

Integrating Metaphor Information into RDF/OWL EuroWordNet

Ernesto William De Luca*, Birte Lönneker-Rodman†

*Otto von Guericke University of Magdeburg
Faculty of Computer Science, Department of Knowledge and Language Engineering
Universitätsplatz 2
39106 Magdeburg, Germany
ernesto.deluca@ovgu.de

†International Computer Science Institute
1947 Center Street, Suite 600
Berkeley, CA 96704, USA
loenneke@icsi.berkeley.edu

Abstract

In this paper, we discuss the integration of metaphor information into the RDF/OWL representation of EuroWordNet. First, the lexical database WordNet and its variants are presented. After a brief description of the Hamburg Metaphor Database, examples of its conversion into the RDF/OWL representation of EuroWordNet are discussed. The metaphor information is added to the general EuroWordNet data and the new resulting RDF/OWL structure is shown in LexiRes, a visualization tool developed and adapted for handling structures of ontological and lexical databases. We show how LexiRes can be used to further edit the newly added metaphor information, and explain some problems with this new type of information on the basis of examples.

1. Introduction

Language Engineering involves the development and application of software systems that perform tasks concerning the processing of human natural language (Cunningham, 1999). Different tools have been designed, constructed, and are used, for tasks like translation, language teaching, information extraction and indexing. Other, more intangible “language engineering tools” are *language resources*. Language resources are essential components of language engineering, containing a wide range of linguistic information with different degrees of complexity. These linguistic resources are sets of language data and descriptions in machine readable form, used for building, improving or evaluating natural language and speech systems or algorithms. In (Cole et al., 1997) various types of language resources, i.e. written and spoken language corpora, lexicons and terminological databases are briefly presented.

In the remainder of this paper, we concentrate on *lexical resources* that provide linguistic information about words. This information can be represented in very diverse data structures, from simple lists to complex repositories with many types of linguistic information and relations attached to each entry, resulting in network-like structures.

Lexical resources are used in Natural Language Processing, for example, to obtain descriptions and usage examples of different word senses. Different word senses refer to different concepts, and concepts can be distinguished from each other not only by their definitions or “glosses”, but also by their specific relations to other concepts. Such disambiguating relations are intuitively used by humans. However, if we want to automate the process of distinguishing between word senses (Word Sense Disambiguation), we have to use resources that include appropriate knowledge, such as relations. One of the most important resources available for

this purpose is WordNet (Fellbaum, 1998) and its multilingual variants, including MultiWordNet (Pianta et al., 2002) and EuroWordNet (Vossen, 1999).

However, since some lexical resources or ontologies, especially WordNet, provide very fine grained word sense distinctions, tools have been implemented that facilitate navigation through the lexical information and modification of the lexicon structure. Moreover, the coverage of word senses in lexical resources cannot be exhaustive; no resource can list all word senses. It is therefore useful to flexibly enlarge a general resource with domain-specific terms, possibly varying with the purpose of the application that will use the resource. The work presented in this contribution exemplifies this idea: We design an add-on to EuroWordNet that contains metaphor information. The LexiRes tool is used for navigating and restructuring the resulting contents of the lexical resource.

The rest of the paper is structured as follows. The lexical database WordNet, its variants including RDF/OWL EuroWordNet, and tools for working with them are introduced in Section 2. After a brief description of the Hamburg Metaphor Database (see Section 3.), examples of its conversion into the RDF/OWL representation of EuroWordNet are discussed. Information on metaphor relations and glosses of metaphorical synsets is added to the general EuroWordNet data by creating new RDF files (see Section 4.). The new resulting RDF/OWL information is then shown in LexiRes, a visualization tool developed and adapted for handling structures of ontological and lexical databases. We discuss how the tool can be used to integrate information on new synsets that we had to create in order to represent all the contents of the metaphor database (see Section 5.). After a brief discussion of remaining problems (see Section 6.), Section 7. concludes the paper.

