Towards a Dependency-Oriented Evaluation for Partial Parsing

Sandra Kübler, Heike Telljohann

Seminar für Sprachwissenschaft
Wilhelmstr. 113
D-72074 Tübingen
Germany
{kuebler,hschulz}@sfs.uni-tuebingen.de

Abstract

Quantitative evaluation of parsers has traditionally centered around the PARSEVAL measures of *crossing brackets, (labeled) precision*, and *(labeled) recall*. However, it is well known that these measures do not give an accurate picture of the quality of the parser's output. Furthermore, we will show that they are especially unsuited for partial parsers. In recent years, research has concentrated on dependency-based evaluation measures. We will show in this paper that such a dependency-based evaluation scheme is particularly suitable for partial parsers. TüBa-D, the treebank used here for evaluation, contains all the necessary dependency information so that the conversion of trees into a dependency structure does not have to rely on heuristics. Therefore, the dependency representations are not only reliable, they are also linguistically motivated and can be used for linguistic purposes.

1. Introduction

Quantitative evaluation of parsers has traditionally centered around the PARSEVAL measures of *crossing brackets, (labeled) precision,* and *(labeled) recall* (Black et al., 1991). However, it is well known that these measures do not give an accurate picture of the quality of the parser's output (cf. Manning and Schütze (1999)), e.g. in cases of attachment errors. Additionally, many phenomena like negation or unary branches are ignored in the original measures in order to allow a comparison between parsers that use incompatible grammars. For this reason, research in recent years has concentrated on dependency-based evaluation measures (cf. e.g. Lin (1995), Lin (1998)). We will show in this paper that such a dependency-based evaluation scheme is particularly suitable for partial parsers since it does not lead to disproportionately high losses in precision and recall for partial parses. Furthermore, the dependency representations are not only reliable, they are also linguistically motivated and can be used for linguistic purposes since the treebank used here for evaluation contains all the necessary dependency information.

2. Deficiencies of Constituency-Based Precision and Recall

It is a well known fact that the PARSEVAL measures do not always give an accurate picture of the quality of a parser's output. Carroll and Brisoce (1996), for example, note that the *crossing brackets* measure is too lenient in case of errors involving the disambiguation of arguments and adjuncts, which in some cases are not recognized as errors. The failure to attach a constituent which should be embedded n levels deep leads to n crossing errors, while this constituent may not be very important to the overall structure. Manning and Schütze (1999) show that this behavior is mirrored in *precision* and *recall*: If a constituent is attached very high in a complex right branching structure, but the parser attached it at a lower point in the structure, both precision and recall will be greatly diminished. An example of such a parsing error for the sentence “ich nehme den Zug nach Frankfurt an der Oder” (I will take the train to Frankfort on the Oder) is shown in Figure 1[1]. There the prepositional phrase “an der Oder” is erroneously grouped as an adjunct of the verb instead of being attached as a postmodifier to the noun phrase “nach Frankfurt” (cf. the following section for a description of the annotation scheme). The correct tree is shown in Figure 2. When using the PARSEVAL measures, the output of the parser shown in Figure 1 results in $10/13 = 76.92\%$ recall[2] and $10/12 = 83.33\%$ precision, the only error being the wrong attachment of the last prepositional phrase.

The same behavior can be observed when the parser attaches a constituent very high in a complex right branching structure instead of very low, or if the constituent is not attached at all. The latter is often the case for chunk parsers (Abney, 1991; Abney, 1996) or partial parsers (cf. e.g. Aït-Mokhtar and Chanod (1997)). These parsers generally aim at annotating only partial, reliably discoverable tree structures, i.e. base phrases and clausal structures. Postmodifications are generally not attached since this decision cannot be taken reliably based on very limited local context. TüSBL (Kübler and Hinrichs, 2001a; Kübler and Hinrichs, 2001b), e.g., a similarity-based parser for German, annotates syntactic structures including function-argument structure in a two-level architecture: in the first phase, a deterministic chunk parser (Abney, 1996) is used to anal-

[1] All syntactic trees shown in this paper follow the data format for trees defined by the NEGRA project of the Sonderforschungsbereich 378 at the University of the Saarland, Saarbrücken. They were printed by the NEGRA graphical annotation tool *Annotate* (Brants and Skut, 1998; Plaehn, 1998).

[2] Contrary to the original PARSEVAL measures, we do count the root node as well since there exist different root nodes in the annotation scheme, and there are cases when a sentence in the treebank is annotated with more than one tree (e.g. interjective utterances).

