Discosuite – A Parser Test Suite for German Discontinuous Structures

Wolfgang Maier[†], Miriam Kaeshammer[†], Peter Baumann[‡], Sandra Kübler[◇]

University of Düsseldorf[†]; Northwestern University[‡]; Indiana University[◇]
Düsseldorf, Germany; Evanston, IL, USA; Bloomington, IN, USA
maierw@hhu.de, kaeshammer@phil.hhu.de, peter.baumann@u.northwestern.edu, skuebler@indiana.edu

Abstract

Parser evaluation traditionally relies on evaluation metrics which deliver a single aggregate score over all sentences in the parser output, such as PARSEVAL. However, for the evaluation of parser performance concerning a particular phenomenon, a test suite of sentences is needed in which this phenomenon has been identified. In recent years, the parsing of discontinuous structures has received a rising interest. Therefore, in this paper, we present a test suite for testing the performance of dependency and constituency parsers on non-projective dependencies and discontinuous constituents for German. The test suite is based on the newly released TIGER treebank version 2.2. It provides a unique possibility of benchmarking parsers on non-local syntactic relationships in German, for constituents and dependencies. We include a linguistic analysis of the phenomena that cause discontinuity in the TIGER annotation, thereby closing gaps in previous literature. The linguistic phenomena we investigate include extraposition, a placeholder/repeated element construction, topicalization, scrambling, local movement, parentheticals, and fronting of pronouns.

Keywords: evaluation, parsing, discontinuity

1. Introduction

The most common way of evaluating the performance of a parser is to compare its output to gold data using an evaluation metric. The standard metric for the evaluation of dependency parsers is the *(labeled) attachment score*, which measures the rate of words that have been assigned the correct head (and the correct dependency label). The standard metric for constituency parsers is *PARSEVAL* (Black et al., 1991), which measures the number of matching brackets in the parser output and the gold data on the basis of precision, recall and F-score. Alternative metrics are, e.g., the *Leaf-Ancestor (LA)* metric (Sampson and Babarczy, 2003) and tree-distance based metrics (Emms, 2008; Tsarfaty et al., 2012).

What all parser evaluation metrics have in common is that they describe the quality of the parser output in terms of a single aggregate score. This score, however, does not necessarily reflect the ability of the parser to recover particular types of constructions that are difficult to parse or are of specific interest. For the evaluation of parser performance on such constructions a test suite is needed, i.e., a collection of sentences in which a certain difficulty of interest has been identified.

In this paper, we present *discosuite*, a test suite of German sentences containing discontinuous structures. The sentences are taken from the TIGER treebank release 2.2 (Brants et al., 2004). In the constituency version of TIGER, discontinuous constituents are annotated directly by allowing crossing branches. This way, all arguments of a predicate can always be grouped under the same node. This contrasts with other treebanks, such as the Penn Treebank (Marcus et al., 1993), where long-distance relations are marked via empty categories and traces, and the German TüBa-D/Z (Telljohann et al., 2012), where such phenomena are marked via specific labels. In the dependency version of TIGER, which is derived automatically from the constituency version, discontinuous constituents generally result in non-projective dependencies, the dependency counterpart of discontinuous constituents. Figure 1 shows an example of the constituent and dependency annotation of (1) in the TIGER treebank.

(1) [Teure   Detektive und kostspielige Elektronik] kann
     [Expensive detectives and costly   electronics] can
     [sich] der Supermarkt an der Ecke   nicht [leisten].
     [itself] the supermarket on the corner not   [afford].
     "The supermarket on the corner cannot afford expensive detectives and costly electronics."

The constituency annotation contains a VP with two gaps, due to the topicalized accusative object and the reflexive pronoun located left of the subject. Since the dependency tree is the result of a conversion of the constituency tree, it shows the same discontinuities.

The selection of sentences is based on our analysis of the linguistic phenomena which are annotated in TIGER as discontinuous constituents and non-projective dependencies, respectively.

