main goal of the ontology is to aid to the tasks carried out by Analyzers and Generators of UNL, and more generally to any task related with the processing of natural language.
- Language independence. The divisions made in the uppermost levels of the ontology (which are presented in this document) try to be based on very general semantic distinctions present in most of the natural languages.
- Exhaustive and disjoint classifications. The ontology should cover the whole range of concepts (universal words) and, at least in its up-most levels, its divisions should be disjoint.
- Clear membership criteria. Though there always be concepts difficult to situate in the ontology, the goal of giving clear criteria for applying the classification is considered central.

One of the main characteristics of UNL is its flexibility both formally and linguistically. From a linguistic point of view, the UNL ontology serves to a wide variety of natural languages. From the formal point of view, its integration with other support languages (HTML, XML, OWL) could be easily achieved. UNL and OWL could be considered as complementary, integrating thus the formal rigour and machine readability of OWL and the expressiveness and language and domain independence of UNL.

Essentially, UNL has the capability of representing knowledge. However the classical problem emerges. It is the semantic validation process, that is, the set of mechanisms able to deduce coherent domain knowledge from existing one. This is still an open problem, which so far has only attained some partial solutions based on the application of the logic verification rules. However verification rules are not enough to establish a model with sufficient semantic coherence.

6 An illustration of our approach

Possibly, one of the best examples of the utility of UNL is its capability to build knowledge bases from texts in an automatic way. In the following example, we have three sentences from a Spanish document about Heritage policies, more specifically it shows some procedures of how to catalogue existing heritage.

The text is:

1. *To integrate the catalogues of all the Spanish museums in the General Catalogue of Historical Heritage*
2. *To establish the necessary mechanisms to integrate all the information from the Autonomous Communities.*

3. *The registry campaigns to make the General Inventory of Moveable Assets of the Church.*

Figure 3, 4 and 5 show the UNL representation of first, second and third sentences respectively.

