The PaGe 2008 Shared Task on Parsing German[*]

Sandra Kübler
Department of Linguistics
Indiana University
Bloomington, IN, USA
`skuebler@indiana.edu`

Abstract

The ACL 2008 Workshop on Parsing German features a shared task on parsing German. The goal of the shared task was to find reasons for the radically different behavior of parsers on the different treebanks and between constituent and dependency representations. In this paper, we describe the task and the data sets. In addition, we provide an overview of the test results and a first analysis.

1 Introduction

German is one of the very few languages for which more than one syntactically annotated resource exists. Other languages for which this is the case include English (with the Penn treebank (Marcus et al., 1993), the Susanne Corpus (Sampson, 1993), and the British section of the ICE Corpus (Wallis and Nelson, 2006)) and Italian (with ISST (Montegmagni et al., 2000) and TUT (Bosco et al., 2000)). The three German treebanks are Negra (Skut et al., 1998), TIGER (Brants et al., 2002), and TüBa-D/Z (Hinrichs et al., 2004). We will concentrate on TIGER and TüBa-D/Z here; Negra is annotated with an annotation scheme very similar to TIGER but is smaller. In contrast to other languages, these two treebanks are similar on many levels: Both treebanks are based on newspaper text, both use the STTS part of speech (POS) tagset (Thielen and Schiller, 1994), and both use an annotation

scheme based on constituent structure augmented with grammatical functions. However, they differ in the choices made in the annotation schemes, which makes them ideally suited for an investigation of how these decisions influence parsing accuracy in different parsers.

On a different level, German is an interesting language for parsing because of the syntactic phenomena in which the language differs from English, the undoubtedly most studied language in parsing: German is often listed as a non-configurational language. However, while the word order is freer than in English, the language exhibits a less flexible word order than more typical non-configurational languages. A short overview of German word order phenomena is given in section 2.

The structure of this paper is as follows: Section 2 discusses three characteristics of German word order, section 3 provides a definition of the shared task, and section 4 gives a short overview of the treebanks and their annotation schemes that were used in the shared task. In section 5, we give an overview of the participating systems and their results.

2 German Word Order

In German, the order of non-verbal phrases is relatively free, but the placement of the verbal elements is determined by the clause type. Thus, we will first describe the placement of the finite verb, then we will explain phrasal ordering, and finally we will look at discontinuous constituents.

---

[*] I am very grateful to Gerald Penn, who suggested this workshop and the shared task, took over the biggest part of the workshop organization and helped with the shared task.

2.1 Verb Placement

In German, the clause type determines the placement of finite verbs: In non-embedded declarative clauses, as in (1a), the finite verb is in second position (V2). In yes/no questions, as in (1b), the finite verb is the clause-initial constituent (V1), and in embedded clauses, as in (1c), it appears clause finally (Vn).

(1) a. Der Mann hat das Auto gekauft.
The man   has the car   bought

'The man has bought the car.'

   b. Hat der Mann das Auto gekauft?
Has the man   the car   bought

'Has the man bought the car?'

   a. dass der Mann das Auto gekauft hat.
that the man   the car   bought has

'. . . that the man has bought the car.'

All non-finite verbs appear at the right periphery of the clause (cf. 2), independently of the clause type.

(2) Der Mann sollte   das Auto gekauft haben.
The man   should the car   bought have

'The man should have bought the car.'

2.2 Flexible Phrase Ordering

Apart from the fixed placement of the verbs, the order of the non-verbal elements is flexible. In (3), any of the four complements and adjuncts of the main verb *(ge)geben* can be in sentence-initial position, depending on the information structure of the sentence.

(3) a. Das Kind hat dem Mann gestern   den
The child has the   man   yesterday the
Ball gegeben.
ball given

'The child has given the ball to the man yesterday.'

   b. Dem Mann hat das Kind gestern den Ball
gegeben.

   c. Gestern hat das Kind dem Mann den Ball
gegeben.

   d. Den Ball hat das Kind gestern dem Mann
gegeben.

