Discriminant-Based MRS Banking

Stephan Oepen♣♠ and Jan Tore Lønning♣

♣Universitetet i Oslo, Boks 1102 Blindern; 0317 Oslo (Norway)
♠Center for the Study of Language and Information, Stanford, CA 94305 (USA)

{ oe@csli.stanford.edu | jtl@ifi.uio.no}

Abstract

We present an approach to discriminant-based MRS banking, i.e. the construction of an annotated corpus where each input item is paired with a logical-form semantics. Semantic annotations are produced by parsing with a broad-coverage precision grammar, followed by manual disambiguation. The selection of the preferred analysis for each item (and hence its semantic form) builds on a notion of semantic discriminants, essentially localized dependencies extracted from a full-fledged, underspecified semantic representation.

1. Background — Motivation

Minimal Recursion Semantics (MRS; Copestake, Flickinger, Sag, & Pollard, 1999) has gained popularity in computational semantics for its balance of formal and computational properties. Much like similar approaches (Reyle, 1993, Bos, 1995, inter alia), MRS facilitates underspecification of common scope ambiguities by means of separating dominance relations and 'regular' argument binding. Based on a designated type of 'handle' variables plus dominance constraints on how these can be realized over an otherwise flat multi-set of semantic relations, scopal relations are effectively factored out from the 'gist' of the semantics. MRS is used as the meaning representation language in the Norwegian LOGON project, a research initiative aiming for high-quality machine translation of tourism texts (Oepen et al., 2004).[1] LOGON pursues a relatively conventional approach to MT based on semantic transfer. A Norwegian sentence is analyzed into its MRS form. This semantic representation of the content of the Norwegian sentence is mapped into an MRS representations of the target language through semantic transfer, and subsequently realized using grammar-based generation for English. For analysis, an existing Norwegian parser (NorGram; Dyvik, 1999, Butt, Dyvik, King, Masuichi, & Rohrer, 2002) based on LFG and implemented in XLE was extended and equipped with an MRS projection (in the LFG co-description approach). For English generation, the LinGO English Resource Grammar (ERG; Flickinger, 2000) and DELPH-IN HPSG tools[2] were applied. All interface terms among LOGON components are MRS formulae. Despite current streams of fashion of using purely satistical methods for MT, the LOGON approach assumes that long-term success in MT will require integration of symbolic and stochastic approaches. In addition to the rule-based semantic transfer architecture, LOGON applies statistical methods for selecting and ranking between alternatives both in its parsing and generation phases.

Both for project-internal diagnostics and for training of domain-specific stochastic processes, the project requires a way of manually identifying intended analyses for a set of outputs obtained from batch parsing a development corpus. Besides, hand-picking the target reading(s) among parser outputs, of course, also helps identify (and thus suppress) legitimate but dis-preferred analyses. Such annotation immediately benefits the work on transfer (since transfer grammarians can focus on the analyses for each input that the NorGram developers find intended). Quite generally, it results in better quality MRSs (because of the regular in-depth scrutiny of each individual output). It further provides a foundation for work on training stochastic parse selection models, and we expect a medium-size MRS bank will suffice to train a domain-specific parse selection model for Norwegian. In a related spirit, Oepen, Flickinger, & Bond (2004) argue that maintaining a set of reference corpora in treebanked form throughout releases can be a valuable grammar engineering and regression testing facility.

2. Discriminant-Based Treebanking

The LinGO Redwoods treebanking environment (Oepen et al., 2002; Oepen, Flickinger, Toutanova, & Manning, 2004) is a combination of two devices, viz. (i) a tree comparison tool for HPSG analyses (similar in spirit to the SRI Cambridge TreeBanker; Carter, 1997) and (ii) the [incr tsdb()] profiling environment (essentially a specialized database recording fine-grained parsing results obtained from a HPSG system; Oepen & Carroll, 2000). The LinGO Redwoods treebank of around 25,000 annotated utterances[3] was constructed by batch processing domain corpora using the LinGO ERG, recording all results in [incr tsdb()], and subsequently having annotators select the preferred analysis for each input. Crucially, both the resulting preferences and all decisions made by annotators are recorded in the [incr tsdb()] database.