The motivation for our work is to provide an important resource for parser evaluation. In the dependency parsing community, much effort has been put into the efficient parsing of non-projective structures (McDonald et al., 2005; McDonald and Pereira, 2006; Nivre et al., 2007; Bohnet and Nivre, 2012); in the constituency parsing community, particularly in the last years, there has been a rising interest in the direct (e.g., van Cranenburgh (2012), Kallmeyer and Maier (2013), and Angelov and Ljunglöf (2014)) or indirect (e.g., Cai et al. (2011)) parsing of discontinuous constituents. However, to our knowledge, there is no data set which allows for cross-framework parser evaluation of discontinuities on both dependency and constituency data in the domain in which the corresponding parsers are usually trained. *Discosuite* provides this possibility.

The rest of this paper is structured as follows. In the following section, we present related work. In section 3., we
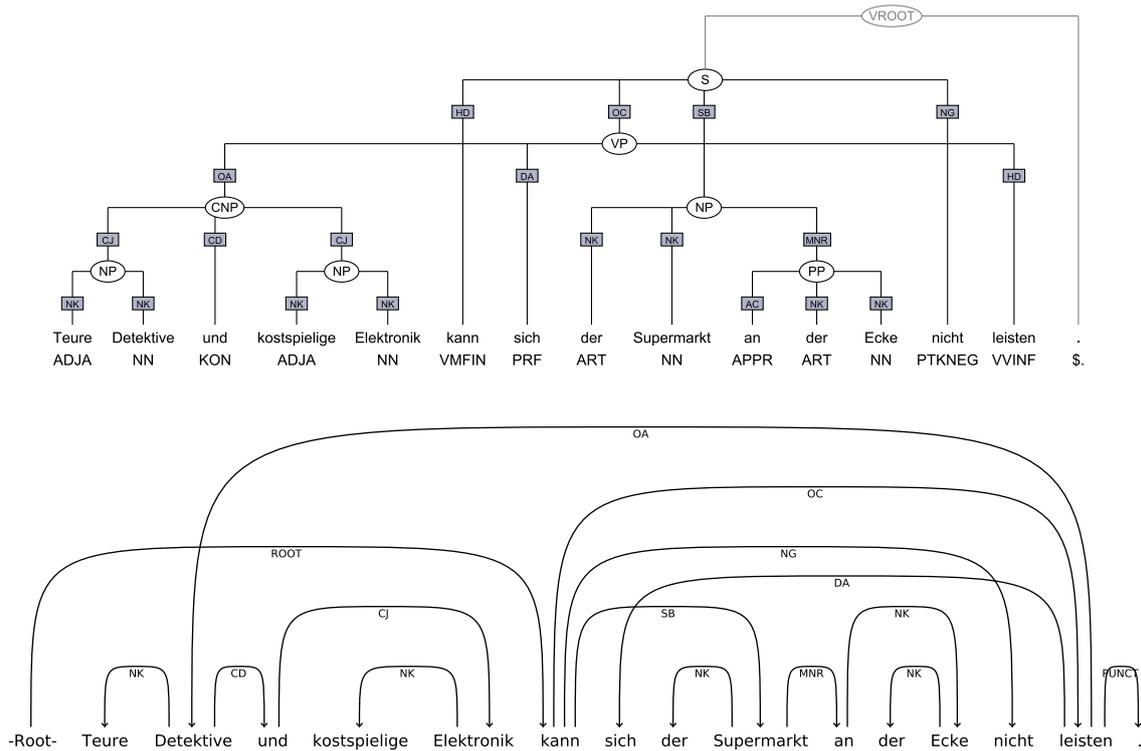
Figure 1: Discontinuity in TIGER (topicalization and scrambling), sentence 227

explain the preprocessing of the treebank, followed by the analysis of the phenomena which cause discontinuity in section 4. In section 5., we present some statistics on the test suite, section 6. closes the paper.

2. Related Work

In the grammar engineering community, several test suites have been created. They are rather different from ours and are not constructed around discontinuous or non-projective structures. They include the test suite of Forst (2003), which is based on the constituency version of the TIGER treebank, but tailored to the evaluation of a broad-coverage LFG parser, such as ParGram. The LKB system (Copestake, 1999) (HPSG) also has a test suite. TSNLP (Lehmann et al., 1996) is a general-purpose test suite, which is not based on a particular grammar formalism. Both test suites cover a range of linguistic phenomena, but do not provide a systematic overview of discontinuous phenomena.

Two of the treebank-based parser test suites we are aware of are directly related to our work.

