

The Hamburg Metaphor Database¹

Carina Eilts (carina.eilts@uni-hamburg.de) / **Birte Lönneker** (birte.loenneker@uni-hamburg.de)

Abstract

This article presents the Hamburg Metaphor Database project, an online database of French and German metaphors which came into being in 2002. In this database, metaphors appearing in different domain-specific corpora collected from mass media are available. The metaphors are annotated with lexical and conceptual information according to standard resources of the field: the EuroWordNet database for lexical information (synonyms) and the Berkeley Master Metaphor List for conceptual information (conceptual domains). The data collected can be explored for language studies and research via a WWW user interface without charge. It can be used for cross-language comparison of metaphors and the technical as well as the conceptual domains they occur in. We believe that it can also give indications on how lexical resources for Natural Language Processing could deal with metaphor representation in a better way.

Dieser Artikel stellt das Projekt der Hamburger Metapherndatenbank vor. Diese Online-Datenbank französischer und deutscher Metaphern wurde im Jahre 2002 ins Leben gerufen und stellt Metaphern zur Verfügung, die in verschiedenen sachgebietsspezifischen Korpora aus Massenmedien auftauchen. Die Metaphern werden mit lexikalischen und konzeptuellen Informationen nach entsprechenden Standards annotiert: Für die lexikalischen Informationen ("Synsets") wird die EuroWordNet-Datenbank herangezogen, für die konzeptuellen Informationen ("konzeptuellen Domänen") die in Berkeley entwickelte Master Metaphor List. Die gesammelten Daten können für Sprachuntersuchungen und Forschungszwecke kostenlos über eine WWW-Schnittstelle abgerufen werden. Vergleiche von Metaphern in verschiedenen Sprachen werden ebenso ermöglicht wie Untersuchungen der Sachgebiete und konzeptuellen Domänen, in denen sie auftreten. Zudem können die Daten als Anhaltspunkte zur Verbesserung der Behandlung von Metaphern in lexikalischen Ressourcen für die maschinelle Sprachverarbeitung Verwendung finden.

1. Aim

The aim of the Hamburg Metaphor Database research project is to make available via the World Wide Web metaphors collected by students for their theses (Master theses and German state examination theses). The theses have been written at the Institute of Romance Languages at Hamburg University since the year 1992. Example sentences are extracted from the corpora of these theses and annotated with lexical and conceptual information. The resulting data is stored in a database and it is available online without charge for purposes of language studies and research. The Hamburg Metaphor Database is reachable via <http://www.rrz.uni-hamburg.de/metaphern>.

This work has been inspired by research on the possibility of metaphor representation in lexical resources for Natural Language Processing (NLP) performed by Alonge/Castelli (2002a, 2002b). They show what kind of information on metaphors is already present in a specific lexical resource, the ItalWordNet, and discuss ways of how to encode more

¹ The authors would like to thank Antonietta Alonge and Wolfgang Settekorn for discussion.

information at various levels. Their basic claim is that not only individual metaphorical word senses should be encoded in NLP resources, but also the conceptual level on which certain regular or conventional metaphorical processes take place. By adding this kind of general information, it would be "possible to infer which words might potentially display a certain metaphoric extension" (Alonge/Castelli 2002b:1951).

The metaphors collected in our database together with the lexical information taken from a lexical database used in NLP, EuroWordNet (EWN), will help to determine some conventional metaphorical senses which should be added to existing lexical resources. Since conceptual information is added to our metaphor database as well, we believe however that its main advantage will be to indicate regularities in conceptual mapping. The data in two European languages, German and French, can further be compared to the results of other work, like the one by Alonge/Castelli, in order to determine whether a common European representation of metaphors in European lexical resources like EWN can be envisaged or not².

2. Resources

By now, our collection contains ten theses about metaphors which have been accomplished at the Institute of Romance Languages under the supervision of Prof. Dr. Settekorn. For their theses, the authors built corpora in French and German language, centered around certain subjects (e.g., political elections, football championship) and analysed the metaphors contained in the texts. The corpora have been extracted from different media (mainly newspapers, magazines and television) by the theses authors. The metaphors in their original immediate context (sentences or parts of sentences) are the input to our database.

