LinGO Redwoods

A Rich and Dynamic Treebank for HPSG

Stephan Oepen, Dan Flickinger, Kristina Toutanova, Christoper D. Manning

{oe|dan|kristina|manning}@csli.stanford.edu

Center for the Study of Language and Information
Stanford University
Ventura Hall, Stanford, CA 94305 (USA)

Abstract

The LinGO Redwoods initiative is a seed activity in the design and development of a new type of treebank. A treebank is a (typically hand-built) collection of natural language utterances and associated linguistic analyses; typical treebanks—as for example the widely recognized Penn Treebank (Marcus, Santorini, & Marcinkiewicz, 1993), the Prague Dependency Treebank (Hajic, 1998), or the German TiGer Corpus (Skut, Krenn, Brants, & Uszkoreit, 1997)—assign syntactic phrase structure or tectogrammatical dependency trees over sentences taken from a naturally-occuring source, often newspaper text. Applications of existing treebanks fall into two broad categories: (i) use of an annotated corpus in empirical linguistics as a source of structured language data and distributional patterns and (ii) use of the treebank for the acquisition (e.g. using stochastic or machine learning approaches) and evaluation of parsing systems.

While several medium- to large-scale treebanks exist for English (and some for other major languages), all pre-existing publicly available resources exhibit the following limitations: (i) the depth of linguistic information recorded in these treebanks is comparatively shallow, (ii) the design and format of linguistic representation in the treebank hard-wires a small, predefined range of ways in which information can be extracted from the treebank, and (iii) representations in existing treebanks are static and over the (often year- or decade-long) evolution of a large-scale treebank tend to fall behind theoretical advances in formal linguistics and grammatical representation.

LinGO Redwoods aims at the development of a novel treebanking methodology, (i) *rich* in nature and *dynamic* in both (ii) the ways linguistic data can be retrieved from the treebank in varying granularity and (iii) the constant evolution and regular updating of the treebank itself, synchronized to the development of ideas in syntactic theory. Starting in October 2001, the project is aiming to build the foundations for this new type of treebank, develop a basic set of tools required for treebank construction and maintenance, and construct an initial set of 10,000 annotated trees to be distributed together with the tools under an open-source license. Building a large-scale treebank, disseminating it, and positioning the corpus as a widely-accepted resource is a multi-year effort; the results of this seeding activity will serve as a proof of concept for the novel approach that is expected to enable the LinGO group at CSLI both to disseminate the approach to the wider academic and industrial audience and to secure appropriate funding for the realization and exploitation of a larger treebank. The purpose of publication at this early stage is three-fold: (i) to encourage feedback on the Redwoods approach from a broader academic audience, (ii) to facilitate exchange with related work at other sites, and (iii) to invite additional collaborators to contribute to the construction of the Redwoods treebank or start its exploitation as early-access versions become available.

This paper is an updated version of an earlier project report published by Oepen, Callahan, Flickinger, and Manning (2002); changes over that version include more recent numbers on the current Redwoods development status, inclusion of an example of discriminator-based disambiguation, and minor adaptations and corrections in various parts of the discussion.

1 Why Another (Type of) Treebank?

For the past decade or more, symbolic, linguistically oriented methods (like those pursued within the HPSG framework; see below) and statistical or machine learning approaches to NLP have typically been perceived as incompatible or even competing paradigms; the former, more traditional approaches are often referred to as 'deep' NLP, in contrast to the comparatively recent branch of language technology focussing on 'shallow' (text) processing methods. Shallow processing techniques have produced useful results in many classes of applications, but have not met the full range of needs for NLP, particularly where precise interpretation is important, or where the variety of linguistic expression is large relative to the amount of training data available. On the other hand, deep approaches to NLP have only recently achieved broad enough grammatical coverage *and* sufficient processing efficiency to allow the use of HPSG-type systems in certain types of real-world applications. Fully-automated, deep grammatical analysis of unrestricted text remains an unresolved challenge.

In particular, applications of analytical grammars for natural language parsing or generation require the use of sophisticated statistical techniques for resolving ambiguities. We observe general consensus on the necessity for bridging activities, combining symbolic and stochastic approaches to NLP; also, the transfer of HPSG resources into industry has amplified the need for general parse ranking, disambiguation, and robust recovery techniques which all require suitable stochastic models for HPSG processing. While we find promising research in stochastic parsing in an number of frameworks, there is a lack of appropriately rich and dynamic language corpora for HPSG. Likewise, stochastic parsing has so far been focussed on IE-type applications and lacks any depth of semantic interpretation. The Redwoods initiative is designed to fill in this gap.