In addition, the ordering of the elements that occur between the finite and the non-finite verb forms is also free so that there are six possible linearizations for each of the examples in (3a-d).

One exception to the free ordering of non-verbal elements is the ordering of pronouns. If the pronouns appear to the right of the finite verb in V1 and V2 clauses, they are adjacent to the finite verb in fixed order.

(4) Gestern   hat sie sie   ihm gegeben.
Yesterday has she her/them him given.

'Yesterday, she gave her/them to him.'

In (4), three pronouns are present. Although the pronoun *sie* is ambiguous between nominative/accusative singular and nominative/accusative plural, the given example is unambiguous with respect to case since the nominative precedes the accusative, which in turn precedes the dative.

Due to the flexible phrase ordering, the grammatical functions of constituents in German, unlike in English, cannot be deduced from the constituents' location in the constituent tree. As a consequence, parsing approaches to German need to be based on treebank data which contain a combination of constituent structure and grammatical functions – for parsing and evaluation. For English, in contrast, grammatical functions are often used internally in parsers but suppressed in the final parser output.

2.3 Discontinuous Constituents

Another characteristic of German word order is the frequency of discontinuous constituents. The sentence in (5) shows an extraposed relative clause that is separated from its head noun *das Buch* by the non-finite verb *gelesen*.

(5) Der Mann hat das Buch gelesen, das   ich
The man   has the book read,   which I
ihm empfohlen   habe.
him recommended have

'The man read the book that I recommended to him.'

In German, it is also possible to partially front VPs, such as in sentence (6). This sentence is taken from the TüBa-D/Z treebank.

(6) Für den Berliner Job qualifiziert hat sich
For the Berlin job qualified has himself
Zimmermann auch durch seinen Blick fürs
Zimmermann also by his view for the
finanziell Machbare.
financially doable

'Zimmermann qualified for the job in Berlin partially because of his view for what is financially feasible.'

Here, the canonical word order would be *Zimmermann hat sich auch durch seinen Blick fürs finanziell Machbare für den Berliner Job qualifiziert.*

Such discontinuous structures occur frequently in the TIGER and TüBa-D/Z treebanks and are handled differently in the two annotation schemes, as will be discussed in more detail in section 4.

3 Task Definition

In this section, we give the definition of the shared task. We provided two subtasks: parsing constituent structure and parsing the dependency representations. Both subtasks involved training and testing on data from the two treebanks, TIGER and TüBa-D/Z. The dependency format was derived from the constituent format so that the sentences were identical in the two versions. The participants were given training sets, development sets, and test sets of the two treebanks. The training sets contained 20894 sentences per treebank, the development and test set consisted of 2611 sentences each. The test sets contained gold standard POS labels. In these sets, sentence length was restricted to a maximum of 40 words. Since for some sentences in both treebanks, the annotation consists of more than one tree, all trees were joined under a virtual root node, *VROOT*.

Since some parsers cannot assign grammatical functions to part of speech tags, these grammatical functions were provided for the test data as attached to the POS tags. Participants were asked to perform a test without these functions if their parser was equipped to provide them. Two participants did submit these results, and in both cases, these results were considerably lower.

Evaluation for the constituent version consisted of the PARSEVAL measures precision, recall, and F_1 measure. All these measures were calculated on combinations of constituent labels and grammatical functions. Part of speech labels were not considered in the evaluation. Evaluation for the dependency version consisted of labeled and unlabeled attachment scores. For this evaluation, we used the scripts provided by the CoNLL shared task 2007 on dependency parsing (Nivre et al., 2007).

4 The Treebanks

The two treebanks used for the shared task were the TIGER Corpus, (Brants et al., 2002) version 2, and the TüBa-D/Z treebank (Hinrichs et al., 2004; Telljohann et al., 2006), version 3. Both treebanks use German newspapers as their data source: the Frankfurter Rundschau newspaper for TIGER and the 'die tageszeitung' (taz) newspaper for TüBa-D/Z. The average sentence length is very similar: In TIGER, sentences have an average length of 17.0, and in TüBa-D/Z, 17.3. This can be regarded as an indication that the complexity of the two texts is comparable. Both treebanks use the same POS tagset, STTS (Thielen and Schiller, 1994), and annotations based on phrase structure grammar, enhanced by a level of predicate-argument structure.