The tree comparison tool presents annotators, one sentence at a time, with the full set of analyses produced by the grammar together with a condensed view of where the ambiguity

[1] See 'http://www.emmtee.net/' for background information on the LOGON initiative, including a comprehensive bibliography and access to the open-source core of the system.

[2] See 'http://www.delph-in.net/' for the DELPH-IN open-source repository of 'deep' NLP components and multilingual resources.

[3] The LinGO Redwoods treebank of English is another component in the open-source DELPH-IN repository; see 'http://www.delph-in.net/redwoods/' for specifics on the material included and its availability.

$$
\left[\begin{array}{ll}
\texttt{TOP} & h_1 \\
\texttt{RELS} & \left\langle
\left[\begin{array}{ll} \multicolumn{2}{l}{\textit{prpstn\_m}} \\ \texttt{LBL} & h_1 \\ \texttt{MARG} & h_2 \end{array}\right]
\left[\begin{array}{ll} \multicolumn{2}{l}{\textit{\_anbefale\_v}} \\ \texttt{LBL} & h_3 \\ \texttt{ARG0} & e_4 \\ \texttt{ARG1} & x_5 \\ \texttt{ARG2} & x_6 \end{array}\right]
\left[\begin{array}{ll} \multicolumn{2}{l}{\textit{pron}} \\ \texttt{LBL} & h_7 \\ \texttt{ARG0} & x_5 \end{array}\right]
\left[\begin{array}{ll} \multicolumn{2}{l}{\textit{\_en\_q}} \\ \texttt{LBL} & h_8 \\ \texttt{ARG0} & x_6 \\ \texttt{RSTR} & h_9 \\ \texttt{BODY} & h_{10} \end{array}\right]
\left[\begin{array}{ll} \multicolumn{2}{l}{\textit{\_tur\_n}} \\ \texttt{LBL} & h_{11} \\ \texttt{ARG0} & x_6 \end{array}\right]
\left[\begin{array}{ll} \multicolumn{2}{l}{\textit{\_rundt\_p}} \\ \texttt{LBL} & h_{11} \\ \texttt{ARG0} & e_{12} \\ \texttt{ARG1} & x_6 \\ \texttt{ARG2} & x_{13} \end{array}\right]
\left[\begin{array}{ll} \multicolumn{2}{l}{\textit{def\_q}} \\ \texttt{LBL} & h_{14} \\ \texttt{ARG0} & x_{13} \\ \texttt{RSTR} & h_{15} \\ \texttt{BODY} & h_{16} \end{array}\right]
\left[\begin{array}{ll} \multicolumn{2}{l}{\textit{\_kilde\_n}} \\ \texttt{LBL} & h_{17} \\ \texttt{ARG0} & x_{13} \end{array}\right]
\cdots \right\rangle \\
\multicolumn{2}{l}{\texttt{HCONS}\ \langle h_2 =_q h_3,\ h_9 =_q h_{11},\ h_{15} =_q h_{17} \rangle}
\end{array}\right]
$$

Figure 1: Partial MRS for the utterance *Vi anbefaler en tur rundt vassdragets kilder* (literally: 'We recommend {a | one} hike around the waterway's sources'). The core of the semantics is captured in the flat RELS bag of elementary predications (EPs), each comprised of a semantic predicate and a set of role – value pairs. While variables of types e and x denote events and referential entities, respectively, the h-type variables and labels (LBL) associated with each EP serve to encode scopal relations. An additional bag of handle constraints (HCONS) is at the core of scope underspecification, but will largely be ignored for our present purposes. — The figure shows only part of the semantics for the example sentence, omitting the EPs introduced by the *vassdrag* ('waterway') entity and its two-place possessive relation to the *kilde* ('source') entity (x_{13}).