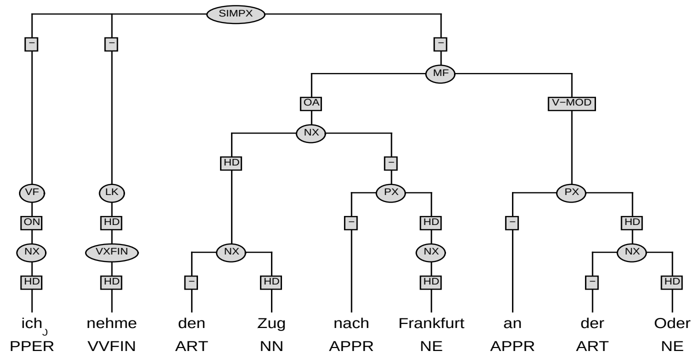

Figure 1: Wrong attachment of the prepositional phrase "an der Oder".

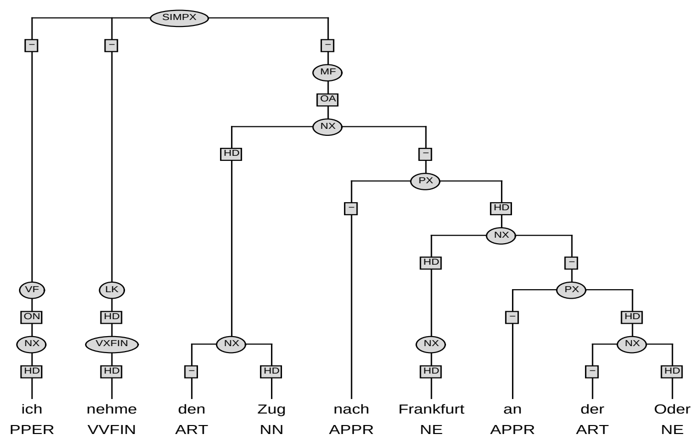

Figure 2: Correct attachment of the prepositional phrase "an der Oder".

yse major syntactic constituents such as non-recursive base phrases and simplex clauses. As a consequence, dependency relations between individual chunks, such as grammatical functions or modification relations, within a clause remain unspecified. In the second step, the attachment ambiguities are resolved, and the partial annotation of the first step are enriched by dependency information. A typical output of this phase is shown in Figure 3. The second phase of analysis is based on a similarity-based machine learning approach, which uses a similarity metric to retrieve the most similar sentence to the input sentence from the instance base and adapts the respective tree to the input sentence. (For a more detailed description of the algorithm cf. Kübler and Hinrichs (2001a) and Kübler and Hinrichs (2001b).) The parser is designed to prefer partial analyses over uncertain ones. In some cases, this strategy leads to unattached phrases, mostly at the end of sentences, which results in high losses in precision and recall. We therefore propose to use a dependency-based evaluation as described by Lin (1995) and Lin (1998), in which both the gold standard and the parser's output are transformed into dependencies and then compared on the basis of dependencies rather than on the basis of the constituent structure.

3. The TüBA-D Treebank

The dependency-based evaluation was based on the German corpus TüBa-D (Stegmann et al., 2000; Hinrichs et al., 2000a; Hinrichs et al., 2000b), which consists of approximately 38,000 syntactically annotated sentences. For this treebank, a theory-neutral and surface-oriented annotation scheme has been adopted that is inspired by the notion of topological fields – in the sense of Herling (1821), Erdmann (1886), Drach (1937), Reis (1980), and Höhle (1985) – and enriched by a level of predicate-argument structure, which guides the conversion into dependencies. The linguistic annotations pertain to the levels of morpho-syntax (part-of-speech tagging) (Schiller et al., 1995), syntactic phrase structure, and function-argument structure.

The tree structure contains different types of syntactic information in the following way: As the primary clustering principle the theory of topological fields (Höhle, 1985) is adopted, which captures the fundamental word order regularities of German sentence structure. In verb-second sentences, the finite verb constitutes the left sentence bracket (LK) and the verb complex the right sentence bracket (VC). This sentence bracket divides the sentence into the following topological order of fields: initial field (VF), LK, middle field (MF), VC, final field (NF). This structuring concept in addition favors bracketings that do not rely on crossing branches and traces to describe discontinuous dependencies.