4.1 The Constituent Data

Despite all the similarities presented above, the constituent annotations differ in four important aspects: 1) TIGER does not allow for unary branching whereas TüBa-D/Z does; 2) in TIGER, phrase internal annotation is flat whereas TüBa-D/Z uses phrase internal structure; 3) TIGER uses crossing branches to represent long-distance relationships whereas TüBa-D/Z uses a pure tree structure combined with functional labels to encode this information. The two treebanks also use different notions of grammatical functions: TüBa-D/Z defines 36 grammatical functions covering head and non-head information, as well as subcategorization for complements and modifiers. TIGER utilizes 51 grammatical functions. Apart from commonly accepted grammatical functions, such as *SB* (subject) or *OA* (accusative object), TIGER grammatical functions in-
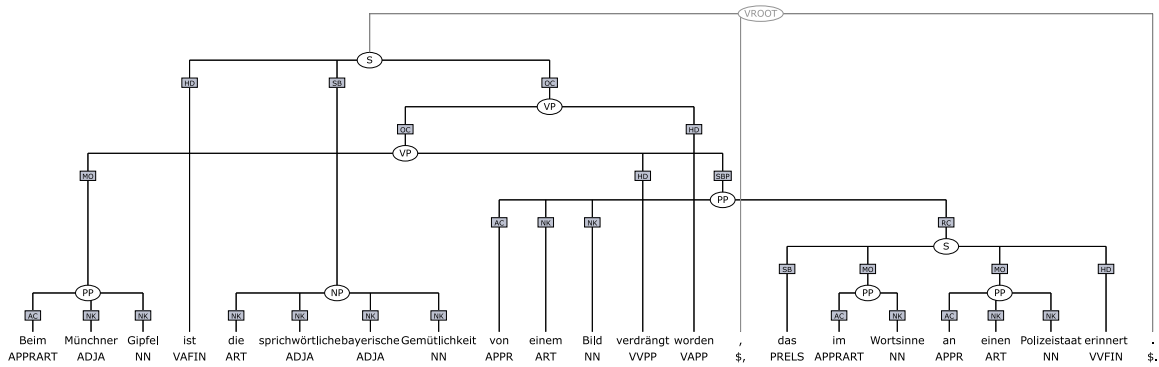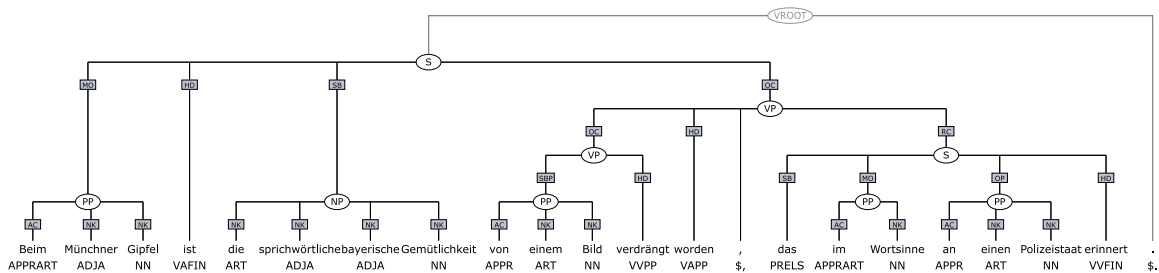
Figure 1: TIGER annotation with crossing branches.

Figure 2: TIGER annotation with resolved crossing branches.

clude others, e.g. *RE* (repeated element) or *RC* (relative clause).