that gives rise to this set of analyses, originates. Put simply, the full set of analyses reflects the product of a series of more local choices—alternation between lexical entries or alternatives for modifier attachment, for example—of which some are independent of each other while others may mutually interact. The tool extracts elementary linguistic properties—called *discriminants*—that correspond to local ambiguity and uses the inference rules of Carter (1997) to determine the smallest possible set of discriminants that fully disambiguates the parse forest. When presented with individual local properties as they indicate choice points in assigning the linguistic analysis to the token sentence, annotators can quickly navigate through the parse forest and identify the correct or preferred analysis in the current context (or, in rare cases, end up rejecting all analyses proposed by the grammar). Using the discriminant-based approach to tree comparison, and given the elementary nature of each decision, annotators need little expert knowledge of the underlying grammar, but instead decide on a range of properties that distinguish competing analyses and are relatively easy to judge.

For each discriminant, annotators can choose whether they require the indicated property in the intended analysis (i.e. positively select a discriminant) or disallow it (i.e. negatively reject a discriminant). Each annotator decision reduces the set of active analyses—trimming down the parse forest—as for positive decisions only trees that have the indicated property remain available, whereas with negative decisions all trees with the rejected property can be excluded. As the set of active analyses is incrementally reduced, so is the set of discriminants. Discriminants from the original set that either have no remaining active parse or are compatible with all remaining parses can be suppressed from the annotator display, as deciding on these properties will not further disambiguate the parse forest.

While the general Redwoods approach makes no implicit commitment as to the exact nature of discriminants, it is important to maintain a fine balance between, on the one hand, sufficient information for effective and full disambiguation and, on the other hand, locality and simplicity of individual decisions.[4]

For a second-year Stanford undergraduate in linguistics, the Redwoods approach to parse selection through minimal discriminators turned out to be not at all hard to learn. It required less training in specifics of the grammatical analyses delivered by the LinGO grammar than could have been expected. After three to four weeks in hands-on training, the annotator was able to disambiguate at a rate of about 2000 sentences per week. Annotator throughput is enhanced by the ability of the treebanking environment to only partially disambiguate a sentence and flag it for later completion, say where annotators do not have sufficient knowledge readily available to fully disambiguate.

For each sentence, not only the resulting preference(s) (or, in rare cases, the conclusion that no correct analysis was available) but also all decisions made by annotators are recorded in the [incr tsdb()] database. Thus, annotator decisions are available as first class data for later semi-automated treebank updates, e.g. following a new release of the analysis grammar. In a nutshell, semi-automatic updating of the treebank for an enhanced version of the underlying grammar can be achieved by re-applying the recorded disambiguating decisions to a new version of the corpus obtained from re-running the parser on the original data set.

While there is parallel research (in a partner project to LOGON) on adapting the Redwoods approach for the LFG framework used for Norwegian analysis in LOGON (Rosén, Smedt, Dyvik, & Meurer, 2005), in the following we develop a generalized, relatively framework-independent notion of discriminant-based 'treebanking', viz. discrimination based on basic contrasts in the universe of logical-form semantics.

[4]In some contexts, it can be the case that a dynamic, 'staged' inventory of discriminants with increasing complexity benefits the annotation process. In particular for highly ambiguous items, it may be feasible to reduce the parse forest in an initial annotation phase by means of unlabeled 'bracketing' discriminants only (which, in turn, could be seeded from a reliable phrase boundary detector if such a tool was available) and only in a later annotation phase increase discriminant granularity to the degree required for full disambiguation. Another scenario we are exploring involves a successive reduction of the packed parse forest itself, i.e. the unfolding and disambiguation of packing nodes, as they correspond to local ambiguity.

3. Variable-Free Semantics

In order to adapt the basic Redwoods approach to cross-framework MRS banking, we propose a procedure for reducing an MRS into 'variable-free' elementary dependencies. The main motivation for variable elimination is our goal of comparing semantic properties across multiple analyses, e.g. the set of competing parses for a token input, since there is no (straightforward) way of making sure that related pieces of semantics across analyses actually use parallel variables. For example, the EPs associated to an NP constituent shared among two analyses might well internally end up using distinct (albeit abstractly equivalent) semantic variables.