Rimell et al. (2009) present a set of English sentences taken from the English Penn Treebank (Marcus et al., 1993), which exhibit unbounded dependencies. While the idea behind their test suite is similar to the idea behind ours, most of the constructions they consider are specific to English.

Kübler et al. (2009) present a test suite of German sentences. Their efforts are not directed at discontinuous structures, but rather at difficult constructions in general. The phenomena they consider include coordination of unlike constituents and forward conjunction reduction, but also phenomena that almost always cause discontinuities, such

as extraposed relative clauses. In contrast to the test suite by Kübler et al., our test suite is constructed explicitly around sentences that exhibit discontinuities in the dependency and the constituency annotation.

3. Data

Our data source is the TIGER treebank release 2.2, which can be obtained from `http://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/tiger.html`. We generate the TIGER dependency version using the script Tiger2Dep from the same webpage. For the punctuation attachment in the dependency version, we use the *easy* algorithm of Tiger2Dep, which attaches each punctuation token to the word to its left, thus making sure that punctuation does not give rise to discontinuity.

In the constituency version, punctuation and a few other elements are not included in the annotation; all resulting root nodes are grouped under a virtual root node. In contrast to previous releases of TIGER, sentences with a unique root node do not have a virtual root node in release 2.2. For reasons of uniformity, we reintroduce the virtual root node in all sentences where it is not present and, again to eliminate spurious gaps, reattach all of its children except the sentence lower in the tree, following common practice in the parsing community. We use the algorithm described in Maier et al. (2012) for this purpose.[1] Since our primary goal is to obtain a test suite for parser evaluation, we

---

[1] The implementation is publicly available at `https://github.com/wmaier/treetools`.

All sentences	50 472
. . . with ≥ 1 gaps (const)	14 008
. . . with ≥ 1 gaps (dep)	14 181
. . . with same no. of gaps (const & dep)	13 959

Table 1: Properties of the TIGER treebank

decided against excluding punctuation, as it was done by Maier and Lichte (2011).

Table 1 shows the ratio of sentences with gaps in the dependency and constituency treebanks. For a formal definition of gaps and gap degree in constituency and dependency trees, consult Maier and Lichte (2011). Note that the concept of gap degree was introduced for dependency structures by Kuhlmann (2007).

4.   Analysis of Phenomena

The goal of our analysis is to discover which annotation principles cause discontinuities in the annotation. Note that due to the assumptions made in the dependency conversion algorithm, not all discontinuous constituents result in non-projective dependencies and vice versa. In some cases of ellipsis, discontinuous constituency structures are converted into projective dependency structures. In some cases of pre-modification, continuous constituency structures are converted into non-projective dependencies. See section 3.1 of Seeker and Kuhn (2012) for details. In order to allow for a comparison across dependency and constituency structures, we only consider sentences in which both versions display discontinuities.

4.1.   Discontinuity and Annotation Principles

The presence or absence of discontinuities depends on the linguistic assumptions on which the annotation is based. Consider relative clause attachment as an example: If the annotation makes the assumption that a relative clause is attached to its referent NP in syntax, we have a discontinuity whenever the relative clause is extraposed to the right of the right sentence bracket.[2] If the annotation assumes relative clauses to always attach high (e.g., for an under-specification of the actual referent phrase), there would be no discontinuity in the case of extraposition. We rely on the TIGER annotation principles, i.e., we accept the annotation decisions without passing judgment concerning the decisions made in the treebank.

In the TIGER annotation, an important assumption in this respect is that all verbs are dominated by a VP except the finite verb which is directly dominated by S. Many cases of discontinuity depend on the presence of a VP: For example, if there is only a finite verb, the extraposition of a sentential modifier does not lead to a discontinuity since the modifier is attached on the S level.

4.2.   Analysis and Categorization

On the basis of an analysis of the first 1 500 sentences of the treebank exhibiting discontinuities, we present a categorization of the phenomena that cause discontinuities in the

[2]See next section concerning *topological field* terminology.

constituency annotation and consequently non-projectivity in the dependency annotation of TIGER. We use the terminology of *topological fields*, a key concept in German linguistics (Drach, 1937; Höhle, 1983). For an introduction to the corresponding terminology, refer to Telljohann et al. (2012, p. 6). Topological fields structure a clause based on the verbal elements, with the finite verb in a main clause in the left sentence bracket and the remaining verbal elements in the right bracket. All other constituents are placed in the initial field (Vorfeld) before the finite verb, the middle field (Mittelfeld) between the brackets and the final field (Nachfeld) after the right bracket.

