Theoretically Motivated Treebank Coverage

Victoria Rosén
University of Bergen and Unifob
victoria@uib.no

Koenraad de Smedt
University of Bergen and Unifob
desmedt@uib.no

Abstract

The question of grammar coverage in a treebank is addressed from the perspective of language description, not corpus description. We argue that a treebanking methodology based on parsing a corpus does not necessarily imply worse coverage than grammar induction based on a manually annotated corpus.

1 Introduction

The need for treebanks as an empirical basis for research on the grammar of a language is well established. While it is often stated that treebanks are useful for linguistic research as well as for language technology (Nivre et al., 2005), linguistic research with treebanks seems underrepresented in the literature (Nivre, 2005). Despite extensive research on the relation between treebanks and grammars, we think that theoretical issues in the relation between treebank annotation and grammar coverage have been underexposed.

Automatic grammar induction from manually annotated treebanks has been explored for over a decade (Krotov et al., 1994; Charniak, 1996; Burke et al., 2004; Cahill et al., 2004). Often, the quality of these grammars is discussed mainly as the percentage of coverage of the corpus by the grammar. There are, however, two obvious problems for inducing grammars from treebanks: one is the existence of annotation errors in the treebank (Dickinson and Meurers, 2005), and the other is the large number of rules that are obtained. For the *Wall Street Journal* part of the Penn Treebank "approximately 17,500 rules are required to analyze just under 50,000 sentences — or about one distinct rule for every three sentences" (Gaizauskas, 1995). According to Gaizauskas, one "possible explanation for the large number of rules is that in contexts where the annotators were unsure of the syntactic structure they created long rules that avoided issues of the internal structure of constituents." Such rules have little empirical content. Even if the number of rules can be dramatically decreased by means of compaction (Dickinson, 2006), the question remains to what extent an induced grammar captures linguistic generalizations about the language or rather describes a particular annotation in a particular corpus.

Grammar induction approaches evaluate the accuracy of the grammar by "measuring the degree to which parser output replicates the analyses assigned to sentences in a manually annotated test corpus" (Carroll et al., 2003), which is thereby treated as a gold standard. Even if some annotation errors can be detected and corrected (Dickinson and Meurers, 2005; Dickinson, 2006), any natural corpus will undoubtedly also include typos and other errors. From the viewpoint of theoretical linguistics, it is undesirable to end up with a grammar that overgenerates (Charniak, 1996) while not capturing the necessary linguistic generalizations.

The problematic status of automatically induced grammars shows that despite the apparent advantage of automatic acquisition, a theoretically motivated handwritten grammar should not be dismissed as a valuable starting point in treebank construction.

Joakim Nivre, Heiki-Jaan Kaalep, Kadri Muischnek and Mare Koit (Eds.)
NODALIDA 2007 Conference Proceedings, pp. 152–159

Parsebanking (a term we heard first from Ron Kaplan) is an approach to treebanking by parsing a corpus with a wide coverage grammar which has been advocated and practiced in recent years. In this paradigm, the question of coverage in the relation between the grammar and the treebank must be explicitly addressed. Treebanks as parsed corpora can be used to evaluate the performance of a parser or track its performance over different versions (Van der Beek et al., 2002; Oepen et al., 2004b).

In the context of the TREPIL project, we are developing a tool called LFG PARSEBANKER for incremental and interactive parsebanking (Rosén et al., 2005b; Rosén et al., 2005a; Rosén et al., 2006) which is compatible with any LFG grammar (Bresnan, 2001) implemented in XLE (Maxwell and Kaplan, 1993). We are therefore not committed to any particular grammar, although we are committed to the LFG formalism. In our approach, manual annotation is still necessary if the correct analysis has to be chosen among a number of possibilities. Much of our current research effort is in fact aimed at maximizing the efficiency of the disambiguation process.

Not every item that is passed to the parser as a 'sentence' can be expected to obtain an analysis (let alone a correct analysis). There may however be several reasons for this. In the following we will examine various coverage issues. It is useful to distinguish clearly between different types of noncoverage. In the first place we draw a distinction between input that one may not expect the grammar to cover and input that one would expect it to cover. The former is treated under sections 2 and 3. Input that ideally should be covered is discussed in sections 4 and 5. The examples given here will be mainly from the Norwegian grammar NorGram (Butt et al., 2002) but the points made are of a more general nature.

