In: Jochen Heinsohn & Bernhard Hollunder (eds.) DFKI Workshop on Taxonomic Reasoning: Proceedings Saarbrücken, 1992.

Natural Language Disambiguation and Taxonomic Reasoning

John Nerbonne *
Deutsches Forschungszentrum für Künstliche Intelligenz
DISCO and ASL Projects
Stuhlsatzenhausweg 3
D-6600 Saarbrücken 11, Germany
nerbonne@dfki.uni-sb.de

March 6, 1992

Précis: Among the expertises relevant for successful natural language understanding are grammar, semantics and background knowledge, all of which must be represented in order to decode messages from text (or speech). In this extended abstract we examine disambiguation—the choice of one of several hypotheses about the meaning of an input string—as an example of the current cooperation of formalisms. It is quite tempting to provide for a cooperation of background knowledge (represented in taxonomic logic) with grammar representations—and this is an example of heterogeneous formalisms in cooperation. The integration of taxonomic and grammatical information has the advantage of allowing a very close integration of background information with grammar—and therefore an early elimination of some analysis hypotheses, but we argue here that the approach also has problems which could be remedied by a closer integration with semantics. The problems are (i) the range of taxonomic reasoning possible in grammar formalisms, and the apparent usefulness of a more sophisticated reasoning in disambiguation; (ii) the need to incorporate information from discourse content in disambiguation; and (iii) the need to distinguish presupposed from asserted information, which calls for mechanisms familiar in meaning representations but otherwise unneeded in grammar formalisms.

Background

Background for some of the remarks below is a proposal for integrating semantics and grammar formalisms found in Shieber 1986, Pollard and Sag 1987, Fenstad et al. 1987,

*Thanks to Rolf Backofen for discussion of the ideas presented here. This work was supported by a research grant, ITW 9002 0, from the German Bundesministerium für Forschung und Technologie to the DFKI DISCO project.

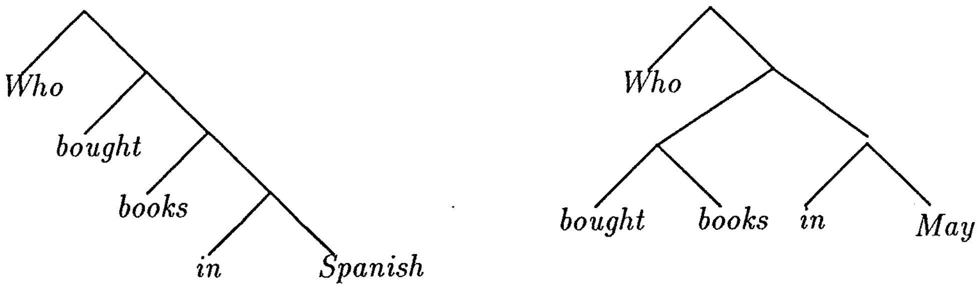

Figure 1: Disambiguation may even require the recognition of distinct constituent structures. Note that the proper names *Spanish* and *May* are not syntactically distinct, nor do that belong to distinct logical types—each denotes an entity. But they do denote objects of different SORTS, as this term is used in sortal logic, since May is a time and Spanish a language. The link to knowledge representation, especially of the kind encoded in KL-ONE, is justified by the emphasis on sorts—taxonomies, or conceptual hierarchies—which distinguishes knowledge representation schemes such as KL-ONE.

and Moore 1989, and elaborated in Nerbonne 1992a, Nerbonne 1992b, and Nerbonne 1992c. The heart of the proposal is a scheme for using feature formalisms as formalized metalanguages for semantic representation languages. In Nerbonne 1992a it is argued that this division between object language and metalanguage specifications is necessary for the characterization of linguistic ambiguity and that it provides a foundation for disambiguation.