A central notion of this reduction step—moving from a full-fledged underspecified logical form to a localized dependency graph—is the concept of a *distinguished variable* in each semantic relation (EP). For most types of relations, the distinguished variable corresponds to its main index (ARG0 in MRSs), e.g. an event variable for verbal relations, a referential index for nominals. Assuming further that, by and large, there is a unique relation for each semantic variable for which that variable serves as the main index (thus assuming, for example, that prepositions, adjectives, and adverbs all introduce event variables of their own, which can be motivated in predicative usages at least), a set of MRSs can be broken down into a set of basic contrasting properties, called semantic *discriminants*. Much like with syntactic discriminants (in the original LinGO Redwoods environment), looking at such basic semantic contrasts can make it (a lot) easier to work out where exactly analyses differ. Adapting the Redwoods paradigm to the MRS universe, discriminants for MRSs come in one of the following three forms:

(a) *relation*$_i$

(b) *relation*$_i$ *role*$_j$ *relation*$_k$

(c) *relation*$_i$ *property*$_j$ *value*$_j$

Here, each *relation* is the predicate name of an EP, and *roles* are the ARG0, ARG1, et al. role labels within EPs. Furthermore, (semantic) *properties* are attributes like GEND, NUM, TENSE, et al. inside of MRS variables, and *values* are appropriate (atomic) instantiations for these properties.

To extract such triples from an MRS, it is converted into a variable-free form, called an *elementary dependency graph*. Based on the distinguished variable notion sketched already, each variable of the full MRS is coupled with its 'representative' relation. In a few corner cases where the uniqueness constraint on the introduction of main indices is not maintained, there usually exist linguistically motivated disambiguation heuristics. We commonly opt for a nominal EP, for example, as the representative relation for a referential index, rather than for the associated quantifier EP.

Furthermore, roles that take scopal arguments (h-type variables) are given a special treatment. Handle constraints of the form $h_i =_q h_j$ ('equal modulo quantifier insertion') in an MRS express that either the two are equal or that h_i outscopes h_j, i.e. formally that the formula depicted by h_j is a subformula of the formula depicted by h_i. An MRS can be viewed as a set of scopal tree fragments associated with a set of constraints (plus some general logical-form wellformedness conditions) on how handles can be equated in order to form one or more fully connected trees. In the localized dependency graphs constraints of the form $h_i =_q h_j$ are treated as if they were actually equating h_i and h_j, so as to directly 'link up' EPs in moving from the underspecified MRS to an elementary dependency graph. For grammars that only use handle constraints of the $=_q$ type, like NorGram and the ERG in LOGON, equating their top and bottom variables is acceptable for the purpose of local dependency extraction as we will not be concerned with further scope specification.

```
{ _1:
  _1:prpstn_m[MARG e4:_anbefale_v]
  e4:_anbefale_v[ARG1 x5:pron, ARG2 x6:_tur_n]
  x5:pronoun_q[]
  e12:_rundt_p[ARG1 x6:_tur_n, ARG2 x13:_kilde_n]
  x6:_en_q[]
  _2:poss[ARG1 x21:_vassdrag_n, ARG2 x13:_kilde_n]
  x13:def_q[]
  x21:def_q[]
}
```

Figure 2: Elementary dependency view on the sample MRS from Figure 1. The nodes are comprised of MRS relations, of which most are contributed by lexical entries but also allowing for semantic contributions from grammatical constructions (e.g. the representation of illocutionary force by virtue of so-called messages; Ginzburg & Sag, 2000). Arcs of the dependency graph are labeled by MRS role labels (ARG1, MARG et al.).