2 Non-syntactic input

Not everything in a corpus that typographically looks like a sentence (starts with an uppercase letter, ends with a period) is necessarily actually a sentence in the grammatical sense. Especially in corpus material such as newspaper text, many text chunks will be headlines, headers, lists, etc.

The PARC 700 Dependency Bank (King et al., 2003) contains dependency structures for 700 sentences randomly extracted from section 23 of the UPenn *Wall Street Journal* treebank (Marcus et al., 1994). The sentences were parsed with PARC's LFG grammar for English and the f-structures were converted into dependency structures. Many of the sentences received more than one parse, and example (1) is cited as an example of a sentence for which "the best parse was far from the desired parse" (King et al., 2003).

(1) *8 13/16% to 8 11/16% one month; 8 13/16% to 8 11/16% two months; 8 13/16% to 8 11/16% three months; 8 3/4% to 8 5/8% four months; 8 11/16% to 8 9/16% five months; 8 5/8% to 8 1/2% six months.*

It is not surprising that the grammar did not produce the desired parse for this item, since it doesn't have the structure of an English sentence. An alternative view would be to consider this a nonsentential item of some sort. It is in fact far from clear how it should be analyzed, if at all. It seems to be a list of some sort, and the readers of the *Wall Street Journal* may know how to process it. But grammatical analysis is not obviously appropriate for sundry lists. Although the *Wall Street Journal* may have a standard way of listing this information, it is clear that this is not a part of the grammar of the English language. We argue that items that have no true syntactic structure should not be annotated as such (but see the possibility of fragment parsing discussed below).

3 Performance phenomena

Performance errors are particularly salient in spoken language corpora, but they are also a feature of written language corpora which cannot be ignored. We will examine these in turn.

3.1 Spoken language corpora

The syntactic annotation of spoken language corpora is complicated by the fact that spoken language is characterized by numerous dysfluencies: false starts, repetitions, repairs, etc. There is no widespread consensus on how dysfluencies should be handled with respect to syntactic annotation.

Johannessen and Jørgensen (2006) discuss different strategies that various researchers have taken to the syntactic annotation of spoken language, ranging from ignoring any performance features in spoken language to including all of them, for instance speech repairs, in the annotation. The example in (2) from Sampson (2003), here in labeled bracket notation, illustrates the latter strategy.

(2) *and that [*$_{NomCl}$ *[*$_{NP}$ *any bonus [*$_{RelCl}$ *he] # anything [*$_{RelCl}$ *he gets* obj *[*$_{PP}$ *over that]]] is [*$_{NP}$ *a bonus]]*

Sampson discusses the difficulties involved in indicating “what is going on in a ‘speech repair’”. In this case, he says that “a speaker embarks on a relative clause modifying *any bonus* and then decides instead to use *anything* as the head of the phrase and to make *bonus* the predicate”. He uses the crosshatch symbol to indicate the point at which there is an interruption. He states that “we need rules for deciding how to fit that symbol, and the words before and after it, into a coherent structure[. . .]” and subsequently asks the question “Where in the tree do we attach the interruption symbol?” He doesn’t go into detail on how such questions are answered, but says: “[This analysis] is based on explicit decisions about these and related questions, and the variety of speech management phenomena found in real-life spontaneous speech is such that these guidelines have had to grow quite complex; but only by virtue of them can thousands of individual speech repairs be annotated in a predictable, consistent fashion.”

A consistent method of annotating speech repairs is certainly a good thing if one is interested in studying speech repairs. At the same time, it is not necessarily a good choice to mix this annotation with the annotation of syntactic structure. A phrase structure tree is normally used to indicate the constituent structure of a phrase or sentence, but the noun phrase indicated in this tree is not a recognizable pattern for a well-formed noun phrase. A better approach would be to let the dysfluency annotation and the syntactic annotation be done on separate levels.

In our approach to syntactic annotation of spoken language, only parts of utterances that have clear constituent structure will be annotated syntactically. For the example in (2), this would be the noun phrase *any bonus* and the clause *anything he gets over that is a bonus*. The rest of the utterance could of course be part of speech tagged, but a syntactic analysis in addition to tagging does not seem warranted when the syntactic structure is unclear.

