

A Current Resource and Future Perspectives for Enriching WordNets with Metaphor Information

Birte Lönneker and Carina Eilts

Institute for Romance Languages, University of Hamburg
Von-Melle-Park 6, D-20146 Hamburg, Germany
Email: birte.loenneker@uni-hamburg.de, carina.eilts@uni-hamburg.de

Abstract. This article deals with the question whether metaphors might be integrated into WordNets in a more systematic way. After outlining the advantages of having more information on metaphors in WordNets, it presents the Hamburg Metaphor Database and a possible method for integrating metaphors and corresponding equivalence relations into monolingual WordNets. Finally, problems are discussed that will have to be faced before more metaphor information could be included in WordNets.

1 Introduction

This article confronts the problem of how information on metaphors might be integrated into WordNets in a more systematic way. In order to decide what this means, certain theoretical assumptions have to be made. We adopt the viewpoint that in most cases, “basic” or “literal” senses of a word can be identified. We then assume that a literal sense can be the basis for different kinds of – attested or hypothetical – metaphorical senses. As pointed out by [1], current WordNets do not display information on the relationship between these different word senses in a systematic way. We furthermore follow a cognitive framework introduced by [2], according to which individual metaphorical word senses illustrate the mapping from a more *concrete* conceptual “*source domain*”, in which the corresponding literal sense is situated, to a more *abstract* conceptual “*target domain*”, in which the metaphorical sense is situated. Several other theoretical viewpoints could be adopted when dealing with metaphors; however, for practical tasks, it is necessary to choose one (main) theoretical framework as a starting point.

The practical task envisioned here consists in adding metaphor information to WordNets. Why and for whom this kind of information would be useful is outlined in Section 2. A resource that will facilitate the enrichment of WordNets with systematic information on metaphors is the Hamburg Metaphor Database, containing metaphorical example sentences in French and German and their annotations with EuroWordNet and conceptual domain data (Section 3). While building and using this resource, we developed ideas of how the actual WordNet enrichment could be performed, but we also detected some points that require clarification before this work can start (Section 4). Accordingly, Section 5 presents directions for future work.

2 Motivation

The fine sense distinctions made in WordNets have sometimes been criticized. However, in the case of metaphors, there are several reasons why a literal-figurative distinction is useful. Especially if this distinction is not only reflected in different synsets, but also documented by a relation between them and by information on the underlying domain mapping, metaphor information in WordNets can enhance a number of applications, for example:

- **Information Retrieval.** Information Retrieval would gain a lot from clearly identified metaphorical senses, because these senses are not valid for the parallel polysemy criterion (cf. [3]).
- **Word Sense Disambiguation.** Word Sense Disambiguation could be improved if lexical resources like WordNets provided senses and glosses for metaphors, enabling the automatic creation of semantically tagged corpora for machine learning (cf. [4]).
- **Language teaching.** Language teaching benefits from a domain-oriented view of metaphors; conceptually structured word/metaphor lists have proved to increase vocabulary retention (cf. [5]).

3 The Hamburg Metaphor Database

In view of the applications mentioned in Section 2 and inspired by work by Alonge and Castelli [1], the Hamburg Metaphor Database¹ (HMD) is being created in order to support studies of metaphors and WordNets. Based on domain-centered corpora, HMD provides both synset-oriented and domain-centered views on French and German metaphors, reachable online through a query interface.