The elementary dependency view on the MRS from Figure 1 is shown in Figure 2. In this form, each relation is prefixed with its distinguished variable, where it is legitimate for multiple relations to share one distinguished variable—a common configuration with nominal EPs and the quantifier binding their instance variable, for example. Conversely, where variables appear as arguments within relations, the elementary dependency representation will consistently show the one 'representative' relation

For overt lexical ambiguity (Norwegian *kort* can be the adjective 'short' or the noun 'card'), type (a) discriminants are often suitable, as picking either the adjectival or nominal predicate is an appropriate localization of the contrast in this case. Regarding our type (b) discriminants, the ambiguity in, say, the Norwegian *en tur* ('a hike' or 'one hike'), would give rise to the following discriminants:

_en_q ARG0 _tur_n

udef_q ARG0 _tur_n

card(1) ARG1 _tur_n

Here the semantic contrast is reflected in the alternation of quantifiers potentially binding the _tur_n entity (where _en_q is a plain indefinite, and udef_q is the grammaticized, covert definite used in conjunction with cardinal adjectives. Likewise, the card(1) intersective modifier on _tur_n is an exclusive property of the 'one hike' reading. In this example the use of type (a) discriminants—simple presense or absence of individual semantic predicates—would in principle suffice, but when choosing among quantifiers it will typically be a lot easier for annotators to judge a contrast when the quantifier is actually coupled with the representative predicate of the entity bound by the quantifier.

Finally, our third type of MRS discriminants, type (c), contrasts properties within the main index (ARG0, i.e. the distinguished variable) of a single relation. For a nominal relation that is ambiguous between a singular or plural interpretation (Norwegian *dyr*, 'animal' or 'animals'), we might see something like:

_dyr_n NUM sg
_dyr_n NUM pl

In terms of actual annotation practice, we find it convenient to present discriminants in a 'staged' process, where annotators can advance from less specific to more specific discriminants—types (a) to (c)—as they see fit.

Using discriminants over a set of MRSs corresponding to the competing analyses of a single input, annotation of the intended reading(s) can now be accomplished by virtue of binary (and more or less independent) decisions on individual discriminants, where each decision selects or rejects a sub-set of the available analyses. Toggling the singular _dyr_n in the example above to 'yes', say, will reduce the set of active trees only to those compatible with this property. Typically, a relatively small number of decisions among discriminants allows one to fully disambiguate (aka identify the preferred analysis from) even large sets of analyses. Finally, to cope with situations where the same semantic predicate is used more than once in an MRS (which is not uncommon for quantifiers, for example), we require that each instantiated semantic relation (EP) be linked to the 'surface' form(s) that gave rise to this piece of semantics in one fashion or another. In LOGON, the analysis system straightforwardly associates each constituent (lexical or phrasal) with a sequence of token identifiers (pointing back to the underlying basic building blocks for this constituent), such that EPs projected off such constituents will remain distinguishable even where they share the same semantic predicate.

4. MRS Banking — Current State of Play

We have implemented the elementary dependency reduction and extraction of semantic discriminants as part of building the LOGON MT system. Already, the resulting discrimant-based MRS comparison tool has been used actively in grammar and system development, specifically by the transfer team when presented with a set of candidate readings for a token input. More recently, we have coupled the MRS comparison tool with the [incr tsdb()] Redwoods environment and thus created the infrastructure for actual MRS banking.

Figure 3 presents the (HTML interface to the) Redwoods MRS banking environment. In the state shown here, our example sentence has already been partially disambiguated: the [4 : 6] display in the summary line indicates that four analyses remain active—i.e. compatible with discriminant decisions made so far—while six have already been rejected. There are two remaining sources of ambiguity in this example, viz. (a) the choice of analyzing *en* as an indefinite or (singleton) cardinal and (b) the contrast of attaching the *rundt* ('around') PP within the object NP or to the verbal projection. In terms of corresponding semantic discriminants, this ambiguity manifests itself in variation for the ARG1 value of the _rundt_p dependency triple, where in one reading the argument of the PP modifier is an entity (_tur_n; 'hike') and in the other it is an event (_anbefale_v; 'recommend').