Even though the problem of identifying grammatical and nongrammatical parts of spoken utterances is nontrivial, it is possible, with the help of the LFG PARSEBANKER, to achieve as good and rich an annotation for the grammatical parts of spoken language as for written language (Rosén, submitted).

3.2 Written language corpora

Performance errors are common in all text types. They are especially frequent in certain types of texts, such as newspapers, which are quickly written and typically proofread cursorily or not at all. For this reason it is somewhat surprising that so many corpora are heavily based on newspaper texts. This question seems to have been given little attention in the literature, although Becker et al. (2003) developed an error typology for a German newspaper corpus.

If one is interested in robustness, performance errors should not be corrected. Since our focus is on grammar, we are, however, not interested in trying to treat ungrammatical input as if it were grammatical. In contrast to some automatically induced grammars, we want to avoid building a grammar that covers ungrammatical input. There are then several possibilities for dealing with ungrammatical input. While one possibility consists of simply rejecting such items, another consists of correcting the errors but retaining information about what actually occurred in the corpus, and a third of assigning partial parsing, which is basically the same approach we adopt for speech performance phenomena, as discussed above.

In the LOGON Norwegian–English machine translation project (Oepen et al., 2004a), the problem of performance errors was dealt with through the “careful copy editor” principle, which may be formulated: “if there is a typographical error in the test corpus which a careful copy editor would have corrected, the test corpus item should be corrected”. A treebank built in the LOGON project contained sentences corrected according to this principle because these sentences were important for coverage in this specific application.

It is however not always easy to determine what kind of performance errors should be corrected if this approach is chosen. Clear typographical errors like misspellings are unproblematic. Grammatical performance errors are more difficult to decide on. What would a careful copy editor do with the sentence in (3)?

(3) *He wants to among other things to go fishing.*

Most editors would probably consider this an unintentional repetition of the infinitival marker, and would delete one of them. In the LFG PARSEBANKER, we can comment out one (or the other) and get a parse, retaining the information on what was actually in the corpus. Consider however the partial sentence in (4) from the LOGON development corpus.

(4) *[. . .] med innlagt solnedgang og*
[. . .] with included sunset and
flott utsyn mot Nesjøen,
beautiful view towards Nesjøen,
Nordskardsfjellet og ikke minst opp
Nordskardsfjellet and not least up
mot Sylmassivet.
towards Sylmassivet.
"[. . .] including a sunset and beautiful view out over Nesjøen, Nordskardsfjellet and of course up toward the Syl massif."

The problem in (4) is that there is a coordination of unlike constituents, a PP, an NP and another PP. Note that the professional idiomatic translation exhibits a parallel construction in English. A really careful copy editor would have corrected this, for instance by adding the preposition *mot* before *Nordskardsfjellet*, thus making a phrase of three coordinated PPs. This kind of construction, on the borderline of grammaticality, is a difficult problem, but not only for us. What do manually constructed treebanks do with this kind of sentence? Perhaps such language use is one reason why there are so many once-only rules in hand-annotated treebanks.

Some may want to consider the language in (4) more or less acceptable, whereas others do not want to include in the grammar. It is of course possible to maintain different versions of the grammar, one that includes it (for instance for robust parsing) and one that does not (for instance for generation). Our point is that providing an analysis or not for such an example must be a conscious choice. One should not contrive an analysis for a single example (thereby raising it from instance to type) if it is not motivated in terms of a linguistic generalization.

If one chooses to exclude borderline constructions from the grammar, one may still obtain *fragment* analyses for the parts that are covered. The simpler example in (5), with a similar pattern, may serve as an illustration. The Norwegian grammar produces the fragment analysis in Figure 1.

(5) *Petter går til butikken, parken og til*
Petter goes to store-the, park-the and to
slottet.
castle-the
"Petter goes to the store, the park and to the castle."