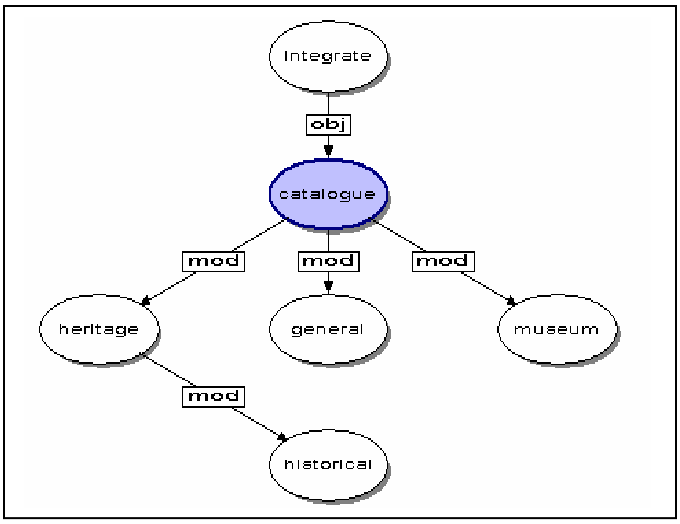

Fig. 3: UNL representation of sentence 1

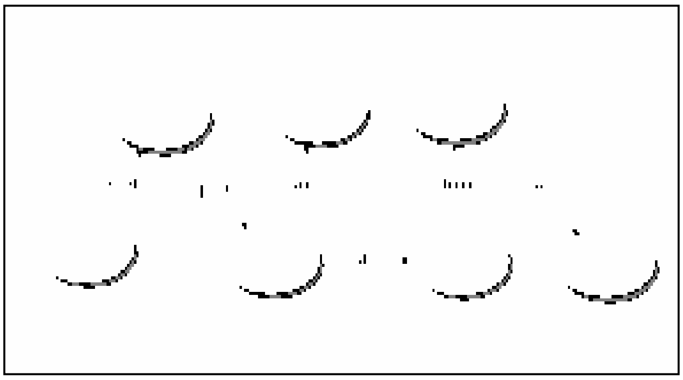

Fig. 4: UNL representation of sentence 2

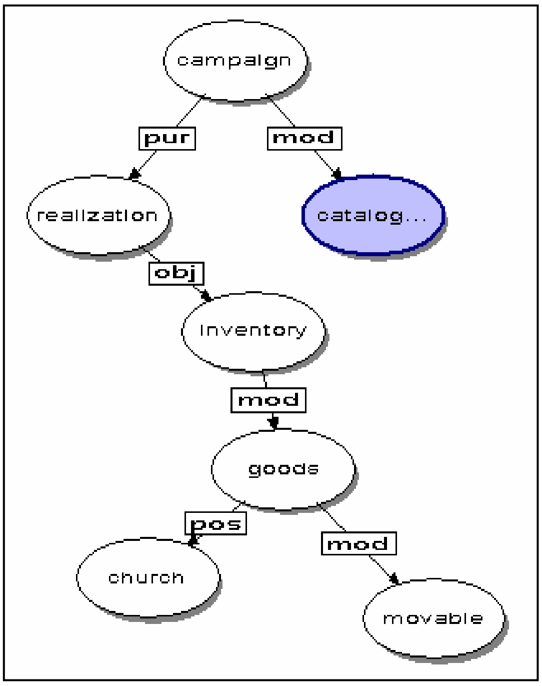

Fig. 5: UNL representation of sentence 3

All these graphs could be merged. Here we are aiming at the capacity of UNL to represent and organize contents, and as a side effect, to generate such contents into different natural languages. By adding all sentence information into a single representation, we can deduct more information about the concepts present in each sentence. Figure 6 represent the merged graph.

The joined representation offers new relations among concepts than those presented in the sentence representations. For example, the term "catalogue" is related to several concepts, some relations may be true whereas others not. For example, here a "catalogue" is described as an entity that could be generic, could belong to museums, to historical heritages, can be a virtual place where some actions are carried out or can be the effective object of a "carrying out" activity. This can be applied to the problem of query expansion in a knowledge based manner, which can complement well known linguistic query expansion (morphological and syntactical variations).

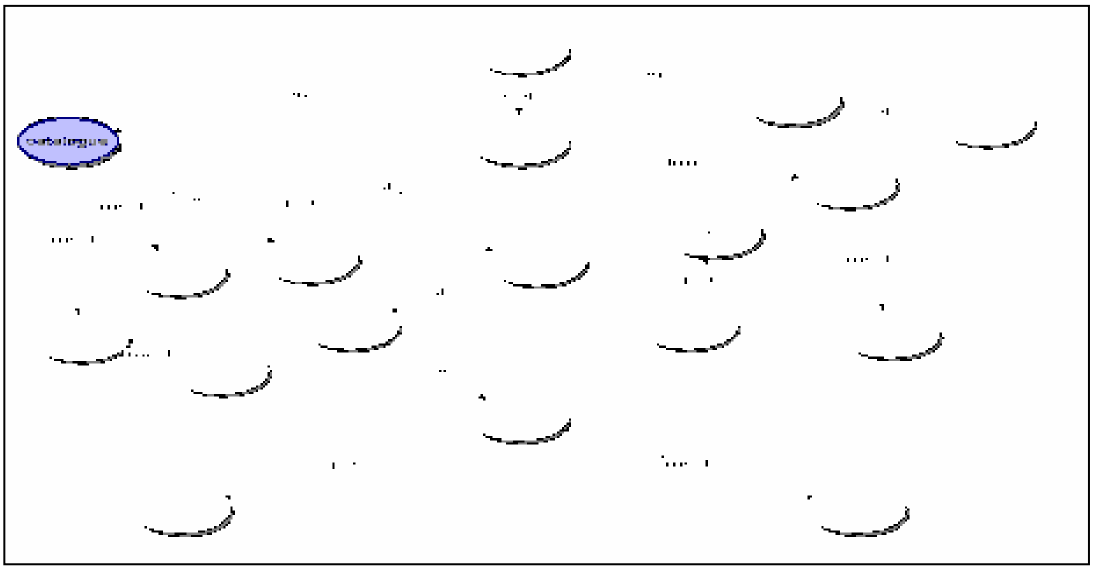

Fig. 6: UNL merged graph

7 CONCLUSIONS

UNL is a language capable of being the formal basis of knowledge representation and also a language capable of representing information coming from textual sources. In order to propose this language as a possible standard for these purposes, we want to bring the attention to the fact that it is maintained by an open world-wide organisation which provides the necessary institutional support.