Most probabilistic parsing research—including, for example, work by by Collins (1997), Charniak (1997), and Manning and Carpenter (2000)—is based on branching process models (Harris, 1963). An important recent advance in this area has been the application of log-linear models (Agresti, 1990) to modeling linguistic systems. These models can deal with the many interacting dependencies and the structural complexity found in constraint-based or unification-based theories of syntax (Johnson, Geman, Canon, Chi, & Riezler, 1999). The availability of even a medium-size treebank would allow us to begin exploring the use of these models for probabilistic disambiguation of HPSG grammars. At the same time, other researchers have started work on stochastic HPSG (or are about to), some pursuing unsupervised approaches, but in many cases using the same grammar or at least the same descriptive formalism and grammar engineering environment. The availability of a reasonably large, hand-disambiguated HPSG treebank is expected to greatly facilitate comparability of results and models obtained by various groups and, eventually, to help define a common evaluation metric.

2 Background

The LinGO Project at CSLI has been conducting research and development in Head-Driven Phrase Structure Grammar (HPSG; Pollard & Sag, 1994) since 1994. In close collaboration with international partners—primarily from Saarbrücken (Germany), Cambridge, Edinburgh, and Sussex (UK), and Tokyo (Japan)—the LinGO Project has developed a broad-coverage, precise HPSG implementation of English (the LinGO English Resource Grammar, ERG; Flickinger, 2000), a framework for semantic composition in large-scale computational grammars (Minimal Recursion Semantics, MRS; Copestake, Lascarides, & Flickinger, 2001), and an advanced grammar development environment (the LKB system; Copestake, 1992, 1999). Through contributions from collaborating partners, a pool of open-source HPSG resources has developed that now includes broad-coverage grammars for several languages, a common profiling and benchmarking environment (Oepen & Callmeier, 2000), and an industrial-strength C^{++} run-time engine for HPSG grammars (Callmeier, 2000). LinGO resources are in use world-wide for teaching, research, and application building. Because of the wide distribution and common acceptance, the HPSG framework and LinGO resources present an excellent anchor point for the Stanford treebanking initiative.

3 A Rich and Dynamic Treebank

The key innovative aspect of the Redwoods approach to treebanking is the anchoring of all linguistic data captured in the treebank to the HPSG framework and a generally-available broad-coverage grammar of English, viz. the LinGO English Resource Grammar, combined with tools for the extraction of various, user-defined representations and a software environment to continuously update the treebank as part of the on-going grammar maintenance and extension. Unlike existing treebanks, there will be no need to define a (new) form of grammatical representation specific to the treebank. Instead, the treebank will record complete syntacto-semantic analyses as defined by the LinGO ERG and provide tools to extract many different types of linguistic information at greatly varying granularity.

In particular, the project centrally draws on the [incr tsdb()] profiling environment (essentially a specialized database recording fine-grained parsing results obtained from diverse HPSG systems; Oepen & Carroll, 2000), constructing the treebank as an extension of the existing data model and tools. In turn building on a pre-existing tree comparison tool in the LKB (similar in kind to the SRI Cambridge TreeBanker; Carter, 1997), the treebanking environment presents annotators, one sentence at a time, with the full set of analyses produced by the grammar. Using the tree comparison tool, annotators can quickly navigate through the parse forest and identify the correct or preferred analysis in the current context (or, in rare cases, reject all analyses proposed by the grammar). The tree selection tools persents users, who need little expert knowledge of the underlying grammar, with a range of properties that distinguish competing analyses and that are relatively easy to judge. Each such property corresponds to the usage of a particular lexical item, semantic relation, or grammar rule applied to a specific substring to form a constituent; unlike the LFG packed f-structure representations discussed by King, Dipper, Frank, Kuhn, and Maxwell (2000), the set of basic discriminating properties reduces the information presented to annotators to the minimal amount of structure required to completely disambiguate a sentence. Figure 1 presents the Redwoods annotation environment.