Below this level of annotation, i.e. strictly within the bounds of topological fields, a phrase level of predicate-argument structure is established with its own descriptive

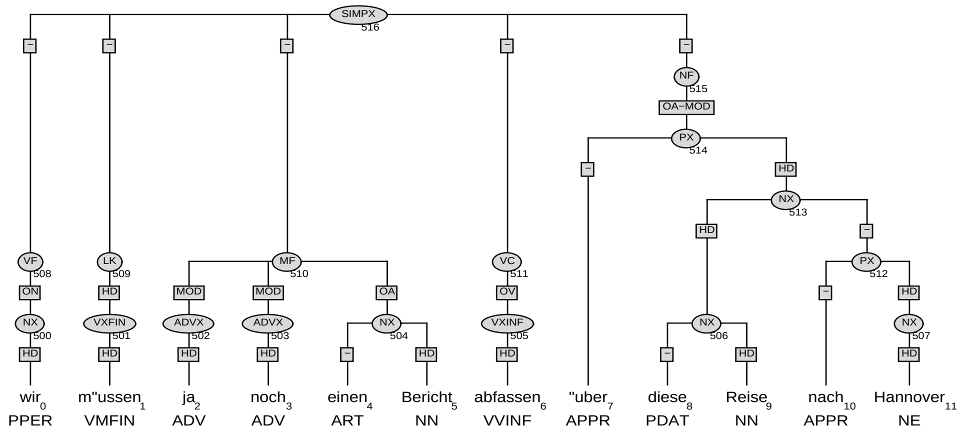

Figure 3: A tree annotated according to the TüBa-D treebank annotation scheme.

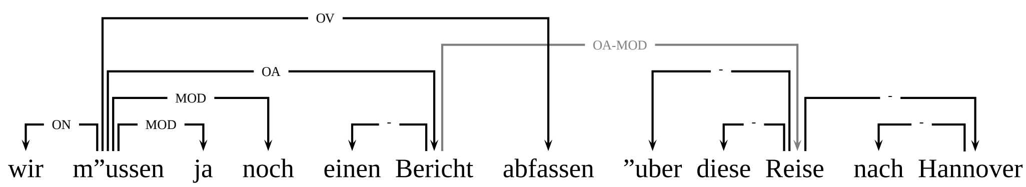

Figure 4: The dependency structure of the tree in Figure 3. The crossing dependency is shown in gray.

inventory based on a minimal set of assumptions concerning constituenthood, phrase attachment, and grammatical functions that have to be captured by any syntactic theory: nodes are labeled with syntactic categories on four different levels of annotation (sentence level, field level, phrase level, and lexical level), edges denote grammatical functions on the phrase level (i.e. immediately below the topological fields) and head/non-head distinctions within phrases. The integrated constituent analysis with its information about grammatical functions ensures that the resulting dependency structures are linguistically motivated and can also be used for linguistic purposes.

An example of such a tree for the sentence “wir müssen ja noch einen Bericht abfassen über diese Reise nach Hannover” (we still need to write a report on this journey to Hanover) is shown in Figure 3 (for more information about the annotation scheme cf. Stegmann et al. (2000)).

Two specific edge labels denote whether a constituent has the function of a head (HD), e.g. a phrase (NX, PX, ADJX, ADVX, VXFIN, VXINF), or a non-head (-), e.g. a determiner or a modifier attached to a phrase. On any annotation level, there is at most one head. The head of a sentence structure (e.g. SIMPX) is always the finite verb, which can be found in the left sentence bracket (LK). If there is no LK, the head is represented by the finite verb in the verb complex (VC). In coordinations, each conjunct depends on the head of the whole construction. Therefore, conjuncts are denoted with the non-head edge label.

The constituents below the topological fields are assigned grammatical functions. A subset of the edge label set consists of labels denoting the grammatical function of complements and modifiers, which depend on the head of the sentence. Another subset consists of labels determining long distance dependencies among these complements or modifiers as well as between conjuncts of split-up coordinations.

In Figure 3, e.g., the first constituent is marked as subject (ON), the finite verb is the head (HD), the two adverbs are modifiers (MOD), and the second noun phrase represents the direct object (OA). The constituent following the verb complex modifies the direct object (OA-MOD). Since the annotation scheme for the TüBa-D treebank facilitates a theory-neutral and surface-oriented representation of syntactic trees, this long distance relation is marked by the label OA-MOD (modifier of the accusative object) which refers to OA (accusative object) in the same tree; instead of using crossing branches and traces. This shows that long distance dependencies, which can even go beyond the border of topological fields, are encoded by special naming conventions for edge labels. Unambiguous edge labels, referring to exactly one non-adjacent constituent in the same tree, are used either for long distance modifications (*X*-MOD) like in the example above or for the rightmost conjunct of split-up coordinations (*X*K) (for an example cf. Figure 5). In both patterns, *X* is a variable for the grammatical function of the constituent to which it refers.