Disambiguation

DISAMBIGUATION is the process of determining (i) which of potentially many meanings was intended in an utterance, but also (ii), with respect to a particular application, which facet is relevant to an NL interaction. The former is a response to the ambiguity of natural language, while the latter exists even where no genuine ambiguity does. We illustrate these in turn below. Disambiguation occupies computational linguists more than theoretical linguists, and is extremely important in applications in which there many be uncertainty about input—e.g., speech. Like parsing itself, at least some disambiguation seems to be automatic, so that untrained speakers are not aware of needing to disambiguate structures. The example below, graphed in Figure 1, suggests how unobtrusive the process is:

(1) a. Who bought books in Spanish?
    b. Who bought books in May?

This sort of example is convenient because it shows how pervasive the effects of disambiguation may be—reaching even into the parsing component. It is simultaneous misleading if it suggests that genuine disambiguation tasks need to be accompanied by such striking consequences. For even if disambiguation MAY be accompanied by striking consequences in application independent ways, the need for disambiguation arises in NLP interface efforts in ways that need have no purely linguistic ramification whatsoever. In particular, NL interfaces need to be sensitive to application distinctions which do not correspond to natural language ambiguities.

Consider the DISCO application, that of consulting with multiple agents who plan shipping. Here the phrase *Schmidts Ladung* 'Schmidt's freight' certainly denotes freight which stands in some relation to Schmidt. For example, we may imagine the freight contracts in the application as organized into a small database, where the freight contract is the basic tuple.

Order Nr.	Contractor	Agent	Destination	?
457	Schmidt	...	...	...
574	...	Schmidt	...	...
745	...	...	Schmidt	...
475	...	...	...	Schmidt

Thus the phrase *Schmidts Ladung* could designate freight which Schmidt contracted to have shipped, freight for which he is the freight agent, freight being sent to him, and perhaps even freight which stands in yet another relation to him (as owner, inspector, as packer, etc.). Now it is unlikely that the relation expressed by the German possessive construction (genitive + $\bar{\mathrm{N}}$) is ambiguous, and it is unthinkable that the construction is ambiguous to just this degree and in just this fashion.

Taxonomic reasoning of the KL-ONE variety (Brachman and Schmolze 1985, Baader and Hollunder 1990) may fruitfully be applied both to the resolution of linguistic ambiguity and to resolution of application-specific distinctions (Bobrow 1979). We consider these in turn. At the heart of taxonomic reasoning is the imposition of a sort hierarchy on the domain, illustrated in Figure 2 for the case where we found linguistic ambiguity.

In addition to the provision of a sort hierarchy, sortal disambiguation requires a characterization of which sorts are appropriate for which (argument positions of) relations. We would then allow that *in* translate into (at least) two relations, one temporally relating eventualities to times, and the other relating documents to media (but not to times). Schematically:

Lexical Item	relation	Argument1-Sort	Argument2-Sort
in	temporal-loc	Eventuality	Time
	use-medium	Document	Medium

Finally, we must enforce the sortal compatibility restrictions. For many applications, it is desirable to enforce these as early as possible, so that unnecessary processing is avoided. Of course, this mechanism takes us outside KL-ONE proper, but the require information—