2. WordNet and EuroWordNet in XML and RDF/OWL

The lexical database WordNet (WN) (Miller et al., 1990; Fellbaum, 1998) contains sets of word senses that are synonyms or near-synonyms of each other (synsets). We can also think of a synset as a concept and consider the word senses it contains as (largely interchangeable) linguistic expressions that can be used to refer to it. For example, the set of nouns in (1) is a synset in WordNet 3.0.

(1) {car:1 auto:1 automobile:1 machine:6 motorcar:1}

Between synsets, semantic relations such as hyponymy (subsumption) and meronymy (part-whole-relation) are defined. For this reason, WordNet has also been called a *lexical ontology*. Although centered around the synset notion, WordNet also includes additional *lexical* relations, defined between individual word senses instead of synsets; antonymy (the relation between opposition pairs) is an example.

EuroWordNet (EWN) (Vossen, 1999) is a multilingual lexical database built along the lines of the WordNet model. In addition to the central relations taken over from WordNet, EuroWordNet offers inter-lingual as well as further semantic relations. EWN data is distributed on CD-Rom in two formats: plain text and binary files. The binary data can be viewed within custom tools.

2.1. XML-based representations

XML representations of the EuroWordNet data along with tools to view and edit them (VisDic, DEBVisDic) were produced in the BalkaNet project (Horák and Smrž, 2004; Horák et al., 2006). The work on VisDic was motivated by several reasons; among others, distribution of the original tool for viewing and editing EWN data (Polaris) had long been discontinued, and a new platform was needed for the work on BalkaNet.

A different XML-based format for WordNet has been proposed by the Semantic Web community (van Assem et al., 2006). Their RDF/OWL representation can be queried and processed by standard Semantic Web tools, thus facilitating the integration of WordNet data into Semantic Web applications.

For representing EuroWordNet data, a multilingual RDF/OWL model of WordNet has been implemented and extended with cross-lingual relations and other EWN relations not defined in WordNet (De Luca et al., 2007a). After an analysis of some problems related to the main semantic limitations of WordNet, De Luca *et al.* (2007a) present the conversion of EuroWordNet into RDF/OWL-EuroWordNet and explain the inclusion of OWL ontologies under the RDF/OWL-EuroWordNet structure.

2.2. RDF/OWL EuroWordNet in the LexiRes tool

For convenient browsing and editing of RDF/OWL (Euro)WordNet, the LexiRes tool has been designed (De Luca and Nürnberger, 2006) and extended (De Luca et al., 2007b). In its word-based interface, the user can search for a word (e.g. English *bank*) in all its senses (19 senses in WordNet) and explore the synsets these senses belong to.

Each synset is shown as a box containing one or more word senses, and having an ID, a part of speech, and a gloss. For the retrieved synsets, WordNet relations to other synsets can be displayed graphically. There are two main motivations behind the LexiRes design:

- An increasing number of ontologies is available in OWL format, some of them containing large lexicons. These can be compared and merged within the LexiRes tool.
- WordNet tends to make fine-grained sense distinctions. For some applications, such as Information Retrieval, it might be advantageous to reduce the number of word senses to distinguish. LexiRes offers the possibility to merge synsets containing different word senses of the same word (according to WordNet), in two modes: manual or automatic.

The LexiRes tool thus supports authors in adding OWL ontologies to the RDF/OWL EuroWordNet representation or to manage the WordNet resources within other external OWL structures. In order to use this tool, an ontology has to be loaded into its scratch framework. The tool currently supports the EuroWordNet structure, but can easily be extended for other ontologies. The user can select the linguistic relations considered when displaying the (senses of the) query word within the lexicon structure, usually a hierarchy. After having selected them, the hierarchy will be displayed. Screenshots of the functionality of the LexiRes tool will be discussed further below (see Figures 4 and 5).

3. The Hamburg Metaphor Database and EuroWordNet

The present contribution describes the integration of additional lexical and semantic data into RDF/OWL EuroWordNet. The data originates from the Hamburg Metaphor Database (HMD), a relational database of French and German corpus attestations containing metaphors (Lönneker and Eilts, 2004; Lönneker-Rodman, *revs*). In HMD, each metaphor is manually analyzed and annotated at several levels: Among other lexical features, HMD provides references to EuroWordNet synsets; in addition, conceptual information is indicated in terms of domain labels from the Berkeley Master Metaphor List (Lakoff et al., 1991). The metaphor database is available for querying via a Web interface.