4. Converting TüBa-D into Dependencies

For TüBa-D, the conversion of the constituent structure into dependencies is in general determined by the head/non-head distinction in the tree. The dependency relations are labeled with the functional labels of the governed constituents. Using these strategies, the tree shown in Figure 3 is converted into the dependency structure in Figure 4. Here, the noun phrase “einen Bericht” is converted into one dependency relation, which denotes that the noun “Bericht”

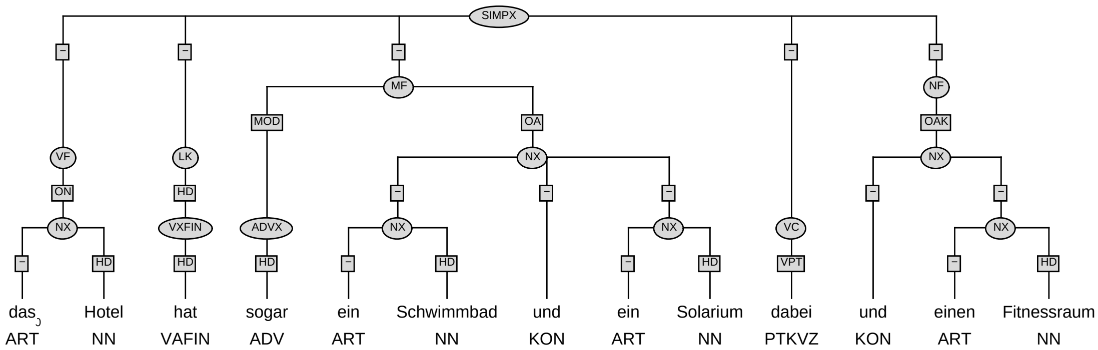

Figure 5: A complex coordination of noun phrases.

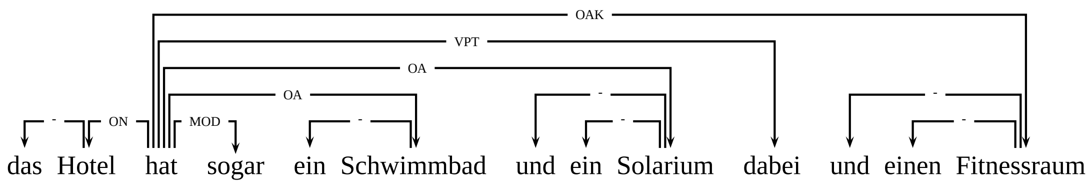

Figure 6: The dependency structure of the tree in Figure 5.

governs the article “den”.

It is evident that the dependency structure contains two different types of dependencies: head/non-head dependencies within phrases (-) and dependencies from the finite verb, i.e. from the head of the clause, to its complements and adjuncts, which are labeled by the grammatical functions of the governed constituents (ON, MOD, OA, OV). This is why e.g. the direct object “einen Bericht” is represented as a dependent of the modal verb “müssen” although it constitutes an argument of the embedded main verb “abfassen”. However, the dependency relations among the finite verb and the (possibly multiple) infinite verbs is explicitly annotated in the syntactic and therefore in the dependency structure. And since information about clausal boundaries is present in the trees, even in this surface-oriented structure, the predicate-argument structure can be recovered.

The long-distance dependency between the direct object and its modifying prepositional phrase was modeled in the syntactic tree by the function label “OA-MOD” instead of by the attachment of the prepositional phrase to the direct object because the latter would have resulted in a *crossing branch*. In the dependency structure, this restriction is suspended, and the dependency is explicitly marked and has now resulted in crossing dependencies. Note that this is the only type of phrase-internal dependency that is not labeled by the head/non-head distinction but by unambiguous labels which denote their specific reference.

Since head information is present on all levels for the majority of constituents, specific decisions for determining dependency have to be taken only in the few cases when dependency relations are not clearly defined in the tree structure, i.e. for the following syntactic phenomena:

1. Conjunctions within coordinations do not depend on the head of the whole construction. Therefore, they are attached to the conjunct on their right hand side. An example of such a coordination is shown in Figure 5, the corresponding dependency structure in Figure 6. Here, the third conjunct is positioned after the verb complex and thus is assigned the label “OAK”.
   Similar constructions with a preposition instead of a conjunction like “der achte bis neunte” (the eighth until the ninth) are treated in the same way. In order to stress the identical syntactic status of conjuncts, all conjuncts depend on the head governing the coordination. This analysis is in contrast to Lin (1998), who relies on the *Single Head Assumption* and proposes a dependency relation between the first and the second conjunct.

2. Sentence-initial coordinative particles such as “und” (and) or “oder” (or) in the KOORD-field depend on the head of the sentence.

3. The annotation of prepositional phrases in the syntactic trees is based on the principles of Dependency Grammar (Heringer, 1996); therefore, the noun phrase constitutes the head. For an example of the dependency structure of a prepositional phrase cf. the phrase “nach Hannover” in Figure 4. Circumpositions and postpositions are treated similarly.