4.2.1.   Extraposition

Extraposition denotes a dislocation of material to the right, from the Mittelfeld into the Nachfeld. The discontinuity is caused by the right sentence bracket being located between the remaining part and the dislocated part. Heavy material is more likely to be dislocated.

1. *Extraposed modifiers* can occur in many kinds of phrases. Extraposition out of NPs (or PPs, which structurally, due to the flat annotation, only differ from NPs by the additional preposition) is most common; the extracted modifiers in this case include relative clauses and manner VPs. In VPs, the extracted modifiers include, e.g., sentences, adjectival phrases (APs) and PPs.

2. *Extraposed arguments* also occur in a variety of phrases. In the case of VPs, clausal and prepositional VP arguments are extraposed. Consider (2) together with its annotation in figure 2 as an example for this case. The verb *müssen* fills the right sentence bracket, which separates the verb *aufpassen* from its argument dominated by S.

   (2)   Man wird sicherlich [aufpassen]   müssen, [daß
          One will surely   [pay attention] have-to [that
          man sie   nicht aufwertet].
          one them not   revalues]
          "One definitely will have to make sure not to revalue them."

   In the case of NPs (and again PPs), clausal and prepositional objects can be extracted. Much more rarely, there are cases of extraction of clausal objects out of APs.

3. The second and/or following conjuncts of a *coordination* can be dislocated into the Nachfeld. In cases where the right sentence bracket is not occupied, we transform the sentence into a verb-final sentence (i.e., we move the finite verb to the right sentence bracket) in order to test if it is possible to place the dislocated element in the Nachfeld. Consider (3) as an example of a sentence to apply the verb-final test. It would be transformed into the acceptable sentence (4), in which the finite verb *ist* occupies the right sentence bracket.

   (3)   Geschäftemachen ist seine Welt und nicht die
          Business   is his   world and not the
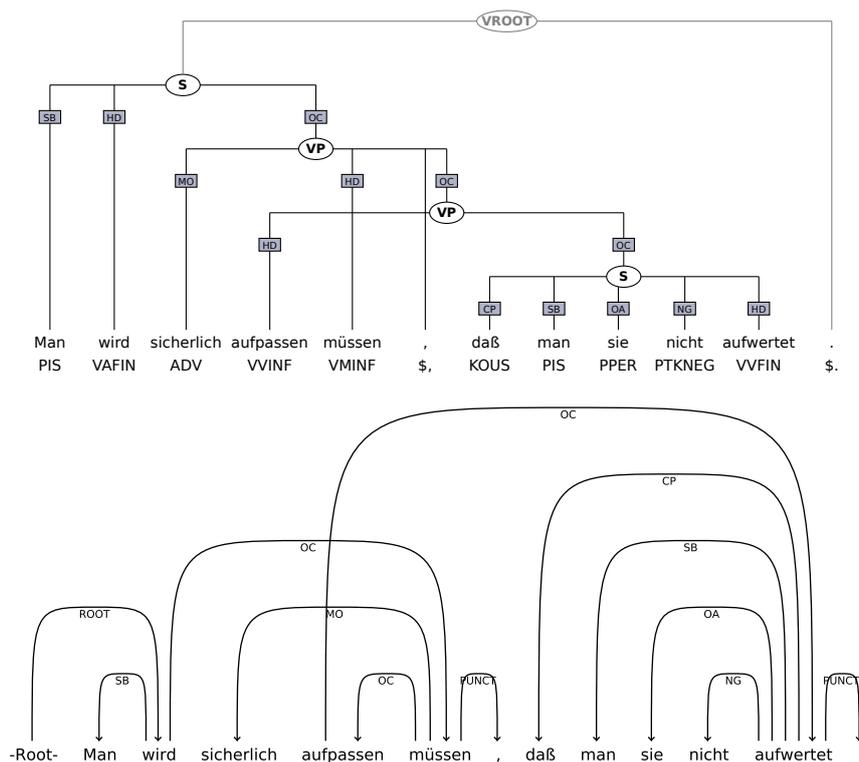
Figure 2: Discontinuity in TIGER (extraposition), sentence 148

Politik
politics
"Business is his world, not politics."