All disambiguating decisions made by annotators are recorded in the [incr tsdb()] database and thus become available for (i) later dynamic extraction from the annotated profile or (ii) dynamic propagation into a more recent profile obtained from re-running an extended version of the grammar on the same corpus. Important innovative research aspects pertaining to this approach to treebanking are (i) enabling users of the treebank to extract information of the type they need and to transform the available representation into a form suited for their needs and (ii) updating the treebank for an enhanced version of the grammar underlying the recorded analyses in an automated fashion, viz. by re-applying the disambiguating decisions to an updated version of the corpus.

Depth of Representation and Transformation of Information Internally, the [incr tsdb()] database records analyses in three different formats, viz. (i) as a derivation tree composed of identifiers of lexical items and constructions used to construct the analysis, (ii) as a traditional phrase structure tree labeled with an inventory of some fifty atomic labels (of the type ‘S’, ‘NP’, ‘VP’ et al.), and (iii) as an underspecified MRS meaning representation. While (ii) will in many cases be similar to the representation found in the Penn Treebank, (iii) subsumses the functor – argument (or tectogrammatical) structure as

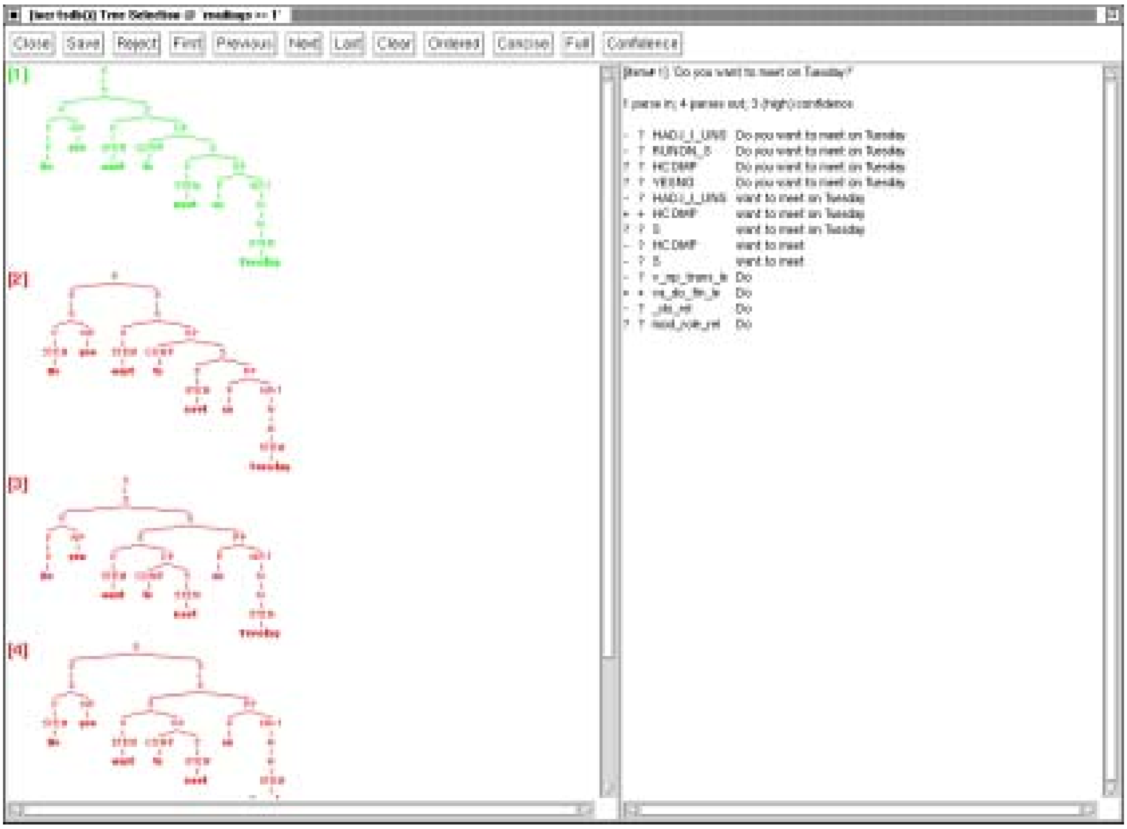

Figure 1: Screenshot of Redwoods treebanking environment: the window on the left presents the full set of analyses as labeled phrase structure trees (often too numerous to fit on a single page), the window on the right shows the minimal set of discriminating properties, based on either a particular lexical item, semantic relation, or construction applied to a specific substring to form a constituent.