4. The single elements of proper names, split cardinal numbers, the spelling of words, and complex conjunctions in the C-field, e.g. “so daß” (so that), are attached on the same level carrying a non-head edge label to indicate that there is no obvious dependency relation between them. Therefore, they are treated like conjuncts in coordinations.

5. A heuristic analysis has to be applied when long distance relations are underspecified – a MOD-MOD la-

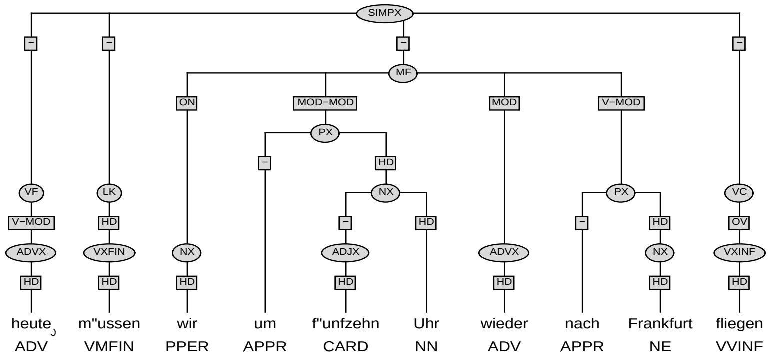

Figure 7: An ambiguous long-distance modifier: MOD-MOD.

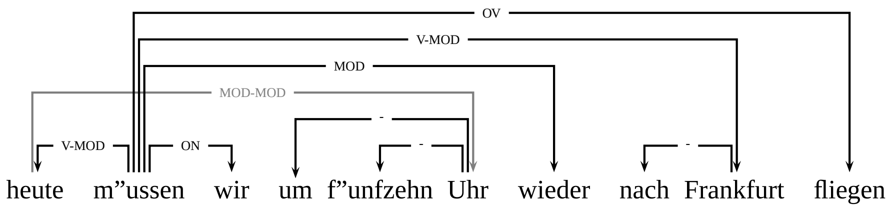

Figure 8: Resolved dependencies for ambiguous long-distance modifiers. The crossing dependency is shown in gray.

bel (modifier of a modifier), e.g., may refer to one of several modifiers in the sentence, such as for the sentence “heute müssen wir um fünfzehn Uhr wieder nach Frankfurt fliegen” (today we need to fly again to Frankfort) in Figure 7. Here, the long-distance modifier MOD-MOD might modify the V-MOD “heute” or the V-MOD “nach Franfurt”. A close inspection of such ambiguous sentences in TüBa-D revealed that in a majority of all cases, the MOD-MOD label refers to the first V-MOD in the clause, or the first MOD if there is no V-MOD present. Exceptions to this rule are MOD-MODs in resumptive constructions, which generally refer to the modifier in the VF. Ambiguous OA-MODs generally refer to the closest OA in the clause. By applying these heuristics, the ambiguities are resolved in the dependency structure, as shown in Figure 8 for the syntactic tree in Figure 7.

5. Dependency-Based Parser Evaluation

Lin (1998) proposed a procedure for converting syntactic trees from the gold standard and from the parser into dependency structures. From these structures, precision and recall are calculated.

Another similar evaluation procedure was suggested by Srinivas et al. (1996), they first convert hierarchical phrasal constituents into chunks, and then compute the dependencies between these chunks. This is a valid approach for the Penn treebank annotation style, which assumes a complete flat annotation of complex noun phrases such as noun compounds. Parsers based on manually developed rules tend to assign more internal structure to such noun phrases, which leads to decreased precision. Reducing such phrases to flat chunks alleviates this problem of comparing these different structures. The TüBa-D annotations, however, assign more complex, non-trivial structures to complex noun phrases. Using the method of Srinivas et al. (1996) would therefore lead to a significant loss in information. Additionally, the flattening of phrases into chunks might introduce errors in the data in such cases, in which the conversion into chunks is not obvious, such as for the noun phrase “wichtige Konferenzen und Besprechungen” in the sentence “da haben wir noch wichtige Konferenzen und Besprechungen” (we still have important conferences and business meetings) shown in Figure 9.

Basili et al. (1998) developed a similar approach for the Italian language. But instead of parsing a sentence completely and then reducing this parse to chunks and dependencies between chunks, Basili et al. apply a chunk parser combined with a module that calculates dependencies between these chunks. For this approach, the same restrictions hold as for the evaluation procedure of Srinivas et al. (1996).

The evaluation method presented here is based on Lin’s (Lin, 1998) approach. Following Lin’s procedure, we first convert both the gold standard tree and the parser’s output into dependency structures and compare these by applying (labeled) precision and (labeled) recall to these dependency structures.