(4) . . . dass Geschäftemachen seine Welt ist und nicht
die Politik

4. *Comparisons* can be discontinuous if the comparative complement (edge label CC) is dislocated into the Nachfeld. We do not apply the verb-final test, even though it would be possible just as for extraposed coordinations. Thus, we restrict this category to only those cases in which the right sentence bracket is occupied. Cases with an empty right sentence bracket are treated under "local movement" (see section 4.2.5.). The reasoning behind this is that, unlike coordinations, comparisons are discontinuous already in their canonical position, as can be seen in (5). This sentence is not grammatical if the two parts of the AP modifying *Vorgänge* are adjacent.

(5) Bei uns gibt es [noch viel unwürdigere]
near us   exist it [even more dishonorable]
Vorgänge [als in der Politik]   .
incidents [than in the politics].
"Here, even more dishonorable incidents happen than in politics."

5. Focus adverbs in the matrix sentence modifying an embedded sentence can cause a discontinuity if the embedded sentence is located in the Nachfeld. Possible adverbs include *nur* (only) and *unter anderem* (among others). (6) is an example of such a sentence.

(6) Andererseits   sei   die Unterstützung
On the other hand would be the support
der   Wirtschaft [nur] sinnvoll , [wenn sie den
of the economy [only] sensible , [if   it the
russischen Reformprozeß beschleunige]
Russian   reform process accelerates]
"On the other hand, a support of the economy would only make sense if it accelerates the Russian reform process."

4.2.2. Placeholder/repeated element construction

The *placeholder/repeated element construction*, a particularity of the TIGER annotation, describes a structure in which a pronoun or an adverb acts as a placeholder for another "repeated" element in the sentence. One phenomenon that falls in this category is the expletive *es* (it) construction in German. The corresponding elements have the edge labels PH and RE, respectively. Since both the placeholder and the repeated element reside under the same node, a discontinuity arises if the repeated element is not directly adjacent to the placeholder. In such cases, the repeated element is mostly located in the Nachfeld. Consider the example in (7): Here, *es* is the placeholder for the repeated element *wer tatsächlich die "krummeren Finger" macht*, which is located in the Nachfeld.

(7) Denn [es] ist gar   nicht so sicher, [wer tatsächlich die
Since [it] is indeed not so sure, [who really   the
"krummeren Finger" macht].
"more bent finger" makes].
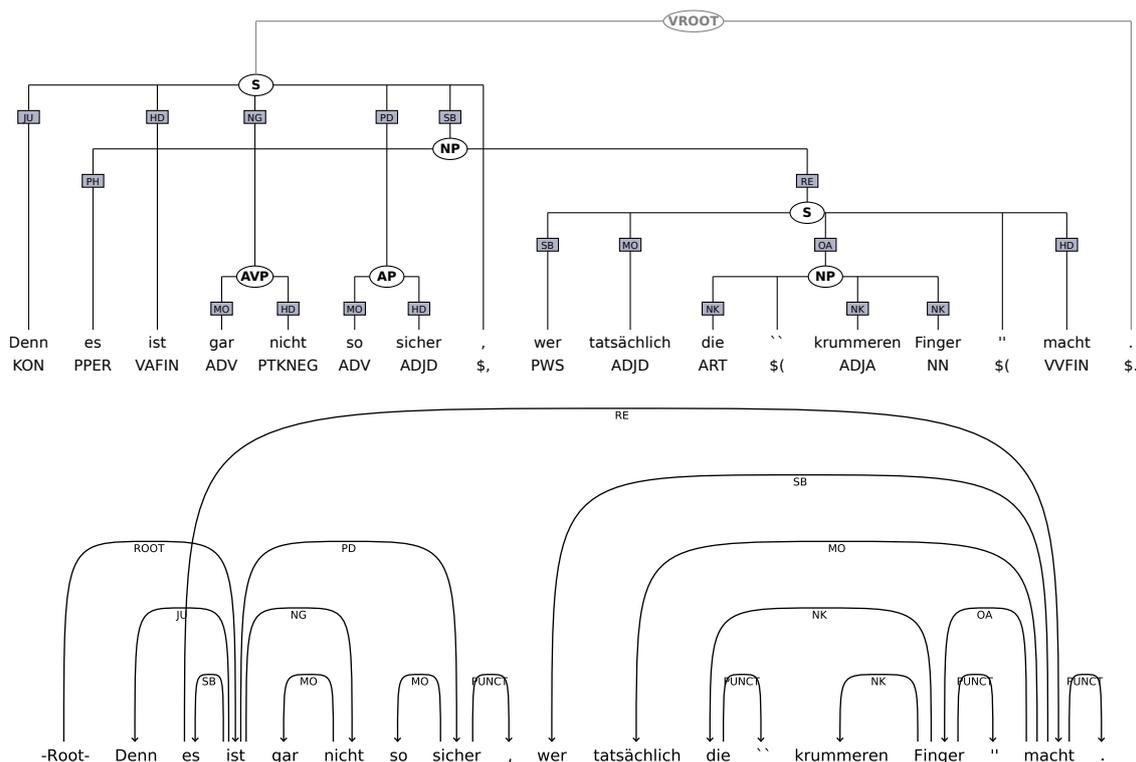"It is not that clear who is actually pilfering more."