HMD is inspired by the conceptual theory of metaphor (Lakoff and Johnson, 1980). According to this framework, popular in Cognitive Linguistics, metaphors are not (primarily) creative, artful expressions. Rather, most of them are commonly used in everyday language, their main feature being that they systematically reflect relations (“mappings”) between concrete and abstract conceptual domains. As abstract domains are understood and talked about in terms of more concrete ones, metaphors illustrate the way we think and interact with the world around us.

For example, a metaphorical usage of French *trionphateur* - ‘triumphant winner’, is illustrated by the HMD example *le chancelier qui est donné comme le triomphateur* ‘the chancellor, who is proclaimed as the triumphant winner’.

```

<owl:ObjectProperty rdf:about="&ewn20schema;extMetaphorOf">
<rdfs:comment>This specifies that the first synset is a metaphorical extension
of the second synset. It should only hold between synsets of the same part of
speech.</rdfs:comment>
<rdfs:domain rdf:resource="Synset"/>
<rdfs:range rdf:resource="Synset"/>
<owl:inverseOf rdf:resource="extendedByMetaphor"/>
</owl:ObjectProperty>

```

Figure 1: The extMetaphorOf relation (RDF schema).

```

<rdf:Description rdf:about="&ewn20instances;synset-gagnant-noun-1">
<ewn20schema:extMetaphorOf rdf:resource="&ewn20instances;synset-maitre-noun-2"/>
</rdf:Description>

```

Figure 2: An instance of the extMetaphorOf relation.

There are two word senses of *trionphateur* in EuroWordNet: one of them belongs to a synset designating winners of war combats and other physical fights (*vainqueur:2 triomphateur:2 maître:2*); this is the synset containing the literal sense of the word. The second, more abstract, synset designates the concept of the most successful person in a non-physical competition or debate (*vainqueur:1 triomphateur:1 gagnant:1*); this synset groups metaphorical word senses. The Hamburg Metaphor Database records these synsets as “literal synset” and “metaphorical synset”. In terms of the conceptual theory of metaphor, the example illustrates the conceptual mapping between the domains of FIGHT and (POLITICAL) DEBATE, exemplified also by many other linguistic expressions.

Metaphors can be used consciously or unconsciously to express individual or culturally influenced attitudes and emotions. Therefore, the recognition of metaphorical senses of a word is useful for Natural Language Processing. Possible ways to add metaphor information to WordNet have been discussed previously (e.g., (Alonge and Lönneker, 2004)). With an RDF/OWL representation of EuroWordNet at hand, some of the information collected in HMD can now be converted into an add-on to the original EWN data in a relatively straightforward way (see next Section); other pieces of information require more sophisticated processing and post-editing of the data, as discussed in Section 5.

4. Representing metaphor relations and glosses in RDF/OWL EuroWordNet

To provide an RDF/OWL representation of HMD data, we started by defining a new relation between synsets, the conceptual relation *extMetaphorOf* (“extension by metaphor of...”). This conceptual relation holds between a synset with a metaphorical meaning and a synset with a literal meaning of at least one of the contained word senses. The relation as such is defined by an RDF schema (see Figure 1).

We then populated the *extMetaphorOf*-relation by deriving 107 instances from the HMD data for French. This was done by converting the data concerning attested

metaphorical mappings between EWN synsets from the HMD relational database into RDF. The 107 instances of the *extMetaphorOf*-relation thus represent cases where both the literal and the metaphorical synset were *already contained* in the original version of EuroWordNet. As with each relation in RDF/OWL EuroWordNet, the resulting information is stored in a separate RDF-file (*extMetaphorOf.rdf*) and can be distributed as such. An example of an entry in the *extMetaphorOf.rdf* file is given in Figure 2. The synset labeled *synset-gagnant-noun-1* has the ID 6432816 and contains the word senses *vainqueur:1*, *trionphateur:1*, and *gagnant:1* (this information is contained in separate RDF files). The synset labeled *synset-maitre-noun-2* contains the word senses *vainqueur:2*, *trionphateur:2*, and *maître:2*. The illustrated relation indicates that the first synset is a metaphorical extension of the second synset; in other words, the first synset refers to a more abstract concept than the second synset.