(7) Beim Münchner Gipfel   ist die
     At the Munich   Summit is the
     sprichwörtliche bayerische Gemütlichkeit
     proverbial   Bavarian 'Gemütlichkeit'
     von einem Bild   verdrängt worden, das
     by a   picture supplanted been,   which
     im   Wortsinne an einen Polizeistaat
     in the literal sense of a   police state
     erinnert.
     reminds

     'At the Munich Summit, the proverbial Bavarian 'Gemütlichkeit' was supplanted by an image that is evocative of a police state.'

Figure 1 shows a typical tree from the TIGER treebank for sentence (7). The syntactic categories are shown in circular nodes, the grammatical functions as edge labels in square boxes. A major phrasal category that serves to structure the sentence as a whole is the verb phrase (VP). It contains non-finite verbs (here: *verdrängt worden*) as well as their complements and adjuncts. The subject NP (*die sprichwörtliche bayerische Gemütlichkeit*) is outside the VP and, depending on its linear position, leads to crossing branches with the VP. This happens in all cases where the subject follows the finite verb as in Figure 1. Notice also that the PPs are completely flat. An additional crossing branch results from the direct attachment of the extraposed relative clause (the lower S node with function RC) to the noun that it modifies.

As mentioned in the previous section, TIGER trees must be transformed into trees without crossing branches prior to training PCFG parsers. The standard approach for this transformation is to re-attach crossing non-head constituents as sisters of the lowest mother node that dominates all the crossing constituent and its sister nodes in the original TIGER tree. Figure 2 shows the result of this transformation
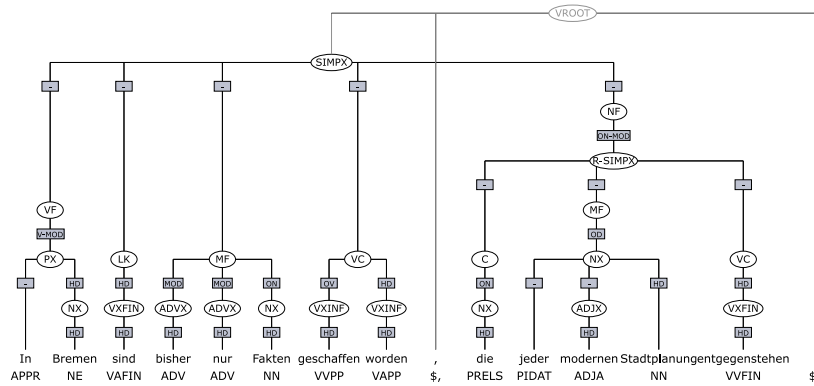
58

Figure 3: TüBa-D/Z annotation without crossing branches.

of the tree in Figure 1. Crossing branches not only arise with respect to the subject at the sentence level but also in cases of extraposition and fronting of partial constituents. As a result, approximately 30% of all TIGER trees contain at least one crossing branch. Thus, tree transformations have a major impact on the type of constituent structures that are used for training probabilistic parsing models.

Figure 3 shows the TüBa-D/Z annotation for sentence (8), a sentence with a very similar structure to the TIGER sentence shown in Figure 1. Crossing branches are avoided by the introduction of topological structures (here: VF, LK, MF, VC, NF, and C) into the tree. Notice also that compared to the TIGER annotation, TüBa-D/Z introduces more internal structure into NPs and PPs. In TüBa-D/Z, long-distance relationships are represented by a pure tree structure and specific functional labels. Thus, the extraposed relative clause is attached to the matrix clause directly, but its functional label *ON-MOD* explicates that it modifies the subject *ON*.

(8) In Bremen sind bisher nur Fakten geschaffen
    In Bremen are so far only facts produced
    worden, die jeder modernen Stadtplanung
    been, which any modern city planning
    entgegenstehen.
    contradict

    'In Bremen, so far only such attempts have been made that are opposed to any modern city planning.'

4.2 The Dependency Data

The constituent representations from both treebanks were converted into dependencies. The conversion aimed at finding dependency representations for both treebanks that are as similar to each other as possible. Complete identity is impossible because the treebanks contain different levels of distinction for different phenomena. The conversion is based on the original formats of the treebanks including crossing branches. The target dependency format was defined based on the dependency grammar by Foth (2003). For the conversion, we used pre-existing dependency converters for TIGER trees (Daum et al., 2004) and for TüBa-D/Z trees (Versley, 2005). The dependency representations of the trees in Figures 1 and 3 are shown in Figures 4 and 5. Note that the long-distance relationships are converted into non-projective dependencies.