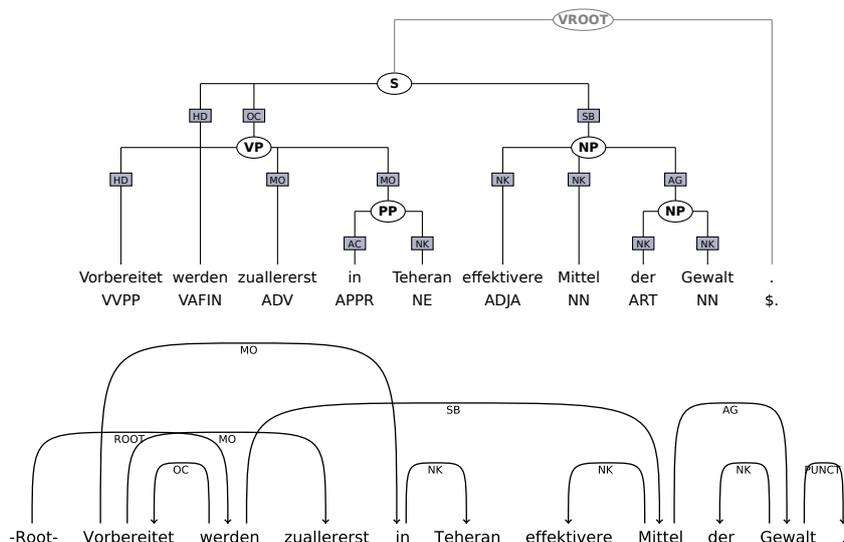Figure 3: Discontinuity in TIGER (PH/RE), sentence 234

Figure 4: Discontinuity in TIGER (topicalization of heads), sentence 395

The annotation of this sentence is shown in figure 3.

4.2.3. Topicalization

In *topicalizations*, material is dislocated to the front, generally from the Mittelfeld into the Vorfeld. The discontinuity is caused by the left sentence bracket, which separates the dislocated material in the Vorfeld from the remaining material. Again, there are several categories.

1. The topicalization of *VP modifiers or arguments* is the most common case. As an example, consider the topicalized accusative object in (1). Note that several ele-

ments can be topicalized, such as two modifiers in (8).

(8) [Unweit von hier] , [in der Gerstäckerstraße
     [Not far from here] , [in the Gerstäckerstraße
     am   berühmten Michel] , möchte der Spiegel
     next to the famous   Michel] , wants   the Spiegel
     [sein neues Domizil   errichten]
     [his new   residence build].
     "Not far from here, in the Gerstäckerstraße next to
     the famous Michel, the Spiegel wants to build his
     new residence."

2. In rare cases, *VP heads* (sometimes together with modifiers or arguments) are topicalized. Consider (9) together with its annotation in figure 4 as an example for this case.

(9) [Vorbereitet] werden [zuallererst   in
     [Prepared]   are   [first and foremost in
     Teheran] effektivere   Mittel der Gewalt.
     Teheran] more effective means of   violence.
     "First and foremost in Teheran, more effective means of violence are prepared."

3. *Modifiers or arguments of non-verbal phrases* can also be topicalized, as is the case, e.g., for the topicalized modifying adverb in (10).