The creation process of entries for the database can be briefly summarized as follows: Sentences or parts of sentences containing metaphors are extracted from a corpus and entered as “examples” into the database. The metaphorically used lexemes are identified in the examples and entered as “lexemes”. Each lexeme is looked up in the respective part of EuroWordNet (EWN) [6]. If the intended metaphorical sense is already encoded in EWN, the corresponding synset is entered into the “metaphorical synset” field, as in the French example in Table 1: For *naissance* ‘birth’, the synset *naissance:3* (glossed as “the time when something begins [...]; ‘they divorced after the birth of the child’ or ‘his election signaled the birth of a new age’”) allows a metaphorical reading. Synsets might also display an exclusively metaphorical sense of a lexeme, e.g. *father:5* ‘a person who holds an important or distinguished position [...]’. However, if a lexeme can only be located in a synset which is interpreted as showing its basic sense, the synset is entered as “literal synset”. Consider the German sentence in Table 1: The verb *verdunkeln* ‘to darken’ appears only in literal synsets; the selected transitive one, *vernebeln:1 verdunkeln:2* is glossed as “make less visible or unclear; ‘The stars are obscured by the clouds’”. Finally, in case the lexeme does not appear in any EWN synset, no synset information is encoded in HMD.

The next step consists in finding conceptual domain information for the metaphorical mapping that is documented by the metaphor, as outlined in Section 1. The “source” domain underlies the literal sense of the lexeme (for instance, BIRTHING for the lexeme *naissance*

¹ <http://www.rrz.uni-hamburg.de/metaphern> [30.08.2003]

Table 1. Selected data from the metaphor table in HMD

Lan-	Example	Lexeme	Meta- phorical synset	Literal synset	Source (Ber- keley terms)	Target (Ber- keley terms)
fr	A l'approche du conseil des 15 et 16 décembre à Madrid [...] Yves-Thibault de Silguy explique [...] que cette ré- union doit constituer l'acte de naissance de la monnaie unique	naissance nais- sance:3			BIR- THING	CREA- TING
de	Ein Aufklärer, der selber ver- dunkelt, ist unglaubwürdig.	ver- dunkeln		verne- beln:1 verdun- keln:2	DARK	BAD

and DARK for the lexeme *verdunkeln*, cf. Table 1), while the “target” domain is the one in which the metaphorical sense is situated (e.g. CREATING for *naissance* and BAD for *verdunkeln*). Two different naming systems for conceptual domains are used in HMD: The one of the Berkeley Master Metaphor List [7], and a proprietary German naming system, in which we add domain names missing from the Berkeley list.

User interfaces to the database allow for a query according to synsets, languages, domains, and corpora. The different corpora can be accessed by selecting one of the Master theses, in which the corpora were collected and documented. The Institute for Romance Languages in Hamburg currently disposes of 15 theses treating figurative language use in a cognitive linguistics framework.² Metaphors from six of these theses have been filed in HMD by August, 2003.

At the time of this writing, the database contains 394 corpus examples, documenting metaphorical uses of 300 distinct lexemes (138 in German, 162 in French). The French lexemes appear in 125 distinct synsets, 66 of them having a metaphorical meaning in EWN, and 59 displaying a literal meaning. The German lexemes appear in much less synsets; one of the reasons for this is that compounds were not split up into their parts. The database contains German synset annotations for 12 metaphorical and 29 literal synsets.

Although there is a large domain overlap in the French and German parts of HMD, the diversity of covered source and target domains is higher in the French part: 49 distinct source domains and 37 target domains have been identified for the French metaphors, while the German ones have been annotated as illustrating 22 source domains and 16 target domains. Metaphorical mappings “highlight” only certain aspects of the target domain which are seen in terms of the source domain [2]; therefore, several source domains might coexist for the same target domain and highlight different aspects: For instance, POLITICS (target) can be seen in terms of FIGHT, SPORTS, THEATER, or STUDY [8].

² Supervisor: Prof. Dr. Wolfgang Settekorn, French Linguistics/Media Science.

Several other databases and searchable lists of metaphors exist on the World Wide Web. For example, the ATT-Meta Project Databank³ developed by John Barnden contains examples of usage of metaphors of mind. The Berkeley MetaNet database MetaDB also includes domain information.⁴ However, to our knowledge, no other metaphor database explicitly includes WordNet information.