For annotating the metaphors, we use two different resources: on the one hand, EuroWordNet (Vossen 1999) in its French and German version, on the other hand the Master Metaphor List compiled by the Berkeley Cognitive Linguistics Group (Lakoff et al. 1991). In the remainder of this section, we will briefly describe these two resources.

EuroWordNet is based on the Princeton WordNet (<http://www.cogsci.princeton.edu/~wn/>) in its 1.5 version (Vossen 1999:8). After the second project phase, which ended in 1999, EuroWordNet contained WordNets in the languages English, Spanish, Italian, French,

² The idea of a European imagery community ("europäische Bildfeldgemeinschaft") is brought forward by Weinrich (1976:287).

Estonian, Czech, Dutch and German. The main notion of a WordNet is that of a synset (Vossen 1999:5):

“A synset is a set of words with the same part-of-speech that can be interchanged in a certain context. For example, {car; auto; automobile; machine; motorcar} form a synset because they can be used to refer to the same concept.”

Between the synsets, there exist further language-internal relationships like hyperonymy and hyponymy. In EuroWordNet, synsets are linked indirectly from one language to the other by means of an Interlingual Index (ILI), which is an unstructured list of English synsets taken from WordNet 1.5 (cf. Vossen 1999:39). For instance, the synsets (1a) and (1b) both point to the same ILI-synset by synonym equivalence relations and are thus valid translations of each other. The same is true for the synsets (2a) and (2b).

(1a) French {univers: 1 nature: 1}

(1b) German {Natur: 2}

(2a) French {nature: 5}

(2b) German {Wesensart: 1 Wesen:1 Naturell:1 Natur:1 Gemütsart:1 Art:1}

In our work, we use the French and the German version of EuroWordNet. The construction of the French EuroWordNet was a joined work of the University of Avignon and the company Memodata (cf. Vossen 1999:4). It contains 17,826 noun and 4,919 verb synsets (Catherin 1999). The German WordNet was built at the University of Tübingen and contains 10,652 noun and 6,904 verb synsets (Kunze/Wagner 1999).

Our second resource, the Master Metaphor List (Lakoff et al. 1991), is a documentation of different kinds of conceptual metaphors. The metaphors are presented according to the cognitive metaphor theory presented in Lakoff/Johnson (1980) and are grouped into the four main sections of the document: EVENT STRUCTURE, MENTAL EVENTS, EMOTIONS and OTHERS (cf. Lakoff et al. 1991). These section headings have to be interpreted as abstract conceptual domains which can be represented by referring to other, more concrete conceptual domains. Conceptual domains are crucial for the understanding of metaphors, as they are instantiated by linguistic expressions in everyday language use. For example, the sentence

(3) She is of pure heart

can be expressed by the formula

(4) MORALITY IS PURITY

and has the conceptual source domain PURITY and the target domain MORALITY (cf. Lakoff et al. 1991:186).

3. Design

In this section, we explain how information is entered into our database. This description will help users to understand in more detail what kind of information is available in the Hamburg Metaphor Database. We concentrate on distinctions that were made while annotating the metaphors, insofar as they are reflected by the user interface for querying the database, or by the presentation of the retrieved data.

Our first step in creating a database entry is to copy an example sentence containing a metaphoric expression out of one of the theses mentioned in section 1. For example, we find the following sentence:

(5) le parti de Helmut Kohl qui doit sortir demain comme le seul et le grand *trionphateur* (transcription of the French television news magazine on channel TF1, 1st December 1990, 20h00)

In the extracted sentences, we individuate the term (or terms) that are used metaphorically. In sentence (5), this is the word *trionphateur*. Now we look for the synset containing this term in the EuroWordNet database. For our example sentence, we find the following synset in the French EWN:

(6) {vainqueur:1 triomphateur:1 gagnant:1}

From the English gloss provided by the ILI (cf. section 2), to which this French synset is linked, we can learn that it may be used in a metaphorical way, not being restricted to physical aggression only:

(7) [eq_synonym] the contestant who wins the contest

We can thus use this synset, which has or may have a metaphorical meaning, to annotate the term *trionphateur* in example sentence (5). However, in many cases, the only matching synsets we find in the EWN database are synsets by which the metaphorical use is not covered. In this case, we choose the best matching synset with the lexical meaning of the term in question for our database entry. For example, the following sentence