In order to meet some of the LOGON objectives sketched in Section 1 above (in-depth scrutiny of semantic forms, identification of unwanted readings, and creating infrastructure for stochastic modelling), we are about to start MRS banking on part of the LOGON development corpus (some 5,000 sentences of running text on back-country activities in Norway). At this phase, the focus of our MRS banking activities will be on confirming the utility of the approach. It will be particularly interesting to compare MRS banking efficiency to 'traditional' Redwoods treebanking, both in terms of the initial learning curve for annotators, as well as in terms of annotator throughput and consistency.

5. Discussion — Outlook

We have presentend an adaption of the discriminant-based Redwoods approach of semi-automated treebank construction to the semantic realm. Based on a notion of reducing a logical-form semantics into a localized, 'variable-free' dependency graph, we have proposed three types of basic semantic discriminants. We conjecture that simple semantic contrasts will prove equally easy to judge by non-expert annotators and, furthermore, that the specific types of discriminants developed in Section 3 will strike a good balance of formal power and simplicity. In other words, we believe that these three basic types will always be sufficient to fully resolve all kinds of ambiguities presented in semantic forms delivered by grammars like NorGram or the ERG while the total set of discriminants for any given input will be small enough for annotators to navigate at ease.

Besides the obvious relations to earlier Redwoods treebanking research (and likewise its ongoing adaptation for Japanese; Bond et al., 2004), there is a close relationship to ongoing work in a collaboration between the LOGON and TrePil projects (Rosén et al., 2005). The latter strain of research is aiming to produce a Redwoods-like Norwegian treebank composed of complete LFG analyses, in turn using the NorGram implementation at its core and aiming to treebank data from the LOGON corpus. In contrast to the present proposal, however, TrePil makes use of LFG-specific discriminants for annotation, e.g. properties extracted from the LFG c- and f-structures. Thus, it primarily targets syntactic disambiguation. For both NorGram (Norwegian LFG) and the ERG (English HPSG), it is possible for multiple analyses (distinct c- or f-structures, say, or distinct HPSG derivations) to project equivalent MRSs. Hence, there exist syntactic ambiguities that are not reflected in the semantics: assume that syntactically the adverbial in a sentence like 'she will arrive on Monday' could either attach to the non-finite base VP or to the constituent built from combining the finite auxiliary with its verbal