4 Towards a Systematic Metaphor Representation in WordNets

The current structure of EWN does not include a relationship which would allow the linking of metaphorical synsets to literal synsets. We therefore envision a method of adding new **eq_metaphor** relations at the level of a composite index, following [8]: A study of HMD example sentences and lexemes taken from several source and target domains led to the conclusion that a domain-centered view with a “central synset” referring to the overall source domain (an event like BIRTHING, in most cases), could be used as a starting point to semi-automatically add metaphorical synsets to existing WordNets. After manually connecting the central synset to its “parallel” metaphorical synset (containing identical literals with different indexes), parallel metaphorical synsets can automatically be created for all synsets that are connected to the central synset by a hyperonym, holonym, role – or possibly other – sense relationship. Glosses [9] for the new synsets could be created using templates to be filled with information like source synset and parts of glosses from the “central” source and target synsets (cf. Figure 1). A computer-assisted manual cleaning should be performed with special attention to those lexemes for which metaphorical senses already exist as synsets in EWN. These, as well as others actually *attested* in HMD, can be specially marked, in order to distinguish them from the remaining automatically created *potential* metaphors.

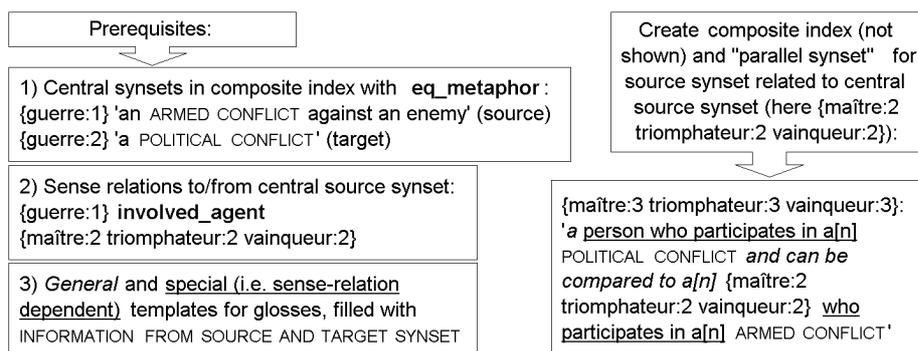


Fig. 1. Automatic creation of metaphorical synsets

³ <http://www.cs.bham.ac.uk/~jab/ATT-Meta/Databank/> [30.08.2003]

⁴ Personal communication from Michael Meisel, ICSI, Berkeley [4 September 2003].

In that way, gaps in EWN corresponding to empty synset fields in HMD would be filled. Still, other problems detected while building HMD need separate consideration and are summarized in what follows.

- **Missing glosses and scarcity of relationships.** Synsets in EuroWordNet do not always have glosses. If glosses are missing or incomplete, only the relations to other synsets might tell which sense is to be attributed to a synset. Given the small range of semantico-conceptual relations actually used in the French EWN – apart from hyperonymy, it contains only some antonymy and meronymy relations –, interpreting a synset is sometimes close to guessing.
- **Incorrect and incomplete synsets.** Incorrect synsets are rare, but they occur: e. g. French {père:2 parent:3 mère:2} ({father parent mother}). If ‘father’ and ‘mother’ were synonyms, they should be interchangeable in the same context without changing its meaning, which is not the case. Incomplete synsets are those from which at least one “literal” seems to be missing, as for instance the French synset {magazine:1 périodique:3} ‘magazine’; there is no obvious reason why the literal *revue* has been omitted.
- **Literal-figurative inconsistencies.** Sometimes, HMD encoders detect a synset with an apparently metaphorical meaning, showing semantico-conceptual relationships to clearly literal synsets (cf. also [8]). As long as metaphors are only documented in the database as explained above, this is not a crucial problem; however, as soon as one would like to create domain views and treat metaphorical mappings using a more or less automated procedure, these inconsistencies will result in errors.
- **Collocations and compounds.** Problems arise when (parts of) collocations or compounds bear a metaphorical meaning. Idioms (as a special case of collocations) are rarely represented in WordNets [10]; it is also difficult to individuate one single constituent in them displaying metaphorical usage, like in German *den Weg freimachen* ‘to clear the way’, French *mettre sur les rails* ‘to put on the rails’ – the whole idiom has a metaphorical meaning. For highly compounding languages like German, some compounds are represented as literals in EWN synsets, others not. Apart from the fact that the searched items might not be found in EWN, the ascription of domain mappings to whole compounds is problematic, because in general only one of the constituents is used figuratively (cf. German *Lügensumpf* ‘swamp of lies’, *Spendensumpf* ‘swamp of donations’ – only ‘swamp’ is metaphorical).