(8) Helmut Kohl, le *géant*, nouveau *maître* incontesté d'Allemagne (transcription of the

French television news magazine on channel TF1, 2nd December 1990, 13h00)

contains two terms which are used metaphorically (*géant*, *maître*). For the term *géant*, only the following synset is available in the database: {géant:1}. It has the English gloss

(9) [eq_synonym] an imaginary figure of superhuman size and strength; appears in folklore and fair [sic] tales

Since in the example sentence, Helmut Kohl is not really an imaginary figure that appears in folklore, this synset has to be coded as showing the literal meaning of the term. A synset with the metaphorical meaning, as it appears in our text, is missing from the EWN database. This is the reason why we have two columns for EWN synsets in our database: "literal" and "metaphorical" synset; results are presented accordingly when the database is used online.

The next step consists in creating labels for source and target domains. The target domain is the concept which is expressed metaphorically by means of the source concept. In general, the conceptual target domain is more abstract than the conceptual source domain. For example, in the metaphorical concept TIME IS MONEY, the abstract concept TIME is understood and experienced in terms of the more concrete concept MONEY, "a kind of thing that can be spent, wasted, budgeted [...]" (Lakoff/Johnson 1980:8). It is a convention to write these concepts in upper case letters.

In our database, two versions of labels for these concepts exist. The first label, a German name, can sometimes be taken over directly from the choice made in the theses, or otherwise is created by us. For instance, the metaphorical concepts underlying the meaning shift of the terms mentioned above (cf. examples 5 & 8) are expressed in (10) and (11), respectively.

(10) POLITIK IST KAMPF

(11) EINFLUSS IST GRÖSSE

In our database, POLITIK ("politics") can be found as target domain, and KAMPF ("fight") can be found as source domain.

In order to have a match with existing resources and naming conventions, we also provide labels of these metaphorical concepts in Berkeley terms, which we take over from Lakoff et al. (1991). In many cases, we do not find an exact match of our German labels in the Berkeley Master Metaphor List, but often a more general metaphorical concept is available and will be encoded in our database. English equivalents or generalizations of the labels mentioned above are the following:

(12) THEORETICAL DEBATE IS COMPETITION

(13) IMPORTANCE IS SIZE

A reason for the higher specificity of the German labels is the fact that the metaphors treated in the theses may be part of technical language, as in the context of political elections or decisions, or in the comments of football matches.

An important remark on our approach is that we regard it as being "top-down": It begins by collecting metaphors of a special "technical" domain or context. The metaphors thus belong to domain-specific corpora. This is strongly reflected by the conceptual domains involved: For some of them, many different metaphors can be found in the database. In this way, our approach differs from the one by Alonge/Castelli (2002a, 2002b), who analyse all occurrences of single words in a general Italian corpus, in order to find out which different metaphorical meanings these words might have.

4. Problems

Problems in building the database occur at all stages of the process described above (cf. section 3), which are: extraction of metaphors (cf. subsection 4.1.), annotation of EWN synsets (cf. subsection 4.2.), and labeling of conceptual domains (cf. subsection 4.3.).

4.1. Selecting metaphors, or: What is (not) a metaphor?

The first problem we encounter is that of different conceptions of metaphor. For example, the authors of the master theses in general also include what we would like to call personifications, metonymies and idioms into their corpora.

For Cruse (1986:37), an idiom is "a lexical complex which is semantically simplex", or – in more traditional terms –, "an idiom is an expression whose meaning cannot be inferred from the meaning of its parts" (Cruse 1986:37). Idioms are a problematic case, because they are not productive any more; they cannot be "revived" any more, as can be "dead" metaphors: "dead metaphors can be 'revived' by substituting for one or more of their constituent parts elements which are near-synonyms, or paraphrases" (Cruse 1986:42). However, in the Berkeley metaphor list (Lakoff et al. 1991), we also find examples of idioms like "*He has a screw loose*" (cf. Lakoff et al. 1991:138). This is why we may, in some cases, also include examples containing idioms in our database.