Each metaphorical synset is illustrated by at least one corpus-based usage example in the Hamburg Metaphor Database. These usage examples have been converted into language-specific glosses as a further add-on to RDF/OWL EuroWordNet. The example in Figure 3 shows the French gloss for the synset labeled *synset-gagnant-noun-1*; it includes all four example sentences registered by the metaphor database for this synset. This way, HMD data is used to add French glosses to EuroWordNet synsets. This is particularly useful because the original EWN data does not contain any French glosses. The derived glosses can be distributed as a separate RDF file as well.

5. Integrating HMD data into RDF/OWL EuroWordNet

The integration of HMD data into RDF/OWL EuroWordNet presents several challenges that we will address with the help of the LexiRes tool. Due to low coverage in EWN, synsets containing either the literal and metaphorical sense of a metaphorically used lexical item are not always available. If such a lexical item occurs in corpus data from the Hamburg Metaphor Database, the respective fields in the

```

<rdf:Description rdf:about="&ewn20instances;synset-gagnant-noun-1">
<ewn20schema:gloss>"le chancelier Kohl qui est donné comme le grand vainqueur";
"le chancelier qui est donné comme le triomphateur"; "le parti de Helmut Kohl qui
doit sortir demain comme le seul et le grand triomphateur"; "pas de surprise pour le
vainqueur, le parti de Monsieur Helmut Kohl"</ewn20schema:gloss>
</rdf:Description>

```

Figure 3: Example sentences as glosses of synsets with metaphorical meaning.

database table are left empty.

As a result, there are 459 HMD entries for French lexical items lacking an EWN synset with metaphorical meaning, but providing a EWN synset with literal meaning; these cover 136 distinct literal synsets and 128 distinct lexemes. In addition, 34 entries could not be linked to an EWN synset with literal meaning, but have been annotated with a metaphorical synset. This is the case with 16 distinct metaphorical synsets and 16 distinct lexemes. Finally, 280 HMD entries document metaphorical usages of a French lexical item lacking literal and metaphorical EWN synset information altogether.

5.1. Conversion of HMD data

We are planning to create “proto-synsets” (preliminary synsets) for each of the HMD entries lacking full synset information. We started converting those entries where one of the synsets is present in EWN and the other one missing. For each of the attestations in HMD documenting a word sense not covered by EWN, we create a new synset containing one (preliminary) word sense; these are proto-synsets. We will illustrate this at the example of the French verb *ébranler*, ‘shake’. One word sense of this word is already included in the EWN synset $\{\textit{ébranler}:1 \textit{secouer}:5\}$. This is the literal one, placed under the synset $\{\textit{bouger}:3\}$ ‘move’ in the EWN hypernym hierarchy. The metaphorical sense, as illustrated in the following Example (2), refers to an event of damaging or harming something abstract, such as values. This sense is not included in any EWN synset.

- (2) *Mais, pour cela, il faut coûte que coûte trouver une solution pour stopper l'hémorragie de citoyens qui vide le pays de ses forces vives, ébranle la crédibilité de l'Etat.* ‘But, to this end, it is necessary – no matter the cost – to find a solution that will stop the hemorrhage of citizens depleting the country of its driving forces, **shaking** the credibility of the State.’

When there is more than one attestation of a metaphorical sense in the metaphor database, several proto-synsets possibly containing the same (metaphorical) word sense are created. We can display all proto-synsets in LexiRes, inspect their glosses, and decide whether the examples document the same sense. If so, the proto-synsets can be merged within the LexiRes tool.