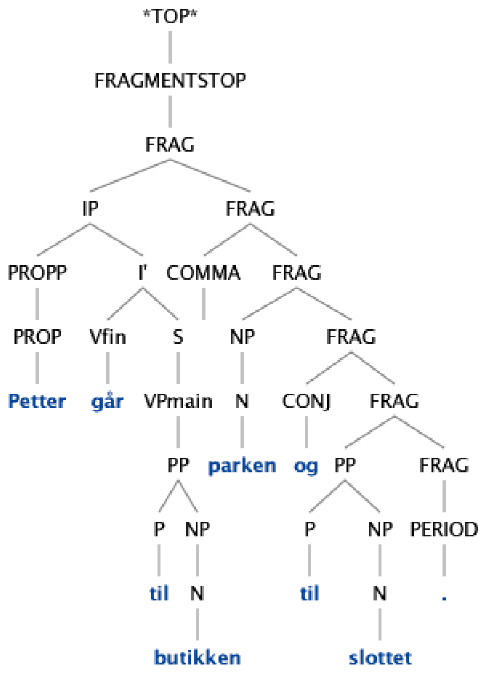

Figure 1: Fragment analysis for (5).

Summing up, we aim at a methodology for treebank construction based on linguistically motivated grammars and therefore ungrammatical input is rejected. In some applications, for instance for the purpose of compiling an error corpus, one might want to parse a corrected version of an item. In other applications that prioritize robustness over grammatical correctness, a more tolerant grammar may be used. If one doesn't want to alter the sentence and doesn't want to include the construction in the grammar, there is the option of producing a fragment analysis. Fragment analyses have proved useful for machine translation in the LOGON project.

4 Open-endedness of language

A handwritten grammar will always miss some constructions, not only by accidental omissions, but also because language is creative and open-ended. As an example of this class of coverage issues, we may consider resultatives, as in (6).

(6) a. *He wiped the table clean.*
b. *She hammered the metal flat.*

A resultative analysis presupposes that the verb subcategorizes for a predicate adjective complement, in addition to a subject and an object. This is not a problem for verbs like those in the examples in (6), where it is well known that they take such complements. Common resultative verbs may be expected to be present in the lexicon with the appropriate subcategorization frame. The problem with this construction is that it is quite creative. Example (7) from a Norwegian newspaper article about a man who threw garbage into his neighbor's yard illustrates this creativity well.

(7) *Det trekker til seg store mengder*
it draws to self large quantities
fugler som deretter skiter
birds that subsequently poop
eiendommen full.
property-the full
"This attracts large quantities of birds who subsequently poop the property full."

If the appropriate subcategorization frame for the verb *skite* is added to the lexicon, the sentence will get the intended analysis. This particular resultative use of this verb is impossible to predict, but with an interactive and incremental approach to parsebanking, it is not difficult to incorporate even such unexpected usages into the grammar and lexicon and thereby include them in the treebank.

Some constructions on the borderline of grammaticality were initially kept outside our Norwegian grammar, but were later included based on insights derived from the LOGON corpus. Example (8) consists of two independent clauses separated by a comma. In school teacher jargon, it would be called a run-on sentence.

(8) *Her går bjørka over 1200 meter,*
here goes birch-the over 1200 meters,
det er høyest i landet.
that is highest in country-the
"Here birches grow over 1200 meters, that is the highest in the country."

According to the normative rules of writing, both in Norwegian and English, independent clauses in the same sentence must be conjoined by a conjunction or a semicolon, not just a comma.

If this type of sentence is to be considered outside of the scope of the grammar, a high quality fragment analysis may still be produced. The c-structure representation of a fragment analysis for (8) is provided in Figure 2.

Our corpus work so far suggests however that authors quite often write sentences like this. Therefore the grammar has been modified to allow the comma to function in the same way as a conjunction in the grammar rules. In that case the sentence may be fully analyzed, as in Figure 3.

Parsebanking is a suitable methodology for tackling these coverage issues due to its incremental nature, so that the treebank is a gradually refined product of testing in the grammar construction process. The parsebanking approach benefits from advanced tools for supporting the communication between annotators and grammar developers (Rosén et al., 2006), for regression testing (Oepen et al., 2004b), etc.

5 Difficult syntactic problems

In the previous section we have demonstrated how some coverage problems, whether syntactic or lexical, come to light when the grammar is confronted with a corpus. Such coverage problems can be remedied by revising the grammar or lexicon and reparsing, for which we are developing efficient tools.