5 Submissions and Results

The shared task drew submissions from 3 groups: the Berkeley group, the Stanford group, and the Växjö group. Four more groups or individuals had registered but did not submit any data. The submitted systems and results are described in detail in papers in this volume (Petrov and Klein, 2008; Rafferty and Manning, 2008; Hall and Nivre, 2008). All three systems submitted results for the constituent task. For the dependency task, the Växjö group had the only submission. For this reason, we will concentrate on the analysis of the constituent results and will mention the dependency results only shortly.
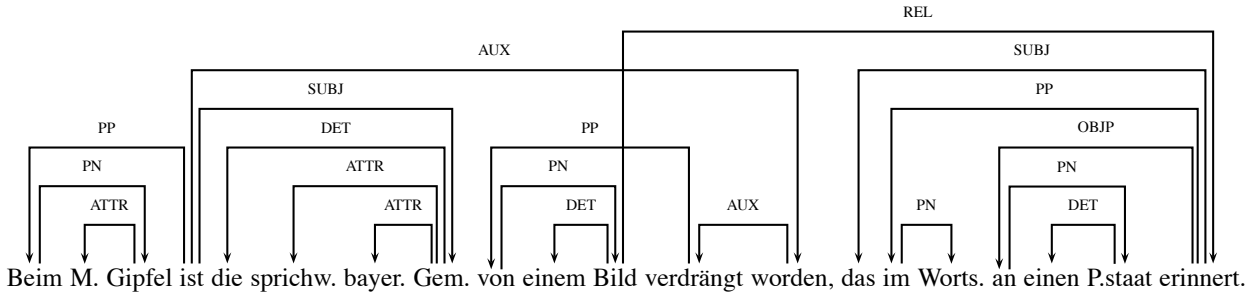
59

Beim M. Gipfel ist die sprichw. bayer. Gem. von einem Bild verdrängt worden, das im Worts. an einen P.staat erinnert.

Figure 4: TIGER dependency annotation.

In Bremen sind bisher nur Fakten geschaffen worden, die jeder modernen Stadtplanung entgegenstehen.
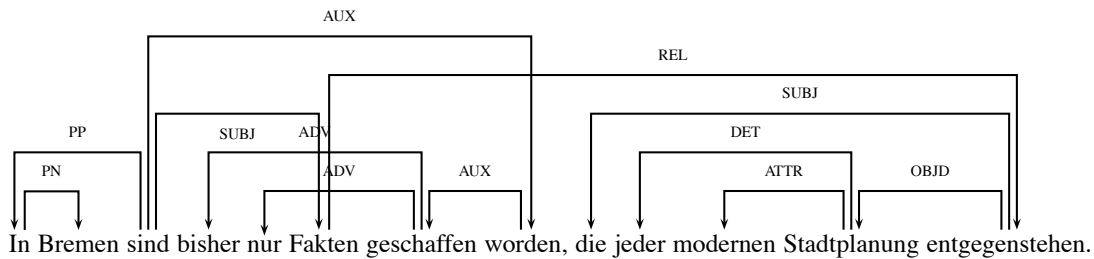
Figure 5: TüBa-D/Z dependency annotation.

5.1 Constituent Evaluation

The results of the constituent analysis are shown in Table 1. The evaluation was performed with regard to labels consisting of a combination of syntactic labels and grammatical functions. A subject noun phrase, for example, is only counted as correct if it has the correct yield, the correct label (i.e. *NP* for TIGER and *NX* for TüBa-D/Z), and the correct grammatical function (i.e. *SB* for TIGER and *ON* for TüBa-D/Z). The results show that the Berkeley parser reaches the best results for both treebanks. The other two parsers compete for second place. For TIGER, the Växjö parser outperforms the Stanford parser, but for TüBa-D/Z, the situation is reversed. This gives an indication that the Växjö parser seems better suited for the flat annotations in TIGER while the Stanford parser is better suited for the more hierarchical structure in TüBa-D/Z. Note that all parsers reach much higher F-scores for TüBa-D/Z.