Figure 4 illustrates such an example for the word *ébranler*. The literal sense, included in an existing EWN synset, is displayed in the middle. Two proto-synsets with glosses are shown to the left and to the right. By accessing the context menu of one of the metaphorical proto-synsets, we can

merge it to another proto-synset. In most cases, all proto-synsets containing a sense of the same lexical item will be merged into one synset. It is only necessary to maintain more than one synset if there are different metaphorical extensions of the same literal expression. Once all attestations have been merged into as many synsets as necessary, clean word senses will have been achieved.

The merging procedure should then be extended to create new synsets containing several synonyms. If one of the participants in an *extMetaphorOf*-relation is an existing literal (or metaphorical) EWN synset, we should in fact inspect all new synsets related to it, each containing one word sense of different words, and see whether they are synonyms. If so, they can be merged. This way, the finalized new metaphorical (or literal) synsets will be created.

5.2. Visualization of HMD data with LexiRes

It is possible to load the *extMetaphorOf.rdf* file and the gloss file into LexiRes as an add-on to EuroWordNet. In LexiRes, the data can be queried and displayed graphically within the context of the EWN synset hierarchy. Figure 5 shows the LexiRes-RDF interface to RDF/OWL EuroWordNet. The language has been set to French, and the word *triomphateur* has been submitted as a search string. Two synsets containing senses of *triomphateur* have been returned and displayed as boxes in the main field of the interface. The synset with the literal meaning is shown at the left side, and the metaphorical one to the right. As illustrated by the screenshot, the user can select a synset (in this case, sense 1 of *triomphateur*) and access more functions by opening the context menu. For example, the RDF source code pertaining to the synset can be shown. The tool currently displays relations between the retrieved synsets and other synsets, as defined in EWN.

In Figure 5, hypernyms and hyponyms were retrieved. The new *extMetaphorOf*-relation, although loaded and accessible via the RDF code, is not yet displayed graphically. Given that it exists between synsets containing two word senses of the same word, we would like to represent it by a horizontal arrow. This would symbolize the mapping between two “parallel” or “analogous” domains, rather than a hierarchical notion as in hypernymy. The French glosses of metaphorical synsets have been loaded and integrated into the EWN data. Note the gloss of the metaphorical synset in Figure 5, corresponding to the RDF source in Figure 3.

6. Discussion

The *extMetaphorOf*-relation should be regarded as a *conceptual relation* that is not always reflected by polysemy of *all* the members of a synset. This means that it