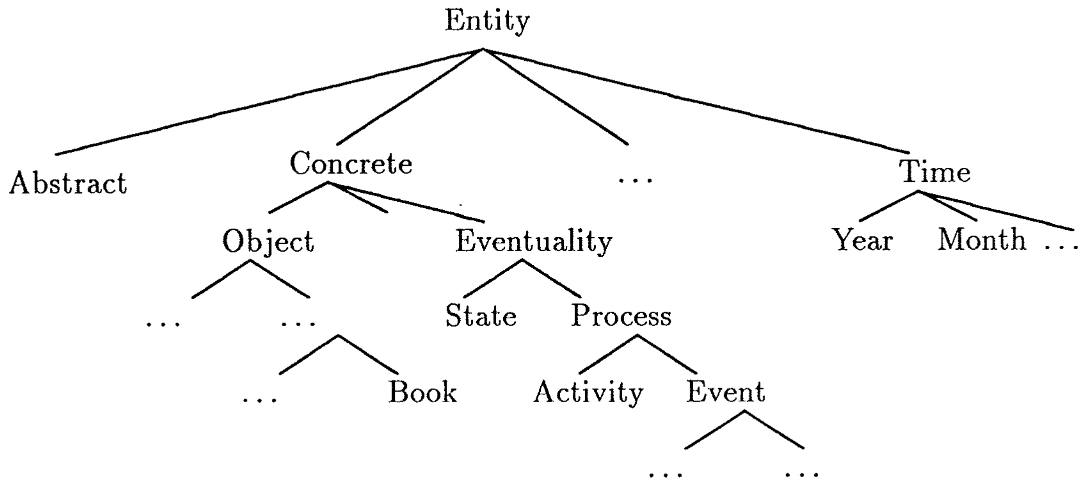

Figure 2: A sort hierarchy which distinguishes enough classes in the domain to illustrate sortal disambiguation for the sentences in (1) (in the text). The hierarchy here is better portrayed not as a tree, but as a directed graph—in which sorts inherit on more than one path back to the root. For example, states and activities might best be viewed as subsorts of both eventualities and nondiscretes, which would include physical substances (water, flour) as well. The move to non-tree-structured hierarchies does not affect the points here however, about the interaction of grammar, semantics and knowledge representations.

compatibility is efficiently provided in KL-ONE-like systems. As the example in Figure 1 suggests, an enforcement of sortal compatibility as early as parse time would be useful (and recall that, e.g., speech applications will rely on disambiguation to prune unlikely hypotheses). This raises the question of how well these constraints can be integrated into other processing—which of course depends on whether they can be expressed in the formalisms of other modules. Here then is a concrete instance of the question of how one relates knowledge representation to grammatical and semantic formalisms.

Earlier work in DISCO demonstrates that semantics can be formulated in an indirect fashion in feature formalisms, so we shall show here that the same is true of knowledge representation—at least within bounds. Cf. Moens et al. 1989 for an earlier proposal along the same lines. That is, once we've taken the step to representing the semantics of *in* in a typed feature description language:

$$\begin{bmatrix} \text{PRED} & \textbf{temporal-location} \\ \text{THEME} & \\ \text{LOCATION} & \end{bmatrix}$$

Then we can also represent the sortal information, relying on unification to enforce sortal compatibility, and thus integrating sortal disambiguation with the unification used in parsing. The following feature structure description represents the ambiguous lexical item *in*:

$$\left[\begin{array}{ll} \text{FORM} & \text{in} \\ \text{SEMANTICS} & \left\{ \begin{array}{c} \left[\begin{array}{ll} \text{PRED} & \textbf{temporal-location} \\ \text{THEME} & \left[\ \textit{Eventuality}\ \right] \\ \text{LOCATION} & \left[\ \textit{Time}\ \right] \end{array}\right] \\ \vdots \\ \left[\begin{array}{ll} \text{PRED} & \textbf{expressive-means} \\ \text{THEME} & \left[\ \textit{Document}\ \right] \\ \text{LOCATION} & \left[\ \textit{Medium}\ \right] \end{array}\right] \end{array}\right\} \end{array}\right]$$

The representation for the word *May*, whose semantics is of the sort Month, and therefore also of the sort Time, can successfully unify with the (location argument of) the first alternative semantics for *in*, but not the last, for which an argument of the sort Medium is expected. Thus the PP *in May* is a seeks to attach where its first argument will be of the sort Eventuality—and this can be a VP attachment, since VP's denote eventualities, but not an NP with the head noun *books*, since this denotes objects of an incompatible sort.