A comparison of how well suited two different annotation schemes are for parsing is a surprisingly difficult task. A first approach would be to compare the parser performance for specific categories, such as for noun phrases, etc. However, this is not possible for TIGER and TüBa-D/Z. On the one

hand, the range of phenomena described as noun phrases, for example, is different in the two treebanks. The most obvious difference in annotation schemes is that TüBa-D/Z annotates unary branching structures while TIGER does not. As a consequence, in TüBa-D/Z, all pronouns and substituting demonstratives are annotated as noun phrases; in TIGER, they are attached directly to the next higher node (cf. the relative pronouns, POS tag PRELS, in Figures 1 and 3). Kübler (2005) and Maier (2006) suggest a method for comparing such different annotation schemes by approximating them stepwise so that the decisions which result in major changes can be isolated. They come to the conclusion that the differences between the two annotation schemes is a least partially due to inconsistencies introduced into TIGER style annotations during the resolution of crossing branches. However, even this method cannot give any indication which annotation scheme provides more useful information for systems that use such parses as input. To answer this question, an *in vivo* evaluation would be necessary. It is, however, rather difficult to find systems into which a parser can be plugged in without too many modifications of the system.

On the other hand, it is a well-known fact that

60

	TIGER			TüBa-D/Z		
system	precision	recall	F-score	precision	recall	F-score
Berkeley	69.23	70.41	69.81	83.91	84.04	83.97
Stanford	58.52	57.63	58.07	79.26	79.22	79.24
Växjö	67.06	63.40	65.18	76.44	74.79	75.60

Table 1: The results of the constituent parsing task.

		TIGER			TüBa-D/Z		
system	GF	precision	recall	F-score	precision	recall	F-score
Berkeley	SB/ON	74.46	78.31	76.34	78.33	77.08	77.70
	OA	60.08	66.61	63.18	58.11	65.81	61.72
	DA/OD	49.28	41.72	43.19	59.46	44.72	51.05
Stanford	SB/ON	64.40	63.11	63.75	71.16	77.76	74.31
	OA	45.52	45.91	45.71	47.23	51.28	49.17
	DA/OD	12.40	9.82	10.96	24.42	8.54	12.65
Växjö	SB/ON	75.33	73.00	74.15	72.37	69.53	70.92
	OA	57.01	57.65	57.33	58.07	57.55	57.81
	DA/OD	55.45	37.42	44.68	63.75	20.73	31.29

Table 2: The results for subjects, accusative objects, and dative objects.

the PARSEVAL measures favor annotation schemes with hierarchical structures, such as in TüBa-D/Z, in comparison to annotation schemes with flat structures (Rehbein and van Genabith, 2007). Here, TIGER and TüBa-D/Z differ significantly: in TIGER, phrases receive a flat annotation. Prepositional phrases, for example, do not contain an explicitly annotated noun phrase. TüBa-D/Z phrases, in contrast, are more hierarchical; preposition phrases do contain a noun phrase, and non phrases distinguish between pre- and post-modification. For this reason, the evaluation presented in Table 1 must be taken with more than a grain of salt as a comparison of annotation schemes. However, it seems safe to follow Kübler et al. (Kübler et al., 2006) in the assumption that the major grammatical functions, subject (*SB/ON*), accusative object (*OA*), and dative object (*DA/OD*) are comparable. Again, this is not completely true because in the case of one-word NPs, these functions are attached to the POS tags and thus are given in the input. Another solution, which was pursued by Rehbein and van Genabith (2007), is the introduction of new unary branching nodes in the tree in cases where such grammatical functions are originally attached to the POS tag. We refrained

from using this solution because it introduces further inconsistencies (only a subset of unary branching nodes are explicitly annotated), which make it difficult for a parser to decide whether to group such phrases or not. The evaluation shown in Table 2 is based on all nodes which were annotated with the grammatical function in question.