TüSBL’s analyses depend heavily on the syntactically annotated sentences contained in the instance base. It is therefore difficult to give examples of errors for specific sentences or linguistic phenomena. It is, however, possible to characterize the typical behavior of the parser and give typical examples of errors.

Attachment errors. Attachment errors as described in Section 1. are not very common for TüSBL. Since TüSBL uses the complete sentence as context to retrieve the most similar tree, it either finds the correct spanning analysis or it does not attach all constituents. In the few cases

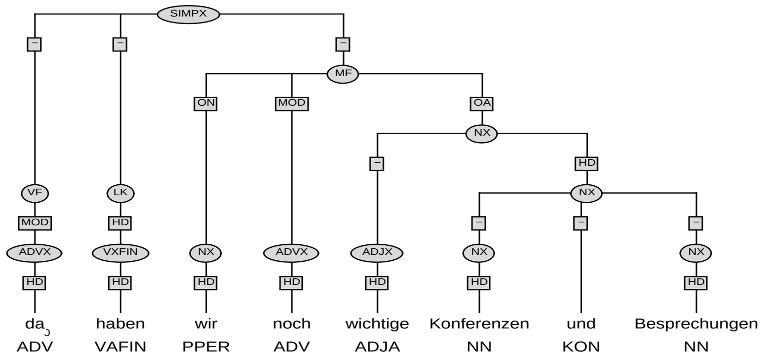

Figure 9: A complex noun phrase in the TüBa-D annotation scheme.

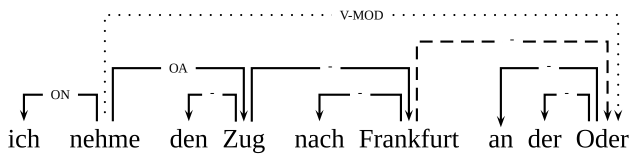

Figure 10: The dependency structure of the trees in Figure 1 and 2. The wrong attachment is shown as a dotted arc whereas the correct attachment is shown as a dashed arc.

where attachment errors are introduced by incorrect adaptations of the retrieved trees or in cases when a wrong tree is found as the most similar one, the parsers evaluation based on constituents suffers from the same problems as decribed in Section 2. above. The parser's output containing the wrong attachment in Figure 1 would result in $10/13 = 76.92\%$ recall and $10/12 = 83.33\%$ precision when using a constituent-based evaluation scheme. The dependency structure of the wrong and the correct attachment is shown in Figure 10. With the dependency-based evaluation, both precision and recall would be calculated as $7/8 = 87.50\%$.

Coordination. Coordination phenomena are in general very difficult to treat with deterministic partial parsers since this type of parsers needs to make the decision on the scope of a coordination early on when there is not enough information available. Two examples of coordination can be found in Figure 11. For both cases, TüSBL would typically retrieve these trees but not be able to attach the conjunction and the second conjunct, as shown in Figure 12 for the second example. For the first example, "am siebten und achten" (on the seventh and the eighth), this would lead to $2/4 = 50.00\%$ recall and $2/3 = 66.67\%$ precision. For the second example, "das wäre Mittwoch der dritte und Donnerstag der vierte August" (that would be Wednesday the third and Thursday the fourth of August), recall would be $9/12 = 75.00\%$ and precision $9/11 = 81.82\%$. If the evaluation is based on dependencies, TüSBL's analysis would deviate from the gold standard by the missing dependencies of the conjunction and the second conjunct. Therefore, recall would be $1/3 = 33.33\%$, for the first example, and $7/9 = 77.78\%$ for the second example. Precision would be $1/1 = 100\%$ for the first example and $7/7 = 100\%$ for the second example.

Another problematic coordination phenomenon constitute split-up coordinations such as in the sentence "das Hotel hat sogar ein Schwimmbad und ein Solarium dabei und einen Fitnessraum" (the hotel even has a swimming pool and a tanning booth – and a fitness room) in Figure 5. A typical error that might occur when parsing such sentences with TüSBL is that the split-up conjunct "und einen Fitnessraum" would not be attached. This would result in $12/14 = 85.71\%$ recall and $12/13 = 92.31\%$ precision. The evaluation based on the dependency structure shown in Figure 6 leads to $11/12 = 91.67\%$ recall and $11/11 = 100\%$ precision.

The comparison shows that dependency-based recall tends to suffer less than constituent-based recall since the unattached part of the coordination does not contribute to errors on higher levels, such as the MF and SIMPX in the second example, which are in principle correct. Dependency-based precision, on the other hand, does not depend on the level of embedding of the coordinations but only on the number of conjuncts that were correctly attached.