However, we try to avoid including examples of personifications and metonymies, because

the underlying semantic processes are different from those of metaphor. Metonymic concepts are already related to each other in the external world: "[metonymy] relies on extralinguistic world knowledge" (Blank 1999:170). For this reason, Lakoff/Johnson (1980) describe metonymies as being more "obvious" relationships between concepts than metaphors are: "[...] [t]he grounding of metonymic concepts is in general more obvious than it is the case with metaphoric concepts, since it usually involves direct physical or causal associations" (Lakoff/Johnson 1980:39). For example, in the sentence "*the buses are on strike*" (cf. Lakoff/Johnson 1980:38), the OBJECT USED stands metonymically for its USER, the driver. Metonymies resulting in regular polysemy are already covered by EuroWordNet by a special composite Interlingual Index (Vossen 1999:40-43). This is why examples of metonymies will not be included in our database.

For Lakoff/Johnson (1980:33), personifications are a kind of ontological metaphor, which provides an understanding of abstract experiences in terms of concrete objects and substances (cf. Lakoff/Johnson 1980:25). In the corpora of the theses, we find many examples of personifications of countries like France or Germany, in which the countries are viewed as entities performing human actions and having human feelings (e.g., "*faire gagner la France*" [in a political context]; "*la France joue comme il faut jouer*" [in a football match]). It is a question whether this case should be regarded as a proper personification, or whether it should count as a metonymy of the type "A COUNTRY FOR THE PERSONS LIVING IN IT." In either case, we will not include these examples in our database.

4.2. EuroWordNet problems

One of the problems of the EuroWordNet database has been explained above (cf. section 3), when we have been treating missing metaphorical senses. Another problem is that adjectives have not been included in EWN (cf. Vossen 1999:6), so we cannot name a synset when an adjective is used metaphorically in the corpus. Sometimes, when we are confronted with polysemous nouns or verbs, it is also difficult to choose the right synset from the English gloss of the related ILI, especially if synsets consist of only one word, as for example

(14) {perte:1} [eq_synonym] something lost especially money lost at gambling

(15) {perte:3} [eq_synonym] something that is lost

More precise glosses, preferably in the language of the respective synset, would be extremely helpful in such a case. As long as they are not available, we follow the hyperonym links of problematic synsets in order to find out more about their meaning, and generally get enough

information this way to be able to decide which synset to use for annotation.

Another problem is the treatment of "creative" compound nouns in German, in which one of the components can be found in its literal meaning in EuroWordNet, but the composed metaphorical expression is not available in any synset. Examples taken from our corpora are *Spendensumpf* and *Lügensumpf*. The noun *Sumpf* (swamp) is present in the synset {*Sumpf:1*}, while the metaphorical compounds *Spendensumpf* ("swamp of donations") and *Lügensumpf* ("swamp of lies") are not represented in EWN. The question is how to treat these compounds: Should they be entered as one term? This is what we are doing now. However, it might turn out to be a better idea to only enter the base noun (e.g. *Sumpf*) as a term, so that a synset would often be available. A third solution is to keep track of both constituents of the compound noun; but in this case, a different status (that is, a different field in the database) would have to be assigned to the modifier parts (e.g. *Spende*, *Lüge*), as these terms usually keep their literal meaning.

Sometimes, spelling mistakes occur in the EuroWordNet synsets. An example is {*combat:2* *bagare:1* *bataille:4* *lutte:1*} from the French EWN, where *bagarre* should in fact be spelled with double *r*. In order to make it possible for other EWN-users to find the synsets we enter and to compare resources, we preserve these spelling mistakes.

4.3. Problems with labels for conceptual domains

We have explained in section 3 that our German concept domain labels are sometimes more specific than labels from the Berkeley Master Metaphor List (Lakoff et al. 1991). Another problem is that some metaphorical concepts we find in our corpora are entirely missing from the Berkeley list. This is especially the case for some social groups which are conceptualized in terms of another social group, like "A POLITICAL PARTY IS A FAMILY". When we do not find any matching label in the Berkeley metaphor list, this field of our database is left empty.