UNL researchers face the same problems than any others in this field, that is, to define mechanisms that guarantee the coherence of inferred knowledge. We, of course, assume that knowledge represented into UNL can transform implicit knowledge into explicit one. As in many other problems, inferred knowledge needs to be validated, and for that it is necessary to design the domain ontology where the valid combination of relations is defined, including the restrictions that cannot be violated.

A possible way for building such domain ontology is to painstakingly encode all the concepts and relations for the application domain. Our approach relies on statistical analysis of UNL representations of domain specific texts. By exploiting these representations, we hope to be able to build such a domain ontology. At the moment we have obtained encouraging initial results in the cultural heritage domain as a side effort of the Spanish Language Center in the Herein project[3]. We hope to have reliable results during this year.

[3] http://www.european-heritage.net

8 REFERENCES

- Ballesteros, L. and Croft, W.B., 1996. "Dictionary Methods for Cross-Lingual Information Retrieval". In Proceedings of the 7th International DEXA Conference on Database and Expert System, 791-801.
- Ballesteros, L. and Croft, W.B., 1997. "Phrasal Translation and Query Expansion Techniques for Cross-Language Information Retrieval", in Proceedings of ACM SIGIR Conference, 20: 84-91.
- Bateman, J.A; Henschel, R. and Rinaldi, F. (1995) The Generalized Upper Model 2.0. http:// www.darmstadt.gmd.de/publish/komet/gen-um/newUM.html
- Bechhofer. S; van Harmelen, F.; Hendler, J.; Horrocks, I.; McGuinness, DL; Patel-Schneider, P.F. and Stein, L.I. (2004) OWL: Web Ontology Language Reference. W3C Recommendation 10 February 2004 http://www.w3.org/TR/owl-ref/
- Buitelaar, Paul and Netter, Klaus and Xu, Feiyu, 1998 "Integrating Different Strategies in Cross-Language Information Retrieval in the MIETTA Project" Proceedings of the 14th Twente Workshop on Language Technology (TWLT 14). Language Technology in Multimedia Information Retrieval, December 1998. Enschede, The Netherlands
- Burg, J.F.M. and van de Riet, R.P. (1997) The impact of linguistics on conceptual models: consistency and understandability. Data & Knowledge Engineering, Vol 21, 131 - 146
- Farquhar, A.; Fikes, R. and Rice, J. (1996) The Ontolingua Server: a Tool for Collaborative Ontology Construction. Proceedings of the 10th knowledge acquisition for knowledge-based systems workshop, Canada.
- Fellbaum, C., editor (1998). WordNet: An Electronic Lexical Database. Language, Speech, and Communication Series. MIT Press.
- Fensel, D.; Horrocks I; Van Harmelen F.; Decker S.; Erdmann, M. and Klein, M. (2000) OIL in a nutshell. Proceedings of the 12th European Workshop on Knowledge Acquisition, Modeling, and Management
- Genesereth, M.R.; Fikes, R.E. (1992) Knowledge Interchange Format. Version 3.0. Reference Manual. Computer Science Department. Standford University. California.
- Gonzalo, J., Verdejo, F., Peters, C. and Calzolari, N., 1998. "Applying EuroWordNet to Cross-Language Text Retrieval". Computers and the Humanities, Special Issue on EuroWordNet.
- Gonzalo, J., Verdejo, F. and Chugur, I., 1999. "Using EuroWordNet in a Concept-Based Approach to Cross-Language Text Retrieval". Applied Artificial Intelligence Special Issue on Multilinguality in the Software Industry: the AI contribution.
- Gruber. T. A. (1993) A translation Approach to portable ontology specifications. Knowledge Acquisition. Vol. 5, 199-220
- Horrocks I.; van Harmelen, F. Reference description of the DAML+OIL ontology markup language. 2001. http://www.daml.org/2001/03/reference.html
- Hovy, E.H., Ide, N. Frederking, R.E., Mariani, J. and Zampolli, A. (editors) . 2000. Multilingual Information Management. Available at http://www.cs.cmu.edu/~ref/mlim - 20-12-00.
- Hull, D., and Grefenstette G., 1996. "Experiments in Multilingual Information Retrieval". Proceedings of ACM, SIGIR'96. Zurich.
- Lassila, O and Swick, R.R (1999) Resource Description Framework (RDF) Model and Syntax Specification . W3C recommendation, 1999. www.w3.org/TR/PR-rdf-syntax