A more in-depth study on metaphors and WordNets, aiming at adding structured information on metaphors to WordNets using well-established EWN-means (composite index, synsets, relations and glosses), will thus have to take into consideration much more topics and issues than those directly related to metaphor.

5 Conclusion and Future Work

The encoding of systematic information on literal-metaphorical-relationships in WordNets necessitates careful analysis of the problems encountered, and step-by-step solutions. We hope to continue our work in two parallel lines:

1. Process those additional interpreted corpora that are available to the Hamburg Metaphor Database, in order to provide more material on metaphors, involved synsets and domain mappings.
2. For some selected source domains, create add-ons to the monolingual parts of EWN. If necessary, we will correct synsets of the source domain and complete the source domain structure by adding semantic relations, with the help of a tool for WordNet editing like VisDic [11] and taking into account further developed resources like GermaNet. Using the semi-automatic methods described above, metaphorical synsets and glosses will then be created. Finally, a script could integrate the add-ons into existing EWN-files.

References

1. Alonge, A., Castelli, M.: Which way should we go? Metaphoric expressions in lexical resources. In: Proceedings LREC-02, Las Palmas, Gran Canaria, 29–31 May. Volume VI. (2002) 1948–1952.
2. Lakoff, G., Johnson, M.: Metaphors we live by. University of Chicago Press, Chicago/London (1980).
3. Chugur, I., Gonzalo, J., Verdejo, F.: A study of sense clustering criteria for information retrieval applications. In: Proceedings OntoLex 2000, Sozopol, Bulgaria, 8–10 September. (2000).
4. Mihalcea, R., Moldovan, D.: An automatic method for generating sense tagged corpora. In: Proceedings AAAI '99, Orlando, Florida, 18–22 July. (1999) 461–466.
5. Boers, F.: Metaphor awareness and vocabulary retention. *Applied Linguistics* **21** (2000) 553–571.
6. Vossen, P.: EuroWordNet general document Version 3, Final (1999) <http://www.i11c.uva.nl/EuroWordNet/docs.html>.
7. Lakoff, G., Espenson, J., Schwartz, A.: Master metaphor list, Second draft copy (1991) <http://cogsci.berkeley.edu>.
8. Lönneker, B.: Is there a way to represent metaphors in WordNets? Insights from the Hamburg Metaphor Database. In: Proceedings of the ACL-03 Workshop on the Lexicon and Figurative Language, Sapporo, Japan, July 11. (2003) 18–26.
9. Pala, K., Smrž, P.: Glosses in WordNet 1.5 and their standardization/consistency. In: Proceedings of the LREC-02 Workshop on WordNet Structures and Standardisation, and How These Affect WordNet Applications and Evaluation, Las Palmas, Gran Canaria, 28 May. (2002) 20–23.
10. Fellbaum, C.: Towards a representation of idioms in WordNet. In: Proceedings of the COLING/ACL-98 Workshop on Usage of WordNet in Natural Language Processing Systems, Montreal, 16 August. (1998) 52–57.
11. Horák, A., Pala, K., Smrž, P.: Lexical semantic networks and ontologies in XML, their viewing and authoring. In: Proceedings of the CIC-03 Workshop on WWW Based Communities For Knowledge Presentation, Sharing, Mining and Protection, Las Vegas, Nevada, 25 June. (2003).