it is advocated in the Prague Dependency Treebank or the German TiGer corpus. Most importantly, however, representation (i) provides all the information required to replay the full HPSG analysis (e.g. using the original HPSG grammar and one of the open-source HPSG processing environments, e.g. the LKB or PET, which already have been interfaced to [incr tsdb()]). Using the latter approach, users of the treebank are enabled to extract information in whatever representation they require, simply by reconstructing the full analysis and adapting the existing mappings (e.g. the inventory of node labels used for phrase structure trees) to their needs. Figures 2 and 3 depict the internal Redwoods encoding and two export representations—labeled constituent trees providing traditional phrase structure and elementary dependency graphs corresponding to functional structure, respectively—derived from existing conversion routines. Labeled phrase structure trees result from reconstructing a derivation (using the original grammar) and matching a user-defined set of underspecified feature structure 'templates' against the HPSG feature structure at each node in the tree. The elementary dependency graph, on the other hand, is an abstraction from the full MRS meaning representation associated to each full analysis; informally, elementary dependencies correspond to the type of tectogrammatical representations found in the Prague Dependency Treebank or the German TiGer corpus and, likewise, resemble the basic relations suggested for parser evaluation by Carroll, Briscoe, and Sanfilippo (1998). Given a rich body of MRS manipuation and conversion software, it is relatively straightforward to adapt the type and form of elementary dependencies to user needs.

Table 1: Redwoods development status as of June 2002: four sets of transcribed and hand-segmented VerbMobil dialogues have been annotated. The columns are, from left to right, the total number of sentences (excluding fragments) for which the LinGO grammar has at least one analysis (‘$\sharp$’), average length (‘$\|$’), lexical and structural ambiguity (‘$\odot$’ and ‘$\times$’, respectively), followed by the last four metrics broken down for the following subsets: sentences (i) for which the annotator rejected all analyses (no active trees), (ii) where annotation resulted in exactly one preferred analysis (one active tree), and (iii) those where full disambiguation was not accomplished through the first round of annotation (more than one active tree); around six per cent of massively ambiguous sentences have yet to be annotated; of the four data sets only VM_{32} has been double-checked by an expert grammarian and (almost) completely disambiguated to date; therefore it exhibits an interestingly higher degree of phrasal ambiguity in the ‘active = 1’ subset.

	total				**active = 0**				**active = 1**				**active > 1**			
	$\sharp$	$\|$	$\odot$	$\times$	$\sharp$	$\|$	$\odot$	$\times$	$\sharp$	$\|$	$\odot$	$\times$	$\sharp$	$\|$	$\odot$	$\times$
VM_6	2422	7·7	4·2	32·9	218	8·0	4·4	9·7	1910	7·0	4·0	7·5	80	10·0	4·8	23·8
VM_{13}	1984	8·5	4·0	37·9	175	8·5	4·1	9·9	1491	7·2	3·9	7·5	85	9·9	4·5	22·1
VM_{31}	1726	6·2	4·5	22·4	164	7·9	4·6	8·0	1360	6·6	4·5	5·9	61	10·1	4·2	14·5
VM_{32}	608	7·4	4·3	25·6	51	10·7	4·3	54·4	549	7·9	4·4	19·0	7	10·4	4·0	20·6

For evaluation purposes, the existing [incr tsdb()] facilities for comparing across competence and performance profiles can be deployed to gauge results of a (stochastic) parse disambiguation system, essentially using the preferences recorded in the treebank as a ‘gold standard’ target for comparison. While the concept of a meta-treebank of the type proposed here has been explored in earlier research (e.g. the AMALGAM project at Leeds University in the UK; Atwell, 1996), previous approaches to the dynamic mapping of treebank representations have built on a static, finite set of hand-constructed mappings.