- LIQUID project: http://liquid.sema.es/
- Luke, S.; Heflin, J. (2000) SHOE 1.01. Proposed specification. http://www.cs.umd.edu/projects/plus/SHOE/spec.html
- MacGregor, R. and Bates R. (1987) The Loom Knowledge Representation Language. Technical Report ISI-RS-87-188, USC Information Sciences Institute, Marina del Rey, CA.
- MIETTA project: http://www.mietta.info/
- Minsky. M. (1974) A Framework for Representing Knowledge. MIT Press
- MuchMore project: http://muchmore.dfki.de/
- MULINEX: Multilingual Indexing, Navigation and Editing Extensions for the World-Wide Web. Proceedings AAAI Spring. Symposium on Cross-Language Text and Speech Retrieval. Menlo Park CA, 1997.
- Neches, R; Fikes, R.E.; Finin, T.; Gruber, T.R.; Senator, T. and Swartout, W (1991) Enabling technology for knowledge sharing. AI Magazine, 12(3):36-56
- Oard, D., 1997. "Alternative Approaches for Cross-Language Text Retrieval". Proceedings of the AAAI Spring 1997 Symposium on Cross-Language Text and Speech Retrieval.
- Quillian M.R. (1968) Semantic Memory. Semantic Information Processing. M.Minsky (Ed.), MIT press
- Sánchez Valderrábanos, Belskis, Iraola, 2002 "Multilingual terminology extraction and validation", LREC 2002, Las Palmas de Gran Canaria 2002.
- Shamsfard, M. and Abdollahzadeh, A. (2004) Learning Ontologies from Natural Language Texts. International Journal of Human Computer studies. Vol. 60, 17-63
- UNL Center. UNL specifications v.3.2. July, 2003. http://www.undl.org/unlsys/unl/UNL%20Specifications.htm
- Vintar, S.; Buitelaar, P.; Volk, M. (2003) Semantic Relations in Concept-Based Cross-Language Medical Information Retrieval. Proceedings of the ECML/PKDD Workshop on Adaptive Text Extraction and Mining (ATEM), September 22nd , 2003, Cavtat-Dubrovnik, Croatia.
- Volk, M.; Vintar, S. and Buitelaar, P. (2003) Ontologies in Cross-Language Information Retrieval. In: Proceedings of WOW2003 (Workshop Ontologie-basiertes Wissensmanagement), Luzern, Switzerland, April 2003.
- Wood, J. (1985) What's in a link ? In R. Brachman and H. Levesque (ed) Readings in Knowledge Representation. Morgan Kaufmann
- www.undl.org
- Yang, Y., Carbonell, J., Brown, R., Frederking,R., 1998. "Translingual Information Retrieval: Learning from Bilingual Corpora". AI Journal Special Issue: Best of IJCAI'97
- Yergeau, F; Bray, T; Paoli, J; Sperberg-McQueen C.M and Maler E. (2004) Extensible Markup Language (XML) 1.0 (Third Edition). W3C Recommendation. http://www.w3.org/TR/REC-xml/