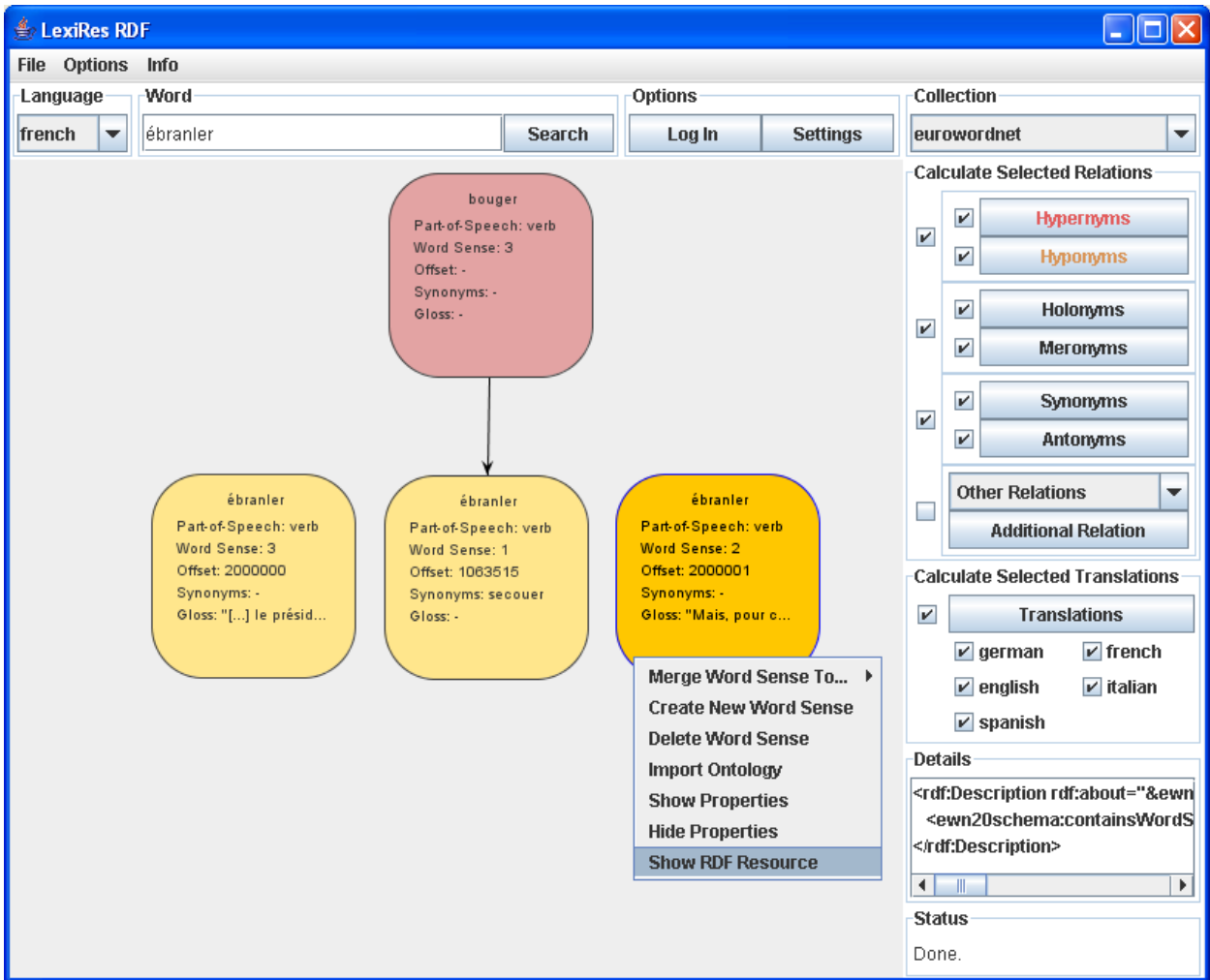


Figure 4: Existing EWN synset (center) and new proto-synsets with metaphorical meaning.

is not necessarily the case that *all* synonyms grouped in a literal synset have a (conventionally used) metaphorical extension. To represent this explicitly in the lexicon, lexical relations between literal and metaphorical word senses (instead of entire synsets), for example a relation called *derivedFromLiteral*, should be defined and added.

We can illustrate this discussion at the example of the synsets containing word senses of *trimphateur*, as introduced in Section 3., above. Lexical relations should exist between each of the senses of *vainqueur* - ‘winner’ and *trionphateur*, because metaphorical usages of both of them are attested in HMD. However, no lexical relations including *gagnant* ‘winner’ (a member of the “metaphorical” synset) or *maître* - ‘master’ (a member of the “literal” synset) can be postulated, based on the data.

The Hamburg Metaphor Database contains further information not yet converted into the RDF/OWL format. In particular, the database lists the conceptual domains (such as FIGHT and DEBATE) covering the mapping underlying a metaphorical expression. Future research will have to deal with the question how this can be represented in RDF/OWL EuroWordNet.

7. Conclusions

In this paper we presented a first integration of metaphor information into the RDF/OWL representation of EuroWordNet. We discussed the conversion of data from the Hamburg Metaphor Database into the RDF/OWL representation of EuroWordNet. We also explained some problems on the basis of examples.

The metaphor information was added to the available RDF/OWL structure of EuroWordNet by using LexiRes, a visualization tool developed and adapted for handling ontology structures. The tool supports users in navigating the hierarchy of an ontology or a lexicon and helps them restructure the resources by merging concepts (synsets), either manually or automatically.