Automating Treebank Construction Although a precise HPSG grammar like the LinGO ERG will typically assign a small number of analyses to a given sentence, choosing among a handful or sometimes a few dozens of readings is time-consuming and error-prone. The project will explore two approaches to automating the disambigutation task, viz. (i) seeding lexical selection from a part-of-speech (POS) tagger and (ii) automated inter-annotator comparison and assisted resolution of conflicts. Ranking lexical ambiguity on the basis of tagger-assigned POS probabilities requires research into generalizations over the rather fine-grained hierarchy of HPSG lexical types and identifying many-to-many correspondences in a standard POS tagset. Conversely, detecting mismatches (i.e. conflicts) between disambiguating decisions made for the same input sentence by two independent annotators will facilitate research into the linguistic nature of the discriminating properties used and existing logical relations (inclusion, implication, inconsistency et al.) among subsets of discriminators. To exemplify the nature of these properties, consider the sentence

(1) *Have her report on my desk by Friday!*

which is (correctly) assigned thirty two readings by the HPSG grammar; while human language users (and correspondingly human annotators) will typically not note most of the alternative analyses, one can contextualize the sentence to emphasize either one of the following ambiguities: the causative vs. possessive *have*, the determiner vs. personal pronoun *her*, the noun vs. verb *report*, the temporal vs. locative preposition *by*, and *Friday* as a day of the week vs. as a proper noun (e.g. the name of a bar). Using the tree comparison tool and our notion of elementary discriminators, annotators can reduce the set of analyses quickly (where full disambiguation requires minimally four decisions for this example); yet, a POS tagger will reliably assign high probability to the pairings ⟨*her*, determiner⟩ and ⟨*report*, noun⟩ which could be used to bias the presentation to annotators.

Treebank Maintenance and Evolution Perhaps the most challenging research aspect of the Redwoods initiative is about developing a methodology for automated updates of the treebank to reflect the continuous evolution of the underlying linguistic framework and of the LinGO grammar. Again building on the notion of elementary linguistic discriminators, it is expected to explore the semi-automatic propagation of recorded disambiguating decisions into newer versions of the parsed corpus. While it can be assumed that the basic phrase structure inventory and granularity of lexical distinctions have stabilized to a certain degree, it is not guaranteed that one set of discriminators will always fully disambiguate a more recent set of analyses for the same utterance (as the grammar may introduce additional distinctions), nor that re-playing a history of disambiguating decisions will necessarily identify the correct, preferred analysis for all sentences. Once more, a better understanding into the nature of discriminators and relations holding among them is expected to provide the foundations for an update procedure that, ultimately, should be fully automated or at least require minimal manual inspection.

Scope and Current State of Seeding Initiative The first 10,000 trees to be hand-annotated as part of the kick-off initiative are taken from a domain for which the English Resource Grammar is known to exhibit broad and accurate coverage, viz. transcribed face-to-face dialogues in an appointment scheduling and travel arrangement domain. Corpora of some 50,000 such utterances are readily available from the VerbMobil project (Wahlster, 2000) and have already been studied extensively among researchers world-wide in the field. For the follow-up phase of the project, it is expected to move into a second domain and text genre, presumably more formal, edited text taken from newspaper text or another widely available on-line source. As of June 2002, the seeding initiative is well underway. The integrated treebanking environment, combining [incr tsdb()] and the LKB tree selection tool, has been established and has been deployed in a first iteration of annotating a corpus of 10,000 VerbMobil utterances. For a second-year Stanford undergraduate in linguistics, the approach to parse selection through minimal discriminators turned out to be not at all hard to learn and required less training in specifics of the grammatical analyses delivered by the LinGO grammar than could have been expected.

Table 1 summarizes the current Redwoods development status; while annotation of a residual fraction of highly ambiguous sentences and inter-annotator cross-validation continue, the current development snapshot of the treebank can be made available upon request. We have just started work on stochastic parse selection models for the Redwoods treebank, so far obtaining a parse selection accuracy of around eighty per cent from a combination of existing methods applied to the Redwoods derivation trees and elementary dependency graphs (see Figures 2 and 3, respectively); details on Redwoods parse selection results are reported in a separate contribution to this volume.