Although we shall not present the details of the treatment of the resolution of application-specific distinctions, it should be clear that the same techniques apply. In the example *Schmidts Ladung*, the relation between Schmidt and the freight is potentially disambiguated by information about whether Schmidt is a shipper, a customer, or the recipient of a customer's shipment. Nor shall we attempt on the basis of this example to argue that sortal restrictions must come from the domain and NOT from the lexicon—the dilemma seems spurious, since the lexicon must in some way be accommodated to the domain for serious applications anyway. Cf. Iida et al. 1989 on the relation between lexicon and disambiguation in complex applications.

0.1 Emerging Issues in Disambiguation

Thus the feature formalism allows the integration of constraints from knowledge representation as well. There are several qualifications needed, however. First, the feature formalism cannot faithfully represent all of the sorts of richer KL-ONE-like languages, in particular not those which allow quantified sort defintions, e.g., definitions such as:

$$\text{Parent}(x) \leftrightarrow \exists\, y.\text{Child-Of}(y, x)$$

Some KL-ONE derivatives allow these without relinquishing decidability, but they are not foreseen in feature formalisms. On the other hand I am unsure of how important this sort of example is—i.e., how frequently one must appeal to sorts of this complexity.

A second qualification is that this sort of treatment will not allow the enforcement of constraints which derive from inferences based on earlier utterances—in order to accomplish this a genuine integration into the semantics representation language would be required. We have in mind the kind of inference possible when information about an individual accumulates during the course of a conversation, but which may be demonstrated even in a single sentence:

(2) Sam talked for two hours in the library and read books for one.

The interesting phrase is the *for one*, and the interesting question is how we account for its VP attachment. Of course, the sortal explanation is available—we simply postulate that *for* denotes a relation between eventualities and durations, but that it denotes no relation between books and durations. But this information cannot be available on the basis of the lexical item one—it must be inferred on the basis of previous content (and the anaphoric link).

One can, as always, attempt alternative explanations, but the ones which immediately come to mind are unconvincing. One could postulate that the choice of attachment site depends on a parallelism to the first clause, but that is not necessary (a). Or one could hypothesize that the VP atachment is strongly preferred. But the $\bar{\mathrm{N}}$-structures of the form $\bar{\mathrm{N}}+$ PP-for are quite possible (b):

(3) a. Two hours elapsed. Sam read books for one, and daydreamed for the other.
b. Sam looked for gifts for his kids. He saw <u>books for one</u>
and <u>T-shirts for the other</u>.

Thus we conclude that a proper account of disambiguation should go beyond the encoding in feature structures illustrated above, and that a more thoroughgoing integration of semantic representation and disambiguating mechanisms is ultimately required. The presentation above stills shows how a great deal of disambiguating information can be integrated into the feature systems and thus arbitrarily deep into a modern NLP system, even if it turns out to be incomplete.

A third qualification about the usefulness of feature-based disambiguation concerns a fundamental pitfall of sortal disambiguation, i.e., that it needs to distinguish between asserted and presupposed sortal information. This is quite clear in the case of application-specific distinctions, and arguably necessary for linguistic ambiguities as well. We examine the case of application-specific distinctions first. We argued above that *Schmidt's freight* might be understood on the basis of a variety of application-specific relations, including 'freight-shipper', 'freight-contractor', etc. In deciding which of these is relevant, it is legitimate to examine the sort to which Schmidt belongs (shipper, contractor, etc.). But notice that in a phrase such as *Schmidt's freight* the relation between Schmidt and the freight is presupposed, not asserted. It would clearly be wrong to apply disambiguation techniques to cases where the relation is asserted or questioned, but not presupposed, e.g., in *Is this freight Schmidt's?* or *If Schmidt sends freight, then his freight will arrive* *today*. (In the latter case, one can imagine blithely disambiguating *his freight* to the 'freight-recipient' relation on the basis of Schmidt's being listed only as recipient—but this would clearly lead to errors.) The case of genuine linguistic ambiguity is similar, but arguably different, in that sortal mismatches remain peculiar enough even in assertion to warrant perhaps being categorized as ill-formed. This tack would regard the following as ill-formed:

The book is for one hour.
Is the book for one hour?
The book cannot be for one hour.
The book is in May.
Is the book in May?
It would be impossible for the book to be in May.