The results presented in Table 2 show that the differences between the two treebanks are inconclusive. While the Stanford parser performs consistently better on TüBa-D/Z, the Berkeley parser handles accusative objects better in TIGER, and the Växjö parser subjects and dative objects. The results indicate that the Berkeley parser profits from the TIGER annotation of accusative objects, which are grouped in the verb phrase while TüBa-D/Z groups all objects in their fields directly without resorting to a verb phrase. However, this does not explain why the Berkeley parser cannot profit from the subject attachment on the clause level in TIGER to the same degree.

5.2 Dependency Evaluation

The results of the dependency evaluation for the Växjö system are shown in Table 3. The results are

	TIGER		TüBa-D/Z	
UAS	92.63		91.45	
LAS	90.80		88.64	
	precision	recall	precision	recall
SUBJ	90.20	89.82	88.99	88.55
OBJA	77.93	82.19	77.18	82.71
OBJD	57.00	44.02	67.88	45.90

Table 3: The results of the dependency evaluation.

important for the comparison of constituent and dependency parsing since in the conversion to dependencies, most of the differences between the annotation schemes, and as a consequence, the preference of the PARSEVAL measures have been neutralized. Therefore, it is interesting to see that the results for TIGER are slightly better than the results for TüBa-D/Z, both for unlabeled (UAS) and labeled attachment scores. The reasons for these differences are unclear: either the TIGER texts are easier to parse, or the (original annotation and) conversion from TIGER is more consistent. Another surprising fact is that the dependency results are clearly better than the constituent ones. This is partly due to the fact that the dependency representation is often less informative than then constituent representation. One example for this can be found in coordinations: In dependency representations, the scope ambiguity in phrases like *young men and women* is not resolved. This gives parsers fewer opportunities to go wrong. However, this cannot explain all the differences. Especially the better performance on the major grammatical functions cannot be explained in this way.

A closer look at the grammatical functions shows that here, precision and recall are higher than for constituent parses. This is a first indication that dependency representation may be more appropriate for languages with freer word order. A comparison between the two treebanks is inconclusive: for the accusative object, the results are similar between the treebanks. For subjects, the results for TIGER are better while for dative objects, the results for TüBa-D/Z are better. This issue requires closer investigation.

6 Conclusion

This is the first shared task on parsing German, which provides training and test sets from both major treebanks for German, TIGER and TüBa-D/Z. For both treebanks, we provided a constituent and a dependency representation. It is our hope that these data sets will spark more interest in the comparison of different annotation schemes and their influence on parsing results. The evaluation of the three participating systems has shown that for both treebanks, the use of a latent variable grammar in the Berkeley system is beneficial. However, many questions remain unanswered and require further investigation: To what extent do the evaluation metrics distort the results? Does a measure exist that is neutral towards the differences in annotation? Is the dependency format better suited for parsing German? Are the differences between the dependency results of the two treebanks indicators that TIGER provides more important information for dependency parsing? Or can the differences be traced back to the conversion algorithms?

References

Cristina Bosco, Vincenzo Lombardo, D. Vassallo, and Leonardo Lesmo. 2000. Building a treebank for Italian: a data-driven annotation scheme. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation, LREC-2000*, Athens, Greece.

Sabine Brants, Stefanie Dipper, Silvia Hansen, Wolfgang Lezius, and George Smith. 2002. The TIGER treebank. In Erhard Hinrichs and Kiril Simov, editors, *Proceedings of the First Workshop on Treebanks*

and Linguistic Theories (TLT 2002), pages 24–41, Sozopol, Bulgaria.

Michael Daum, Kilian Foth, and Wolfgang Menzel. 2004. Automatic transformation of phrase treebanks to dependency trees. In *Proceedings of the 4th International Conference on Language Resources and Evaluation, LREC-2004*, Lisbon, Portugal.