5. Current status and future work

By October 2002, around 160 examples have been extracted from three theses, annotated and entered into the Hamburg Metaphor Database. Sometimes, the same term with the same metaphorical meaning occurred quite often in the corpus; in this case, we decided not to enter all the examples from the corpus, which would lead to repetition. For the moment being, the metaphor database has thus to be seen as a qualitative rather than a quantitative resource. In

addition to the information described above, all example sentences have been provided with information about the language they are written in, the Institute of Romance Languages theses they are treated in (their authors and titles), and their original source (bibliographic information on newspaper articles or other sources).

An online interface has been implemented which allows users to query the database. Two different kinds of queries are possible: In the first one, displayed in figure 1, the user can choose conceptual domains or synsets (or combinations of these) for which he wants to display all examples found in all theses; using the other one, all examples from one thesis can be viewed. The second choice results in a more domain-centered view, because the theses are centered around certain topics, as explained in section 2.

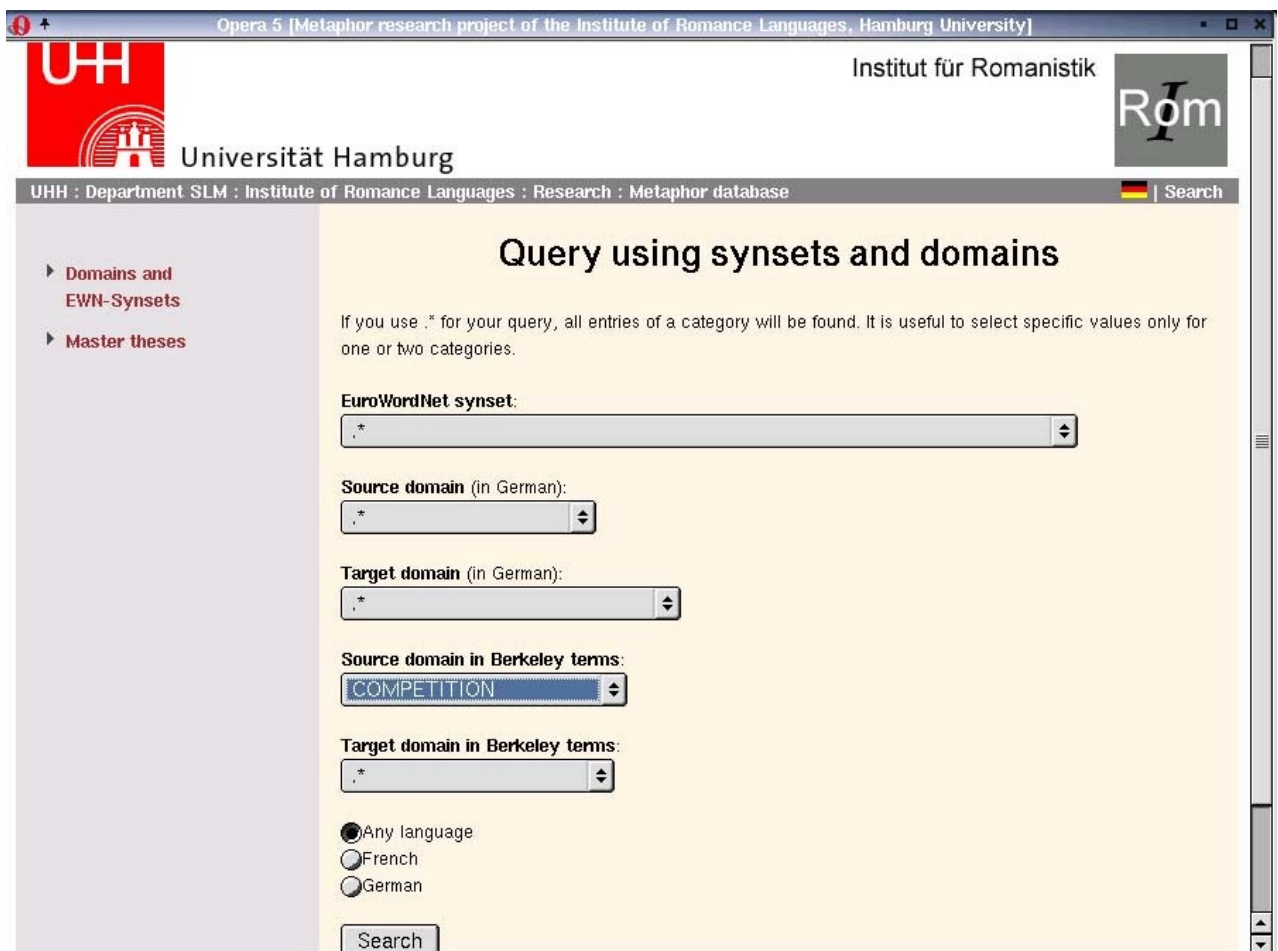


Figure 1: One of the online query interfaces to the Hamburg Metaphor Database.

Our first goal is to store the information from all ten theses which have been written so far at the Institute of Romance Languages (cf. section 1). In this first step, we have to ignore some of the problems mentioned above, like missing synsets or conceptual domain labels.

In a second step, a revision of the whole database will be performed adding the missing

information in a consistent way. Regarding the synsets, the German part of EWN is in fact based on the GermaNet project (<http://www.sfs.uni-tuebingen.de/lsd/Intro.html>) (cf. Wagner/Kunze 1999), which has been further developed since. For our purposes, it might well be possible to integrate synsets from that resource even if the structure of the GermaNet and the EuroWordNet databases are not entirely the same. However, it is not clear how these synsets might be treated in an analysis which would take into account EWN structure. As far as the conceptual domain labels are concerned, a closer analysis of those metaphors for which they are missing will be necessary in order to create consistent new labels.