4 Related Work

To our best knowledge, no prior research has been conducted exploring both the linguistic depth, flexibility in available information, and dynamic nature of treebanks as proposed presently. Earlier work on building corpora of hand-selected analyses relative to an existing broad-coverage grammar

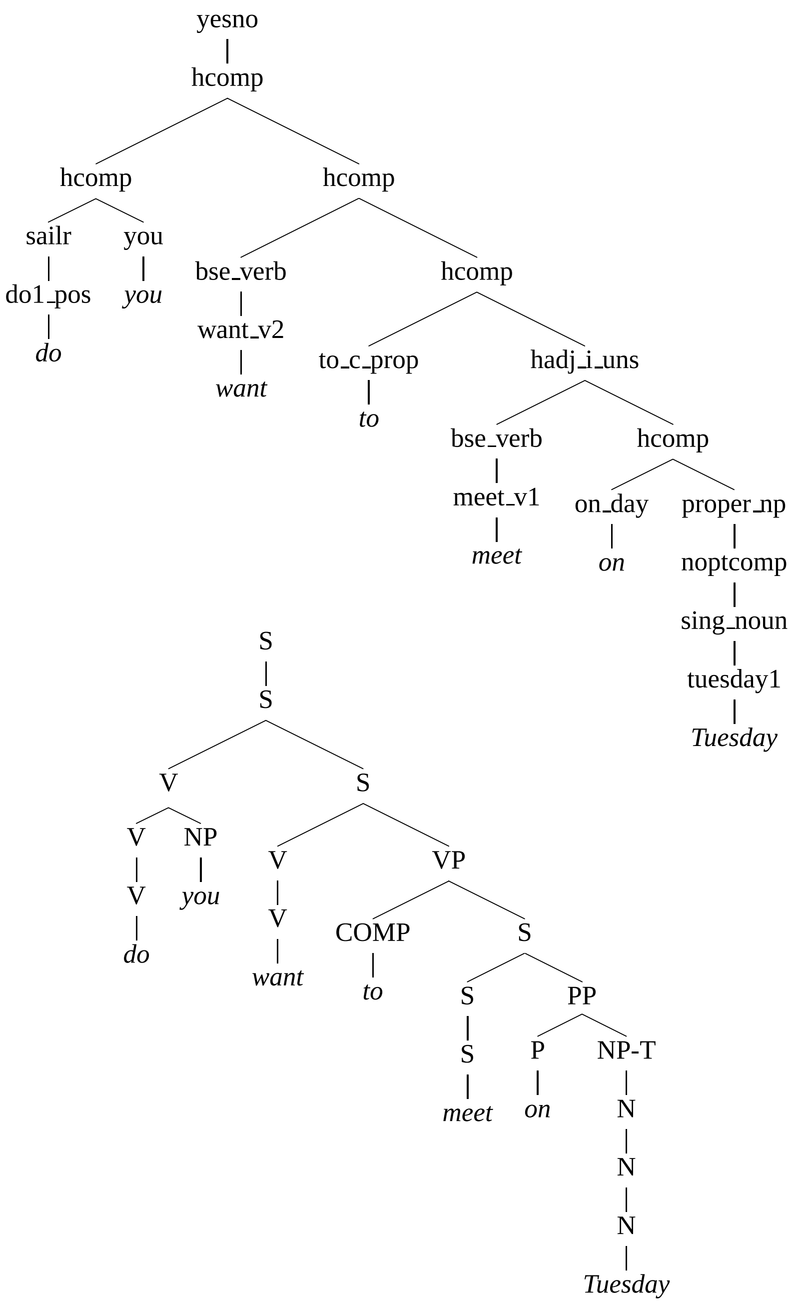

Figure 2: Native and derived Redwoods representations for the sentence *Do you want to meet on Tuesday?* (taken from the current development corpus): (a) derivation tree using unique rule and lexical item identifiers of the source grammar (top) and (b) phrase structure tree labelled with user-defined, parameterizable category abbreviations (bottom)

was carried out at Xerox PARC, SRI Cambridge, and Microsoft Research; as all these resources are tuned to proprietary grammars and analysis engines, the resulting treebanks are not publicly available, nor have research results reported been reproducible. Yet, especially in the light of the successful LinGO open-source repository, it seems vital that both the treebank and associated processing schemes and stochastic models be made available to the general (academic) public.

```
_4:{
  _4:int_rel[SOA e2:_want2_rel]
  e2:_want2_rel[ARG1 x4:pron_rel, ARG4 _2:hypo_rel]
  _1:def_rel[BV x4:pron_rel]
  _2:hypo_rel[SOA e18:_meet_v_rel]
  e18:_meet_v_rel[ARG1 x4:pron_rel]
  e19:_on_temp_rel[ARG e18:_meet_v_rel, ARG3 x21:dofw_rel]
  x21:dofw_rel[NAMED :tue]
  _3:def_np_rel[BV x21:dofw_rel]
}
```

Figure 3: Another derived Redwoods encoding: elementary dependency graph extracted from MRS meaning representation. The nodes are comprised by MRS relations, of which most are contributed by lexical entries but allowing for semantic contributions from non-lexical elements in the full HPSG derivation (e.g. the representation of illocutionary force by virtue of MRS messages). Arcs of the dependency graph are labeled by uninterpreted MRS role label (ARG1, SOA et al.) which could be assigned user-level interpretations as, for example, thematic roles relative to the lexicon and various MRS relation types.