Unattached phrases. The failure to attach constituents at the end of an input sentence is the most common error type when evaluating partial parsers. It is generally part of the design decisions to prefer partial analyses which can be gained with a small amount of effort but which will be correct in a majority of cases to complete analyses which involve a high degree of manual labor and a higher error rate for attachment decisions. A typical analysis of TüSBL for the input sentence "wir müssen ja noch einen Bericht abfassen über diese Reise nach Hannover" would be similar to the tree in Figure 3; one possible error might be that the last PX ("nach Hannover") could not be attached to the NX ("diese Reise"). Thus, the NX node 513 would be missing, and the PX node 514 would then immediately dominate the NX node 506. Using the PARSEVAL measures, this

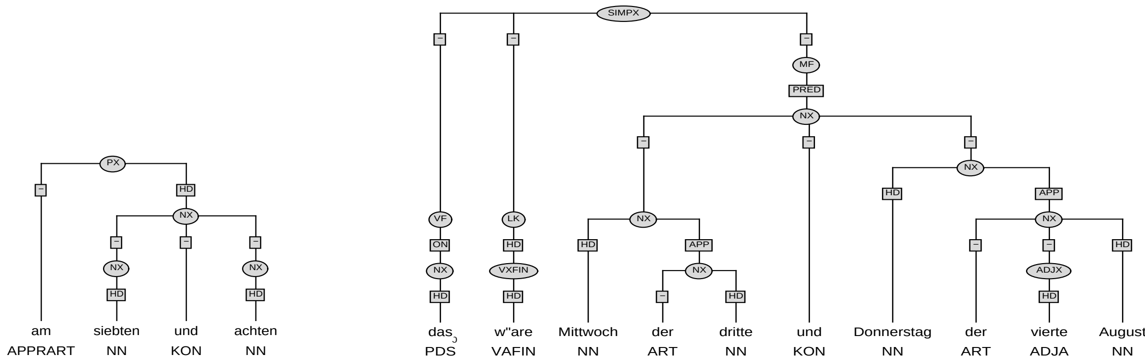

Figure 11: Two trees containing coordination.

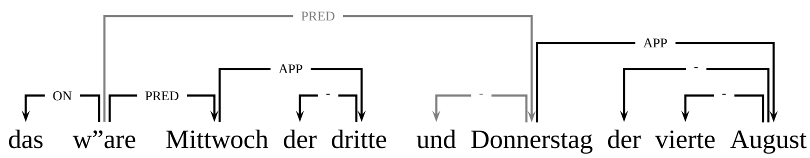

Figure 12: The dependency-based representation of the second example in Figure 11. TüSBL's analysis is shown in black, the missing dependencies in gray.

error would result in $13/17 = 76.47\%$ labeled recall and $13/16 = 81.25\%$ labeled precision. The evaluation based on the dependency structure would give $10/11 = 90.90\%$ labeled recall and $10/10 = 100\%$ labeled precision. Considering that only the attachment of the final PX is missing and that the analysis of the sentence is otherwise correct and complete, the latter figures give a better picture of the quality of the partial parse.

6. Conclusion

We have shown that the PARSEVAL measures do not allow a suitable evaluation of partial parsers. If the evaluation is based on constituency, missing information in the partial parses leads to precision and recall errors in several constituents, and the losses in both measures are disproportionately high. We therefore proposed a dependency-based evaluation. TüBa-D, the treebank used here, contains all the necessary dependency information so that the conversion of trees into a dependency structure does not have to rely on heuristics. Therefore, the dependency representations are not only reliable, they are also linguistically motivated and can be used for linguistic purposes. Using these structures for evaluation ensures that missing information will not decrease the evaluation measures disproportionately, which allows a more suitable evaluation of partial information.

7. Acknowledgments

The research reported here was supported by the German Research Council (DFG) as part of the Sonderforschungsbereich 441 “Linguistische Datenstrukturen” (Linguistic Data Structures). The authors are grateful to Steven Abney, who made the chunk parser available, and to Prof. Hans Uszkoreit and his colleagues at the Universität des Saarlandes, who kindly provided us with the graphical annotation tool *Annotate*. The authors are also grateful to the anonymous reviewers for their helpful comments.

8. References

Steven Abney. 1991. Parsing by chunks. In Robert Berwick, Steven Abney, and Caroll Tenney, editors, *Principle-Based Parsing*. Kluwer Academic Publishers.

Steven Abney. 1996. Partial parsing via finite-state cascades. In John Carroll, editor, *Workshop on Robust Parsing (ESSLLI ’96)*, pages 8 – 15, Prague, Czech Republic.