Finally, the collected data will be compared to data for other languages (e.g. Alonge 2002a, Alonge 2002b) in order to determine which kinds of representation might be adequate for metaphorical expressions in NLP resources like WordNets.

Bibliography

- Alonge, Antonietta/Castelli, Margherita (2002a): "Metaphoric expressions: an analysis of data from a corpus and the ItalWordNet database." In: *Proceedings of the First Global WordNet Conference. Mysore, India*. Mysore: Central Institute of Indian Languages. 342-350.
- Alonge, Antonietta/Castelli, Margherita (2002b): "Which way should we go? Metaphoric expressions in lexical resources." In: *Proceedings of the third Language Resources and Evaluation Conference. Las Palmas, Gran Canaria*. Paris: European Language Resources Association. VI: 1948-1952.
- Blank, Andreas (1999): "Co-presence and Succession. A Cognitive Typology of Metonymy." In: Panther, Klaus-Uwe/Radden, Günter (edd.): *Metonymy in Language and Thought*. (= Human Cognitive Processing, 4.) Amsterdam/Philadelphia: John Benjamins. 169-191.
- Catherin, Laurent (1999): *The French Wordnet*, EuroWordNet (LE-8328) Deliverable 2D014 Part B2. Available <http://www.illc.uva.nl/EuroWordNet/docs.html> [15. September 2002]
- Cruse, D. Alan (1986): *Lexical semantics*. Cambridge et al.: Cambridge University Press.
- Kunze, Claudia/Wagner, Andreas (1999): *The German Wordnet*, EuroWordNet (LE-8328) Deliverable 2D014 Part B1. Available <http://www.illc.uva.nl/EuroWordNet/docs.html> [15. September 2002]
- Lakoff, George/Johnson, Mark (1980): *Metaphors we live by*. Chicago/London: University of Chicago Press.
- Lakoff, George/Espenson, Jane/Schwartz, Alan (1991): *Master metaphor list. Second draft copy*. Cognitive Linguistics Group. University of California Berkeley. Available <http://cogsci.berkeley.edu> [15. September 2002]
- Vossen, Piek (1999): *EuroWordNet General Document*. Version 3, Final. University of

Amsterdam. Available <http://www.illc.uva.nl/EuroWordNet/docs.html> [15. September 2002]

Wagner, Andreas/Kunze, Claudia (1999): "Integrating GermaNet into EuroWordNet, a Multilingual Lexical-Semantic Database." In: *Sprache und Datenverarbeitung* 23(2), 5-20.

Weinrich, Harald (1976): *Sprache in Texten*. Stuttgart: Klett.