Kilian Foth. 2003. Eine umfassende Dependenzgrammatik des Deutschen. Technical report, Fachbereich Informatik, Universität Hamburg.

Johan Hall and Joakim Nivre. 2008. A dependency-driven parser for German dependency and constituency representations. In *Proceedings of the ACL Workshop on Parsing German*, Columbus, OH.

Erhard Hinrichs, Sandra Kübler, Karin Naumann, Heike Telljohann, and Julia Trushkina. 2004. Recent developments in linguistic annotations of the TüBa-D/Z treebank. In *Proceedings of the Third Workshop on Treebanks and Linguistic Theories*, pages 51–62, Tübingen, Germany.

Sandra Kübler, Erhard W. Hinrichs, and Wolfgang Maier. 2006. Is it really that difficult to parse German? In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, EMNLP 2006*, pages 111–119, Sydney, Australia.

Sandra Kübler. 2005. How do treebank annotation schemes influence parsing results? Or how not to compare apples and oranges. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing, RANLP 2005*, pages 293–300, Borovets, Bulgaria.

Wolfgang Maier. 2006. Annotation schemes and their influence on parsing results. In *Proceedings of the ACL-2006 Student Research Workshop*, Sydney, Australia.

Mitchell Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330.

S. Montegmagni, F. Barsotti, M. Battista, N. Calzolari, O. Corazzari, A. Zampolli, F. Fanciulli, M. Massetani, R. Raffaelli, R. Basili, M. T. Pazienza, D. Saracino, F. Zanzotto, N. Mana, F. Pianesi, and R. Delmonte. 2000. The Italian syntactic-semantic treebank: Architecture, annotation, tools and evaluation. In *Proceedings of the Workshop on Linguistically Interpreted Corpora LINC-2000*, pages 18–27, Luxembourg.

Joakim Nivre, Johan Hall, Sandra Kübler, Ryan McDonald, Jens Nilsson, Sebastian Riedel, and Deniz Yuret. 2007. The CoNLL 2007 shared task on dependency parsing. In *Proceedings of the CoNLL 2007 Shared Task. Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, EMNLP-CoNLL 2007*, Prague, Czech Republic.

Slav Petrov and Dan Klein. 2008. Parsing German with language agnostic latent variable grammars. In *Proceedings of the ACL Workshop on Parsing German*, Columbus, OH.

Anna Rafferty and Christopher Manning. 2008. Parsing three German treebanks: Lexicalized and unlexicalized baselines. In *Proceedings of the ACL Workshop on Parsing German*, Columbus, OH.

Ines Rehbein and Josef van Genabith. 2007. Treebank annotation schemes and parser evaluation for German. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, EMNLP-CoNLL*, pages 630–639, Prague, Czech Republic.

Geoffrey Sampson. 1993. The SUSANNE corpus. *ICAME Journal*, 17:125 – 127.

Wojciech Skut, Thorsten Brants, Brigitte Krenn, and Hans Uszkoreit. 1998. A linguistically interpreted corpus of German newspaper texts. In *ESSLLI Workshop on Recent Advances in Corpus Annotation*, Saarbrücken, Germany.

Heike Telljohann, Erhard W. Hinrichs, Sandra Kübler, and Heike Zinsmeister, 2006. *Stylebook for the Tübingen Treebank of Written German (TüBa-D/Z)*. Seminar für Sprachwissenschaft, Universität Tübingen, Germany.

Christine Thielen and Anne Schiller. 1994. Ein kleines und erweitertes Tagset fürs Deutsche. In Helmut Feldweg and Erhard Hinrichs, editors, *Lexikon & Text*, pages 215–226. Niemeyer, Tübingen.

Yannick Versley. 2005. Parser evaluation across text types. In *Proceedings of the Fourth Workshop on Treebanks and Linguistic Theories, TLT 2005*, pages 209–220, Barcelona, Spain.

Sean Wallis and Gerald Nelson. 2006. The British component of the International Corpus of English. Release 2. CD-ROM. London: Survey of English Usage, UCL.