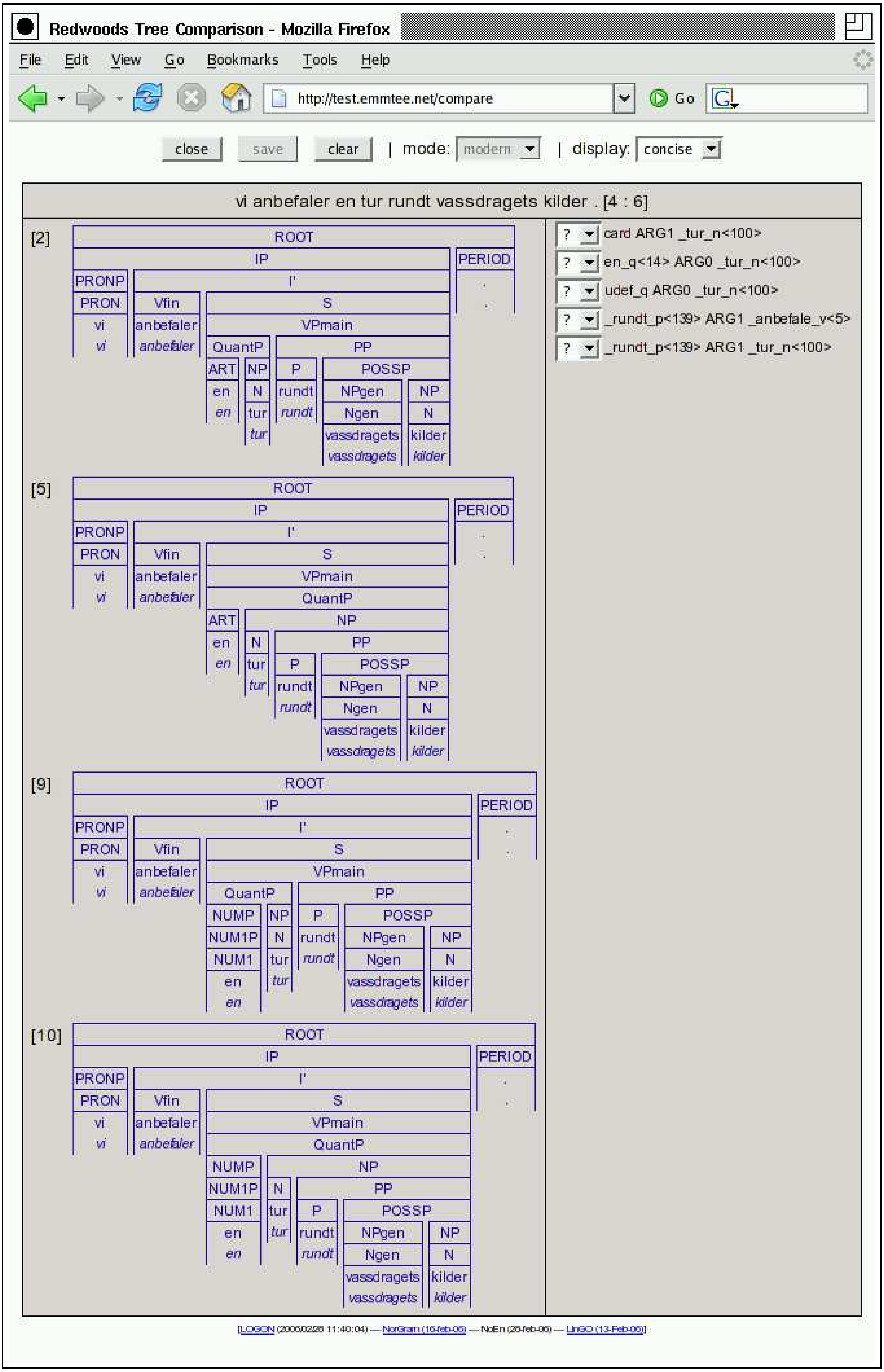

Figure 3: Screenshot of the MRS banking annotation tool after partial disambiguation.

complement. Semantically, there is only one event though (the ARG0 of an _arrive_v EP or something, whose tense and aspect properties may be further specified by the auxiliary). Thus, in terms of their semantics, the two distinct parse trees would collapse at the level of the MRS projection, and there would not be semantic discriminants to choose either analysis. From a purist (semanticist) point of view, such an ambiguity could be considered spurious, and at least in the LOGON approach to MT, there is no way for grammar-internal distinctions that are not reflected in the MRS interface terms to affect downstream processing. At the same time, there will often be good linguistic (or grammar-internal) reasons calling for such 'spurious' ambiguity, or the granularity of semantic description simply remains insufficient.

To the extent that one reason to treebank (or MRS bank) is to build annotated training material for stochastic processes—a parse selection model, say, to identify likely readings—there is no way of predicting which level of representation will be best-suited for a stochastic model to capture frequency distributions. The TrePil treebank of full LFG analyses (and likewise the original Redwoods HPSG treebanks) facilitates training of richer stochastic models, in the sense that they can condition on arbitrary c- or f-structure properties (and potentially MRS aspects too, of course). Conversely, an MRS bank built using the approach presented presently would limit the inventory of features accessible to a stochastic MRS selection model to just properties of MRSs. On the one hand, such a model would only see a reduced granularity of linguistic variation among competing outputs; on the other hand, it would also be confronted with less ambiguity (as 'spurious' MRS duplicates can be eliminated mechanically) and it would, at the same time, be trained on the actual, downstream application task—viz. ranking competing semantic hypotheses, irrespective of underlying syntactic structures. It is impossible to predict strong and weak points of either approach to parse selection, and we expect to investigate both paradigms in the remaining project duration.