A new relation describing metaphors (*extMetaphorOf*) and some of the French data from the Hamburg Metaphor Database, including illustrative usage examples, have already been successfully integrated into RDF/OWL EuroWordNet. Additional data is ready for merging within the LexiRes tool. Future work will deal with the conversion of further data types from the metaphor database.

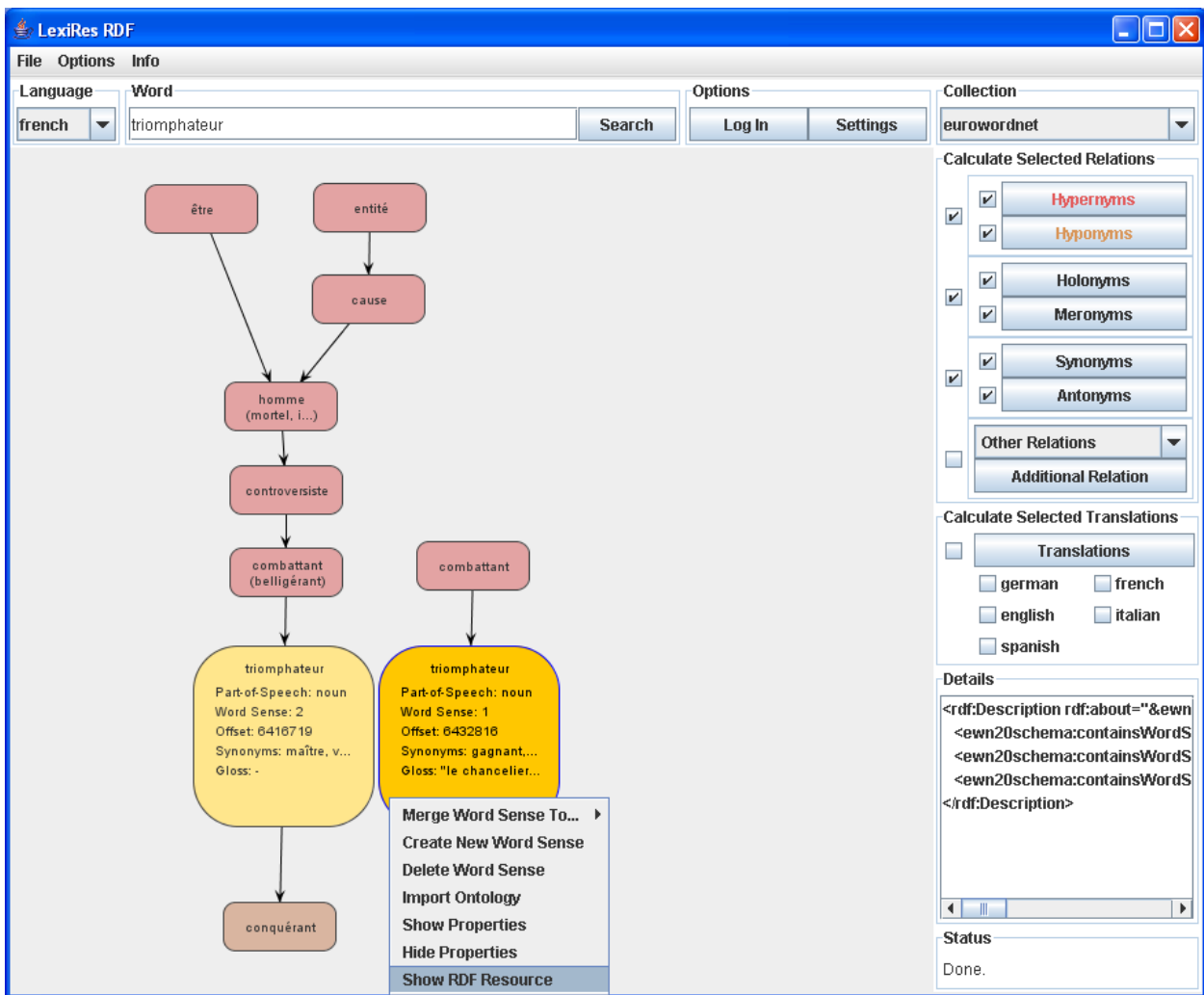


Figure 5: EWN synsets with literal (left) and metaphorical (right) meaning.