An on-going initiative at Rijksuniversiteit Groningen (NL) is developing a treebank of dependency structures (Mullen, Malouf, & Noord, 2001), as they are derived from an HPSG-like grammar of Dutch (Bouma, Noord, & Malouf, 2001). While the general approach resembles the Redwoods initiative (specifically the discriminator-based method used in selecting trees from the set of analyses proposed by the grammar; the LKB tree selection tool was originally developed by Malouf, after all), there are three important differences. Firstly, the Groningen decision to compose the treebank from dependency structures commits the resulting resource to a single stratum of representation, tectogrammatical structure essentially, and thus eliminates some of the flexibility in extracting various types of linguistic structure that the Stanford initiative foresees. Secondly, and in a similar vein, recording dependency structures means that the (stochastic) disambiguation component has to consider two syntactically different analyses equivalent whenever they project identical dependency structures; hence, there is a mismatch of granularity between the disambiguated treebank structures and the primary structures (i.e. derivation trees) constructed by the grammar. Finally, the Groningen initiative is making the assumption that the dependency structures, once they are stored in the treebank, are correct and do not change over time (or as an effect of grammar evolution); from the available publications, at least, there is no evidence that the disambiguating decisions made by annotators are recorded in the treebank or that the project expects to dynamically update the treebank with future revisions of the underlying grammar.

Another closely related approach is the work reported by Dipper (2000), essentially the application of a broad-coverage LFG grammar for German to constructing tectogrammatical structures for the TiGer corpus. While many of the basic assumptions about the value of a systematic, broad-coverage grammar for the treebank construction are shared, the strategy followed by Dipper (2000) exhibits the same limitations as the Groningen initiative: the TiGer target representation, still, is mono-stratal and the approach to hand-disambiguation and subsequent transfer of result structures into the TiGer corpus looses the linkage to the original analyses and basic properties used in the disambiugation, hence the potential for dynamic adaptation of the data or automatic updates.

Acknowledgements

The Redwoods initiative is part of the LinGO Laboratory at CSLI and many people, both at Stanford and at partner sites, have contributed to its design and (given small amounts of resources) relative success so far. Ivan A. Sag, Tom Wasow, Emily M. Bender, Tim Baldwin, John Beavers, and Kathryn Campbell-Kibler all have participated in our regular tree conferences, helping annotators select parses and offering productive critiques on analyses provided by the LinGO grammar. Ann Copestake, John Carroll, Rob Malouf, and Stephan Oepen are the main developers of the LKB and [incr tsdb()] software packages from which the Redwoods treebanking environment has been built and, in various capacities, have influenced the Redwoods approach significantly. During a three-month visit to Stanford, Stuart Shieber has been among the driving forces for applications of the existing development version of the treebank, helping us develop and fine-tune suitable stochastic parse selection models. The Redwoods initiative has been partially funded by an internal opportunity grant from CSLI Stanford and by a donation from YY Technologies (Mountain View, CA).

References

Agresti, A. (1990). *Categorical data analysis.* John Wiley & Sons.

Atwell, E. (1996). Comparative evaluation of grammatical annotation models. In R. Sutcliffe, H.-D. Koch, & A. McElligott (Eds.), *Proceedings of the Workshop on Industrial Parsing of Software Manuals* (pp. 25 – 46). Amsterdam, The Netherlands: Rodopi.

Bouma, G., Noord, G. van, & Malouf, R. (2001). Alpino. Wide-coverage computational analysis of Dutch. In W. Daelemans, K. Sima-an, J. Veenstra, & J. Zavrel (Eds.), *Computational linguistics in the netherlands* (pp. 45 – 59). Amsterdam, The Netherlands: Rodopi.