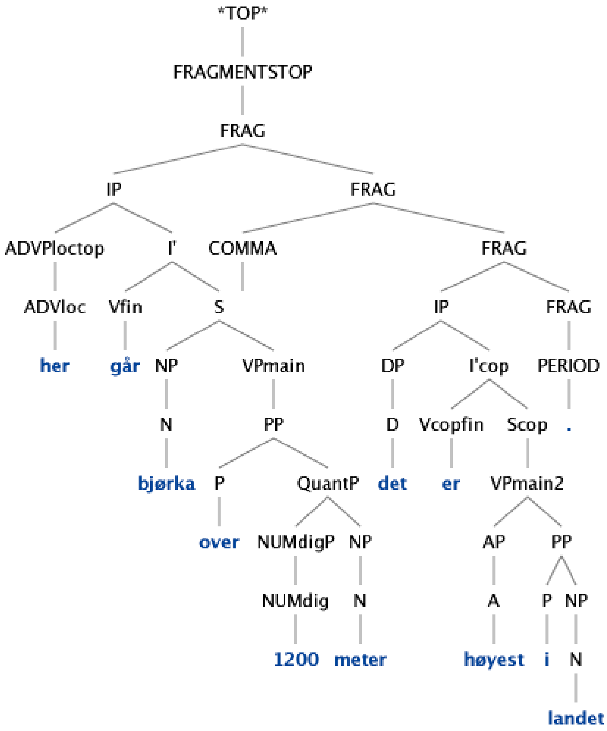

Figure 2: Fragment analysis for (8).

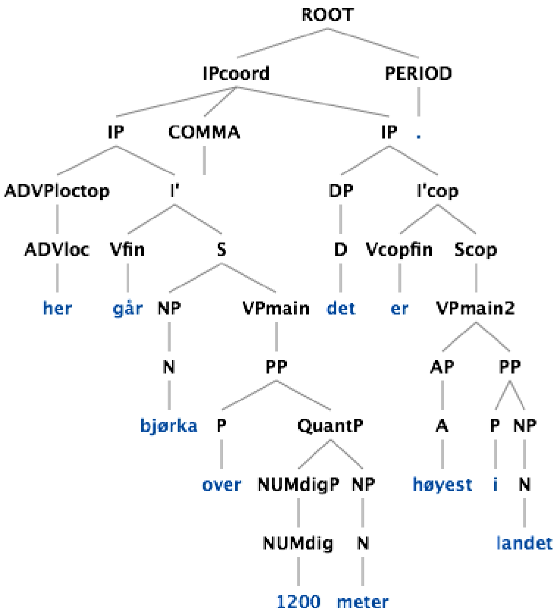

Figure 3: Full analysis for (8).

This is not to say that there are not real coverage problems for a handwritten grammar. There are genuinely difficult syntactic problems for all approaches to syntactic annotation.

A construction that has been the subject of much debate is the so-called “the more the merrier” construction (also called the “covariational conditional construction” (Goldberg and Casenhiser, 2006) and “correlative *the*-clauses”), examples of which are given in (9)–(11).

(9) *The more the merrier.*

(10) *The more chips you eat, the more you want.*

(11) *The bigger they come, the harder they fall.*

Among other issues, it has been debated how the definite article should be analyzed here, since it is not clear that it is a determiner in this construction. In the Penn Treebank II’s *Bracketing Guidelines* (Bies et al., 1995), we find an overview of bracket labels for annotating phrases. This explanation is provided for how the label *X* is to be used: “Unknown, uncertain or unbracketable. X is often used for bracketing typos and in bracketing *the*. . . *the*-constructions (see section 10 [Subordinate Clauses] and section 25 [Correlative *the*-Clauses]).” In section 10.7 we read: “There is no definitive policy for handling these cases. Most analyses involve the use of SBAR.” Then some sample analyses are provided for this construction. The analysis of the sentence in (12) is given in (13).

(12) *The more he muzzles his colleagues, the more leaks will pop up all around Washington.*

(13)
```
(S (SBAR-ADV (X the more)
             (S (NP-SBJ he)
                (VP muzzles
                    (NP his
                        colleagues))))
   ,
   (X the more)
   (NP-SBJ leaks)
   (VP will
       (VP pop
           (PRT up)
           (PP-LOC all around
                   (NP Washington)))))
```

Here the *X* is used to label a part of the tree for which the analysis is undecided. Although in a sense a tree has been drawn, it's a bit paradoxical to bracket a constituent with a label that means "unknown, uncertain or unbracketable". Furthermore, it is strange that the same annotation is used for both typos and challenging constructions. The question is whether this can be considered meaningful coverage. An alternative view would be that this is analogous to our fragment parsing, and that it only partially provides an analysis for the sentence.