Salah Aït-Mokhtar and Jean-Pierre Chanod. 1997. Incremental finite-state parsing. In *Proceedings of ANLP’97*, pages 72 – 79, Washington, D.C.

Roberto Basili, Maria Teresa Pazienza, and Fabio Massimo Zanzotto. 1998. Evaluating a robust parser for Italian language. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 1998*, Granada, Spain.

E. Black, S. Abney, D. Flickinger, C. Gdaniec, R. Grishman, P. Harrison, D. Hindle, R. Ingria, F. Jelinek, J. Klavans, M. Liberman, M. Marcus, S. Roukos, B. Santorini, and T. Strzalkowski. 1991. A procedure for quantitatively comparing the syntactic coverage of English grammars. In *Proceedings of the DARPA Speech and Natural Language Workshop 1991*, Pacific Grove, CA.

Thorsten Brants and Wojciech Skut. 1998. Automation of treebank annotation. In *Proceedings of NeMLaP-3/CoNLL98*, pages 49 – 57, Sydney, Australia.

John Carroll and Ted Brisoce. 1996. Apportioning development effort in a probabilistic LR parsing system through evaluation. In *Proceedings of the ACL/SIGDAT Conference on Empirical Methods in Natural Language Processing*, pages 92 – 100, University of Pennsylvania, Philadelphia, PA.

Erich Drach. 1937. *Grundgedanken der Deutschen Satzlehre*. Frankfurt/M.

Oskar Erdmann. 1886. *Grundzüge der deutschen Syntax nach ihrer geschichtlichen Entwicklung dargestellt.* Stuttgart. Erste Abteilung.

Hans Jürgen Heringer. 1996. *Deutsche Syntax Dependentiell*. Stauffenburg-Verlag, Tübingen.

Simon Heinrich Adolf Herling. 1821. Über die Topik der deutschen Sprache. In *Abhandlungen des frankfurterischen Gelehrtenvereins für deutsche Sprache*, pages 296–362, 394. Frankfurt/M. Drittes Stück.

Erhard W. Hinrichs, Julia Bartels, Yasuhiro Kawata, Valia Kordoni, and Heike Telljohann. 2000a. The Tübingen treebanks for spoken German, English, and Japanese. In Wolfgang Wahlster, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer, Berlin.

Erhard W. Hinrichs, Julia Bartels, Yasuhiro Kawata, Valia Kordoni, and Heike Telljohann. 2000b. The Verbmobil treebanks. In *5. Konferenz zur Verarbeitung natürlicher Sprache (KONVENS 2000)*, pages 107 – 112, Ilmenau, Germany.

Tilman Höhle. 1985. Der Begriff "Mittelfeld, Anmerkungen über die Theorie der topologischen Felder. In *Akten des Siebten Internationalen Germanistenkongresses*, pages 329–340, Göttingen.

Sandra Kübler and Erhard W. Hinrichs. 2001a. From chunks to function-argument structure: A similarity-based approach. In *Proceedings of ACL-EACL 2001*, pages 338 – 345, Toulouse, France.

Sandra Kübler and Erhard W. Hinrichs. 2001b. TüSBL: A similarity-based chunk parser for robust syntactic processing. In *Proceedings of the First International Human Language Technology Conference, HLT-2001*, San Diego, CA, March.

Dekang Lin. 1995. A dependency-based method for evaluating broad-coverage parsers. In *Proceedings of IJCAI-95*, pages 1420 – 1425, Montreal, Canada.

Dekang Lin. 1998. A depedency-based method for evaluating broad-coverage parsers. *Natural Language Engineering*, 4(2):97 – 114.

Christopher D. Manning and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA.

Oliver Plaehn, 1998. *Annotate Bedienungsanleitung*. Universität des Saarlandes, Sonderforschungsbereich 378, Projekt C3, Saarbrücken, Germany, April.

Marga Reis. 1980. On justifying topological frames: 'Positional field' and the order of nonverbal constituents in German. *DRLAV: Revue de Linguistique*, 22/23:59 – 85.

Anne Schiller, Simone Teufel, and Christine Thielen. 1995. Guidelines für das Tagging deutscher Textkorpora mit STTS. Technical report, Universität Stuttgart and Universität Tübingen, September.

B. Srinivas, Christine Doran, Beth Ann Hockey, and Aravind Joshi. 1996. An approach to robust partial parsing and evaluation metrics. In John Carrol, editor, *Eight European Summer School in Language, Logic, and Information*, pages 70 – 82, Prague, Czech Republic.

Rosmary Stegmann, Heike Telljohann, and Erhard W. Hinrichs. 2000. Stylebook for the German Treebank in VERBMOBIL. Technical Report 239, Verbmobil.