(10) [Dafür]   gibt es [Gründe].
      [For this] gives it [reasons].
      "There are reasons for this."

4. More rarely, *heads of non-verbal phrases* are topicalized, also together with modifiers. As an example, consider (11), in which the head plus modifiers of an AP is topicalized.

(11) [Nicht sehr auskunftsfreudig] zeigte sich
      [Not   very willing to talk]   showed himself
      Karlheinz Kaske [über die von Siemens
      Karlheinz Kaske [about the by Siemens
      eingeleitete "Schlankheitskur"].
      introduced   "slimming diet"].
      "Karlheinz Kaske was not willing to talk much about the "slimming diet" introduced by Siemens."

Note that topicalization in non-verbal phrases occurs less frequently than in verbal phrases.

4.2.4. Scrambling

In the TIGER treebank, the subject, negation, or modifiers on the S/VP level can cross into lower VP-OCs or AP-PDs in the Mittelfeld (i.e., into clause level). This corresponds to *scrambling*, which in German linguistics usually denotes a displacement of arguments of verbs and/or nouns within the Mittelfeld to a non-canonical position (see, e.g., Müller (1995)), with the addition that discontinuities can also arise from displaced modifiers. In other words, arguments are not necessarily involved. As an example, consider the dative reflexive pronoun *sich* in (1).

4.2.5. Local Movement

Discontinuities can also occur in a rather local context, in the sense that neither of the sentence brackets is part of the intervening material which causes the discontinuity. Unlike scrambling as described in section 4.2.4., local movement covers the discontinuities on the phrase level, not on the clause level. We distinguish two types of local movement, depending on where the interrupting material comes from.

1. The interrupting material comes from the clause level. Consider (12) as an example. In this sentence, the subject *das alles* is interrupted by the sentence modifier *wohl*.

(12) Dir kommt [das] wohl   [alles] wie ein Spiel
      You seems [this] no doubt [all]   like a   game
      vor.
      VPart.
      "No doubt, all of this seems to be just a game for you."

2. The interrupting material comes from the phrase level. Consider (13) as an example where the discontinuity happens within one noun phrase.

(13) ... das [größte] geheime Sondergericht [seit
      ... the [biggest] secret   special tribunal [since
      der Unabhängigkeit des   Landes]...
      the independence   of the country]...
      "... the biggest secret special tribunal since the independence of the country..."

(13) is also an example for a comparison as described in item 4. of section 4.2.1.

4.2.6. Pronouns

Interrogative and relative pronouns are usually analyzed as occupying the left sentence bracket. A VP with a pronoun is discontinuous if the subject or modifiers (including negation) of the matrix sentence cross into it. (14) is an example for such an embedded phrase.

(14) ... [was] er eigentlich [machen] will
      ... [what] he in fact   [to do] wants
      "... what he really wants to do"

In some cases, the pronoun can be modified by a non-adjacent modifier, such as in (15).

(15) ... [was] man sich   hierzulande [alles]
      ... [what] one himself in these parts [everything]
      gefallen lasse
      like   let.
      "... what one is willing to put up with in these parts"

4.2.7. Parentheticals

Parentheticals are analyzed as embedding the subordinate clause, which therefore becomes discontinuous, see (16).

(16) [" Die Frage   ist nur "] , meint ein Finanzexperte ,
      [" The question is only "] , says a   financial expert ,
      [ob   er ins   Weiße Haus einziehen kann ...]
      [whether he into the White House move in   can   ...]
      "The question is, says a financial expert, whether he can make it to the White House ..."

In some cases, such as in (17) due to *so*, the parenthetical can dominate an AVP-PH/RE construction in which the *so* is the placeholder, and the subordinate clause is the repeated element.

(17) [Von modernem Management hat Perot] , so Fortune ,
      [Of modern   management has Perot] , so Fortune ,
      [nie   etwas   gehalten] .
      [never anything held]   .
      "Perot never held modern management in great esteem, says Fortune."

5. The Test Suite

For the creation of our test suite, we have categorized all discontinuities in the 1 500 sentences mentioned before, using the categories presented in the preceding section. Out of those sentences, for each category, we have selected 15 sentences, or less if the phenomenon occurred less than 15 times. The sentence numbers of the corresponding sentences are published at `http://phil.hhu.de/beyond-cfg/resources/discosuite`.