If the treebank were to consist only of trees, it would be possible to construct a tool which would allow an annotator to edit a fragment analysis, erasing all the fragment nodes and creating new nodes and branches. However, manipulating the output of the grammar would mean that the treebank no longer was in sync with the grammar. Rather than devising ad hoc structures, we think it is fairer to admit that certain constructions are simply not covered yet.

Moreover, in our multilevel framework based on LFG, manual editing is not a desirable solution, since the c-structure is co-described by the grammar with the f-structure and the mrs-structure. Thus, manipulation of one structure would cause the different levels to no longer correspond, and since this correspondence is grammar dependent, it is not possible to make a tool that assures these correspondences for items that fall outside of the grammar.

6 Conclusion

In our work on treebanking as automatic parsing of a corpus, we have met criticism that not everything in the corpus is analyzed, or even can be analyzed. It might however be counterproductive to analyze everything just for the sake of the analysis. We have suggested that building a treebank is not just a matter of assigning some analysis to everything, but also of making grammaticality judgments. Analyses that contribute nothing to a linguistically motivated grammatical description are at best superfluous and increase grammar size. For items outside of the coverage of the grammar, a fragment analysis will often be useful. The parser then assigns structures to the largest chunks that it can analyze. This makes it easy for annotators and grammar developers to see what is covered and what is not. Legitimate analyses that are missing will be detected in our incremental and interactive approach linking treebanking to grammar development (Rosén et al., 2006), and their detection will lead to better coverage after revision of the grammar. Our aim is to develop treebanking methods that provide a correct and theoretically motivated account of a *language*, not of a corpus. With this aim, automatic parsing yields better quality by avoiding inconsistencies and other errors associated with a manual approach and eliminates complicated postprocessing steps for error detection and compaction.

References

Markus Becker, Andrew Bredenkamp, Berthold Crysmann, and Judith Klein. 2003. Annotation of error types for German Newsgroup Corpus. In Anne Abeillé, editor, *Treebanks: Building and Using Parsed Corpora*, chapter 6, pages 89–100. Kluwer Academic Publishers.

Ann Bies, Mark Ferguson, Karen Katz, Robert MacIntyre, Victoria Tredennick, Grace Kim, Mary Ann Marcinkiewicz, and Britta Schasberger. 1995. Bracketing guidelines for Treebank II style Penn Treebank Project. Technical report, University of Pennsylvania.

Joan Bresnan. 2001. *Lexical-Functional Syntax*. Blackwell, Malden, MA.

Michael Burke, Aoife Cahill, Ruth O' Donovan, Josef Van Genabith, and Andy Way. 2004. Treebank-based acquisition of wide-coverage, probabilistic LFG resources: Project overview, results and evaluation. In *The First International Joint Conference on Natural Language Processing (IJCNLP-04), Workshop "Beyond shallow analyses – Formalisms and statistical modeling for deep analyses", March 22-24, 2004 Sanya City, Hainan Island, China.*

Miriam Butt, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi, and Christian Rohrer. 2002. The Parallel Grammar project. In *Proceedings of COLING-2002 Workshop on Grammar Engineering and Evaluation, Taipei, Taiwan.*

Aoife Cahill, Michael Burke, Ruth O'Donovan, Josef Van Genabith, and Andy Way. 2004. Long-distance dependency resolution in automatically acquired wide-coverage PCFG-based LFG approximations. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics.*

John Carroll, Guido Minnen, and Ted Briscoe. 2003. Parser evaluation: Using a grammatical relation annotation scheme. In Anne Abeillé, editor, *Treebanks: Building and Using Parsed Corpora*, chapter 17, pages 299–316. Kluwer Academic Publishers.

Eugene Charniak. 1996. Tree-bank grammars. Technical Report CS-96-02, Dept. of Computer Science, Brown University, Providence, Rhode Island, January.

Markus Dickinson and W. Detmar Meurers. 2005. Prune diseased branches to get healthy trees! In *Proceedings of the Fourth Workshop on Treebanks and Linguistic Theories (TLT 2005)*.

Markus Dickinson. 2006. Rule equivalence for error detection. In Jan Hajič and Joakim Nivre, editors, *Proceedings of the Fifth Workshop on Treebanks and Linguistic Theories*.