While the examples are undoubtedly peculiar, it still seems wrong to regard them as ill-formed as opposed to unusually formulated or simply concerned with unusual circumstances. Presented with a sentence such as one of these, it would seem that the appropriate reaction would be to try to interpret it metaphorically or, if possible, to clarify it with a user. This conclusion suggests that both sorts of disambiguation—that of resoving genuine ambiguities and that of resolving application-specific distinctions—benefit from the distinction between assertion and presupposition, which therefore ought to be part of a comprehensive disambiguation scheme. We would propose an integration into a logic for presupposition of the sort proposed for dynamic logic (Beaver 1992) or discourse representation theory (van der Sandt and Geurts 1992).

Summary and Conclusions

The purpose of the present abstract has been the investigation of the use of taxonomic information for the purpose of disambiguation. We have argued that taxonomic information is clearly very useful in disambiguation, and that, while it can be incorporated into feature formalisms (allowing an integration of disambiguation with feature unification), still there are several indications that a more thorough-going integration with semantic representation would be sensible.

References

Baader, F., and B. Hollunder. 1990. KRIS: Knowledge Representation and Inference System—System Description. Technical memo, Deutsches Forschungszentrum für Künstliche Intelligenz, Kaiserslautern.

Beaver, D. 1992. Presuppoposition in Update Semantics and 3-Valued Logic. In *Proceedings of the 8th Amsterdam Colloquium*, ed. P. Dekker and J. van der Does, xxx–yyy.

Bobrow, R. 1979. Semantic Interpretation in PSI-KLONE. Annual Report 4274, Bolt, Beranek and Newman, Cambridge, MA.

Brachman, R. J., and J. G. Schmolze. 1985. An overview of the KL-ONE knowledge representation system. *Cognitive Science* 9:171–216.

Fenstad, J. E., P.-K. Halvorsen, T. Langholm, and J. van Benthem. 1987. *Situations, Language, and Logic.* Dordrecht: Reidel.

Iida, M., J. Nerbonne, D. Proudian, and D. Roberts. 1989. Accommodating Complex Applications. In *Proceedings of the First International Workshop on Lexical Acquisition, IJCAI 1989*, xxx–yyy. Menlo Park. Morgan Kaufman.

Moens, M., J. Calder, E. Klein, M. Reape, and H. Zeevat. 1989. Expressing Generalizations in Unification-Based Grammar Formalisms. In *Proceedings of the 4th Meeting of the European Chapter of the Association for Computational Linguistics*, 174–181.

Moore, R. C. 1989. Unification-Based Semantic Interpretation. In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, 33–41.

Nerbonne, J. 1992a. Constraint-Based Semantics. In *Proceedings of the 8th Amsterdam Colloquium*, ed. P. Dekker and J. van der Does, xxx–yyy.

Nerbonne, J. 1992b. A Feature-Based Syntax/Semantics Interface. In *Proceedings of the Second International Conference on the Mathematics of Language*, ed. A. Manaster-Ramer and W. Zadrozny, xxx–yyy. Annals of Mathematics and Artificial Intelligence.

Nerbonne, J. 1992c. Representing Grammar, Meaning and Knowledge. In *Proceedings of the Berlin Workshop on Natural Language Processing and Knowledge Representation*, ed. S. Preuss and B. Schmitz. Technische Universität Berlin. KIT FAST.

Pollard, C., and I. Sag. 1987. *An Information-Based Theory of Syntax and Semantics, Vol.I.* Stanford: CSLI.

Shieber, S. 1986. *An Introduction to Unification-Based Approaches to Grammar.* Stanford University: Center for the Study of Language and Information.

van der Sandt, R., and B. Geurts. 1992. Presupposition, Anaphora and Lexical Content. unpublished ms.