Acknowledgments

The work presented here is embedded in a large collaborative effort, specifically the original Redwoods research, the LOGON consortium, and the larger DELPH-IN community. We are grateful to numerous colleagues and friends for inspiration and helpful discussions. Specifically, Dan Flickinger, Victoria Rosén, Francis Bond, and Rob Malouf have had a great deal of influence on this work.

References

Bond, F., Fujita, S., Hashimoto, C., Kasahara, K., Nariyama, S., Nichols, E., Ohtani, A., Tanaka, T., & Amano, S. (2004). The Hinoki Treebank. A treebank for text understanding. In *Proceedings of the 1st International Joint Conference on Natural Language Processing* (pp. 158 – 168). Hainan, China.

Bos, J. (1995). Predicate logic unplugged. In *Proceedings of the tenth Amsterdam colloquium* (pp. 133 – 142). Amsterdam, The Netherlands.

Butt, M., Dyvik, H., King, T. H., Masuichi, H., & Rohrer, C. (2002). The Parallel Grammar project. In *Proceedings of the COLING Workshop on Grammar Engineering and Evaluation* (pp. 1 – 7). Taipei, Taiwan.

Carter, D. (1997). The TreeBanker. A tool for supervised training of parsed corpora. In *Proceedings of the Workshop on Computational Environments for Grammar Development and Linguistic Engineering.* Madrid, Spain.

Copestake, A., Flickinger, D., Sag, I. A., & Pollard, C. (1999). *Minimal Recursion Semantics. An introduction.* In preparation, CSLI Stanford, Stanford, CA.

Dyvik, H. (1999). The universality of f-structure. Discovery or stipulation? The case of modals. In *Proceedings of the 4th International Lexical Functional Grammar Conference.* Manchester, UK.

Flickinger, D. (2000). On building a more efficient grammar by exploiting types. *Natural Language Engineering*, *6 (1)*, 15 – 28.

Ginzburg, J., & Sag, I. A. (2000). *Interrogative investigations. The form, meaning, and use of English interrogatives.* Stanford, CA: CSLI Publications.

Oepen, S., & Carroll, J. (2000). Performance profiling for parser engineering. *Natural Language Engineering*, *6 (1)*, 81 – 97.

Oepen, S., Dyvik, H., Lønning, J. T., Velldal, E., Beermann, D., Carroll, J., Flickinger, D., Hellan, L., Johannessen, J. B., Meurer, P., Nordgård, T., & Rosén, V. (2004). Som å kapp-ete med trollet? Towards MRS-based Norwegian – English Machine Translation. In *Proceedings of the 10th International Conference on Theoretical and Methodological Issues in Machine Translation.* Baltimore, MD.

Oepen, S., Flickinger, D., & Bond, F. (2004). Towards holistic grammar engineering and testing. Grafting treebank maintenance into the grammar revision cycle. In *Proceedings of the IJCNLP workshop Beyond Shallow Analysis.* Hainan,China.

Oepen, S., Flickinger, D., Toutanova, K., & Manning, C. D. (2004). LinGO Redwoods. A rich and dynamic treebank for HPSG. *Journal of Research on Language and Computation*, *2*(4), 575 – 596.

Oepen, S., Toutanova, K., Shieber, S., Manning, C., Flickinger, D., & Brants, T. (2002). The LinGO Redwoods treebank. Motivation and preliminary applications. In *Proceedings of the 19th International Conference on Computational Linguistics.* Taipei, Taiwan.

Reyle, U. (1993). Dealing with ambiguity by underspecification. Construction, representation, and deduction. *Journal of Semantics*, *10*, 123 – 179.

Rosén, V., Smedt, K. D., Dyvik, H., & Meurer, P. (2005). TrePil. Developing methods and tools for multilevel treebank construction. In *Proceedings of the 4th Workshop on Treebanks and Linguistic Theories* (pp. 161 – 172). Barcelona, Spain.