Robert Gaizauskas. 1995. Investigations into the grammar underlying the Penn Treebank II. Technical Report Research Memorandum CS-95-25, University of Sheffield.

Adele E. Goldberg and Devin Casenhiser. 2006. English constructions. In Bas Aarts and April McMahon, editors, *The Handbook of English Linguistics*, Blackwell Handbooks in Linguistics, chapter 15. Blackwell.

Janne Bondi Johannessen and Fredrik Jørgensen. 2006. Annotating and parsing spoken language. In Peter Juel Henrichsen and Peter Rossen Skadhauge, editors, *Treebanking for Discourse and Speech: Proceedings of the NODALIDA 2005 Special Session on Treebanks for Spoken Language and Discourse*, pages 83–104, Copenhagen. Samfundslitteratur.

Tracy Holloway King, Richard Crouch, Stefan Riezler, Mary Dalrymple, and Ronald M. Kaplan. 2003. The PARC 700 dependency bank. In *Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora, held at the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL'03), Budapest*.

Alexander Krotov, Robert Gaizauskas, and Yorick Wilks. 1994. Acquiring a stochastic context-free grammar from the Penn Treebank. In *Proceedings of Third Conference on the Cognitive Science of Natural Language Processing*, pages 79–86.

Mitchell Marcus, Grace Kim, Mary Ann Marcinkiewicz, Robert MacIntyre, Ann Bies, Mark Ferguson, Karen Katz, and Britta Schasberger. 1994. The Penn Treebank: Annotating predicate argument structure. In *Proceedings of the ARPA Human Language Technology Workshop*.

John Maxwell and Ronald M. Kaplan. 1993. The interface between phrasal and functional constraints. *Computational Linguistics*, 19(4):571–589.

Joakim Nivre, Koenraad De Smedt, and Martin Volk. 2005. Treebanking in Northern Europe: A white paper. In Henrik Holmboe, editor, *Nordisk Sprogteknologi 2004. Årbog for Nordisk Sprogteknologisk Forskningsprogram 2000-2004*, pages 97–112. Museum Tusculanums Forlag, Copenhagen.

Joakim Nivre. 2005. Book review of Anne Abeillé, editor, Treebanks: Building and using parsed corpora, Kluwer AP, 2003. *Machine Translation*, 18:373–376.

Stephan Oepen, Helge Dyvik, Jan Tore Lønning, Erik Velldal, Dorothee Beermann, John Carroll, Dan Flickinger, Lars Hellan, Janne Bondi Johannessen, Paul Meurer, Torbjørn Nordgård, and Victoria Rosén. 2004a. 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, October.

Stephan Oepen, Dan Flickinger, Kristina Toutanova, and Christopher D. Manning. 2004b. LinGO Redwoods, a rich and dynamic treebank for HPSG. *Research on Language & Computation*, 2(4):575–596, December.

Victoria Rosén, Koenraad De Smedt, Helge Dyvik, and Paul Meurer. 2005a. TREPIL: Developing methods and tools for multilevel treebank construction. In Montserrat Civit, Sandra Kübler, and Ma. Antònia Martí, editors, *Proceedings of the Fourth Workshop on Treebanks and Linguistic Theories (TLT 2005)*, pages 161–172.

Victoria Rosén, Paul Meurer, and Koenraad De Smedt. 2005b. Constructing a parsed corpus with a large LFG grammar. In *Proceedings of LFG'05*, pages 371–387. CSLI Publications.

Victoria Rosén, Koenraad De Smedt, and Paul Meurer. 2006. Towards a toolkit linking treebanking to grammar development. In *Proceedings of the Fifth Workshop on Treebanks and Linguistic Theories*, pages 55–66.

Victoria Rosén. submitted. Mot en trebank for talespråk. In Janne Bondi Johannessen and Kristin Hagen, editors, *Språk i Oslo. Ny forskning omkring talespråk*. Novus forlag, Oslo.

Geoffrey Sampson. 2003. Thoughts on two decades of drawing trees. In Anne Abeillé, editor, *Treebanks: Building and Using Parsed Corpora*, chapter 2, pages 23–41. Kluwer Academic Publishers.

Leonoor Van der Beek, Gosse Bouma, Robert Malouf, and Gertjan Van Noord. 2002. The Alpino dependency treebank. In *Computational Linguistics in the Netherlands (CLIN) 2001*, Twente University.