8. Acknowledgments

The authors would like to thank Thies Lönneker, University of Würzburg, Germany, who assisted with converting data from the Hamburg Metaphor Database into the RDF/OWL representation of EuroWordNet.

The second author is supported by a fellowship within the Postdoc-Programme of the German Academic Exchange Service (DAAD).

9. References

- Antonietta Alonge and Birte Lönneker. 2004. Metaphors in Wordnets: from theory to practice. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, pages 165–168, Lisbon, Portugal, May. ELRA.
- Ron Cole, Joseph Mariani, Hans Uszkoreit, Annie Zaenen, and Victor Zue. 1997. *Survey of the State of the Art in Human Language Technology*. Cambridge University Press.
- Hamish Cunningham. 1999. A definition and short history of Language Engineering. *Natural Language Engineering*, 5(1):1–16.
- Ernesto William De Luca and Andreas Nürnberger. 2006. LexiRes: A tool for exploring and restructuring EuroWordNet for Information Retrieval. In *Proceedings of the Workshop on Text-based Information Retrieval (TIR-06). In conjunction with ECAI 2006*, pages 12–17, Riva del Garda, Italy.
- Ernesto William De Luca, Martin Eul, and Andreas Nürnberger. 2007a. Converting EuroWordNet in OWL and extending it with domain ontologies. In *Proceedings of the Workshop on Lexical-Semantic and Ontological Resources. In conjunction with GLDV 2007*, Tübingen, Germany.
- Ernesto William De Luca, Martin Eul, and Andreas Nürnberger. 2007b. Multilingual query-reformulation using an RDF-OWL EuroWordNet representation. In *Proceedings of the Workshop on Improving Non-English Web Searching (iNEWS07). In conjunction with SIGIR 2007*, pages 55–61, Amsterdam, The Netherlands.
- Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. The MIT Press, Cambridge, Mass.
- Aleš Horák and Pavel Smrž. 2004. VisDic - Wordnet browsing and editing tool. In *Proceedings of the Second International Conference of the Global WordNet Associ-*

- ation (GWC 2004), pages 136–141, Brno, Czech Republic, January.
- Aleš Horák, Karel Pala, Adam Rambousek, and Martin Povolný. 2006. DEBVisDic - first version of new client-server Wordnet browsing and editing tool. In *Proceedings of the Third International WordNet Conference (GWC 2006)*, pages 325–328, Seogwipo, Jeju Island, Korea, January.
- George Lakoff and Mark Johnson. 1980. *Metaphors we live by*. University of Chicago Press, Chicago.
- George Lakoff, Jane Espenson, and Alan Schwartz. 1991. Master metaphor list. Second draft copy. Technical report, Cognitive Linguistics Group, University of California Berkeley. <http://cogsci.berkeley.edu>.
- Birte Lönneker and Carina Eilts. 2004. A current resource and future perspectives for enriching WordNets with metaphor information. In *Proceedings of the Second International Conference of the Global WordNet Association (GWC 2004)*, pages 157–162, Brno, Czech Republic, January.
- Birte Lönneker-Rodman. revs. The Hamburg Metaphor Database project: Issues in resource creation. *Language Resources and Evaluation*. Under revision.
- George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. 1990. Five papers on WordNet. *International Journal of Lexicology*, 3(4).
- Emanuele Pianta, Luisa Bentivogli, and Christian Girardi. 2002. MultiWordNet: developing an aligned multilingual database. In *First International Conference on Global WordNet*, pages 293–302, Mysore, India.
- Mark van Assem, Aldo Gangemi, and Guus Schreiber. 2006. RDF/OWL Representation of WordNet. Editor's Draft, 23 April 2006. <http://www.w3.org/2001/sw/bestpractices/wnet/wn-conversion.html>.
- Piek Vossen. 1999. EuroWordNet General Document. Version 3. Technical report, University of Amsterdam.