Callmeier, U. (2000). PET — A platform for experimentation with efficient HPSG processing techniques. *Natural Language Engineering, 6 (1) (Special Issue on Efficient Processing with HPSG),* 99 – 108.

Carroll, J., Briscoe, E., & Sanfilippo, A. (1998). Parser evaluation: a survey and a new proposal. In *Proceedings of the 1st International Conference on Language Resources and Evaluation* (pp. 447 – 454). Granada, Spain.

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.

Charniak, E. (1997). Statistical parsing with a context-free grammar and word statistics. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence* (pp. 598 – 603). Providence, RI.

Collins, M. J. (1997). Three generative, lexicalised models for statistical parsing. In *Proceedings of the 35th Meeting of the Association for Computational Linguistics and the 7th Conference of the European Chapter of the ACL* (pp. 16 – 23). Madrid, Spain.

Copestake, A. (1992). The ACQUILEX LKB. Representation issues in semi-automatic acquisition of large lexicons. In *Proceedings of the 3rd ACL Conference on Applied Natural Language Processing* (pp. 88 – 96). Trento, Italy.

Copestake, A. (1999). *The (new) LKB system.* (CSLI, Stanford University: `http://www-csli.stanford.edu/~aac/doc5-2.pdf`)

Copestake, A., Lascarides, A., & Flickinger, D. (2001). An algebra for semantic construction in constraint-based grammars. In *Proceedings of the 39th Meeting of the Association for Computational Linguistics.* Toulouse, France.

Dipper, S. (2000). Grammar-based corpus annotation. In *Workshop on linguistically interpreted corpora linc-2000* (pp. 56 – 64). Luxembourg.

Flickinger, D. (2000). On building a more efficient grammar by exploiting types. *Natural Language Engineering, 6 (1) (Special Issue on Efficient Processing with HPSG),* 15 – 28.

Hajic, J. (1998). Building a syntactically annotated corpus. the Prague dependency treebank. In *Issues of valency and meaning* (pp. 106 – 132). Prague, Czech Republic: Karolinum.

Harris, T. E. (1963). *The theory of branching processes.* Berlin, Germany: Springer.

Johnson, M., Geman, S., Canon, S., Chi, Z., & Riezler, S. (1999). Estimators for stochastic ‘unification-based’ grammars. In *Proceedings of the 37th Meeting of the Association for Computational Linguistics* (pp. 535 – 541). College Park, MD.

King, T. H., Dipper, S., Frank, A., Kuhn, J., & Maxwell, J. (2000). Ambiguity management in grammar writing. In *Workshop on linguistic theory and grammar implementation* (pp. 5 – 19). Birmingham, UK.

Manning, C. D., & Carpenter, B. (2000). Probabilistic parsing using left corner language models. In H. Bunt & A. Nijholt (Eds.), *Advances in probabilistic and other parsing technologies* (pp. 105 – 124). Kluwer Academic Publishers.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English. The Penn Treebank. *Computational Linguistics*, *19*, 313 – 330.

Mullen, T., Malouf, R., & Noord, G. van. (2001). Statistical parsing of Dutch using Maximum Entropy models with feature merging. In *Proceedings of the Natural Language Processing Pacific Rim Symposium.* Tokyo, Japan.

Oepen, S., Callahan, E., Flickinger, D., & Manning, C. D. (2002). LinGO Redwoods. A rich and dynamic treebank for HPSG. In *LREC workshop on parser evaluation.* Las Palmas, Spain.

Oepen, S., & Callmeier, U. (2000). Measure for measure: Parser cross-fertilization. Towards increased component comparability and exchange. In *Proceedings of the 6th International Workshop on Parsing Technologies* (pp. 183 – 194). Trento, Italy.

Oepen, S., & Carroll, J. (2000). Performance profiling for parser engineering. *Natural Language Engineering, 6 (1) (Special Issue on Efficient Processing with HPSG)*, 81 – 97.

Pollard, C., & Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar.* Chicago, IL and Stanford, CA: The Univeristy of Chicago Press and CSLI Publications.

Skut, W., Krenn, B., Brants, T., & Uszkoreit, H. (1997). An annotation scheme for free word order languages. In *Proceedings of the 5th ACL Conference on Applied Natural Language Processing.* Washington, DC.

Wahlster, W. (Ed.). (2000). *Verbmobil. Foundations of speech-to-speech translation.* Berlin, Germany: Springer.