6. Conclusion

We have presented an analysis of phenomena causing discontinuity in the annotation of the TIGER treebank. The analysis is the basis for a test suite for evaluating data-driven constituency and dependency parsers on German discontinuous structures. The test suite closes an important gap in the literature on parser evaluation. The resource is publicly available.

7. References

Krasimir Angelov and Peter Ljunglöf. 2014. Fast statistical parsing with parallel multiple context-free grammars. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*. to appear.

Ezra Black, Steven Abney, Dan Flickenger, Claudia Gdaniec, Ralph Grishman, Philip Harrison, Donald Hindle, Robert Ingria, Frederick Jelinek, Judith Klavans, Mark Liberman, Mitchell Marcus, Salim Roukos, Beatrice Santorini, and Tomek Strzalkowski. 1991. A procedure for quantitatively comparing the syntactic coverage of English grammars. In Patti Price, editor, *Fourth DARPA Speech and Natural Language Workshop*, pages 306–311, San Mateo. Morgan Kaufmann.

Bernd Bohnet and Joakim Nivre. 2012. A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1455–1465, Jeju Island, Korea, July. Association for Computational Linguistics.

Sabine Brants, Stefanie Dipper, Peter Eisenberg, Silvia Hansen, Esther König, Wolfgang Lezius, Christian Rohrer, George Smith, and Hans Uszkoreit. 2004. TIGER: Linguistic interpretation of a German corpus. *Journal of Language and Computation*, 2:597–620.

Shu Cai, David Chiang, and Yoav Goldberg. 2011. Language-independent parsing with empty elements. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 212–216, Portland, OR.

Ann Copestake. 1999. The (new) LKB system. Technical report, CSLI, Stanford University.

Erich Drach. 1937. *Grundgedanken der deutschen Satzlehre*. Diesterweg, Frankfurt a. M.

Martin Emms. 2008. Tree distance and some other variants of Evalb. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, pages 1373–1379, Marrakech, Morocco.

Martin Forst. 2003. Treebank conversion – Establishing a testsuite for a broad-coverage LFG from the TIGER treebank. In *Proceedings of the EACL workshop on linguistically interpreted corpora (LINC03)*, pages 25–32.

Tilman Höhle. 1983. *Topologische Felder*. Ph.D. thesis, Universität zu Köln.

Laura Kallmeyer and Wolfgang Maier. 2013. Data-driven parsing using probabilistic linear context-free rewriting systems. *Computational Linguistics*, 39(1).

Sandra Kübler, Ines Rehbein, and Josef van Genabith. 2009. TePaCoC - a corpus for testing parser performance on complex German grammatical constructions. In *Proceedings of the Seventh International Workshop on Treebanks and Linguistic Theories*, pages 15–28, Groningen, The Netherlands.

Marco Kuhlmann. 2007. *Dependency Structures and Lexicalized Grammars*. Doctoral dissertation, Saarland University, Saarbrücken, Germany.

Sabine Lehmann, Stephan Oepen, Sylvie Regnier-Prost, Klaus Netter, Veronika Lux, Judith Klein, Kirsten Falkedal, Frederik Fouvry, Dominique Estival, Eva Dauphin, Hervè Compagnion, Judith Baur, Lorna Balkan, and Doug Arnold. 1996. Tsnlp: Test suites for natural language processing. In *Proceedings of the 16th conference on Computational linguistics - Volume 2*, COLING '96, pages 711–716, Stroudsburg, PA, USA. Association for Computational Linguistics.

Wolfgang Maier and Timm Lichte. 2011. Characterizing discontinuity in constituent treebanks. In *Formal Grammar. 14th International Conference, FG 2009. Bordeaux, France, July 25-26, 2009. Revised Selected Papers*, volume 5591 of *LNCS/LNAI*, pages 167–182, Berlin, Heidelberg, New York. Springer-Verlag.

Wolfgang Maier, Miriam Kaeshammer, and Laura Kallmeyer. 2012. Data-driven PLCFRS parsing revisited: Restricting the fan-out to two. In *Proceedings of the Eleventh International Conference on Tree Adjoining Grammars and Related Formalisms (TAG+11)*, Paris, France.

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

Ryan McDonald and Fernando Pereira. 2006. Online learning of approximate dependency parsing algorithms. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, pages 81–88, Trento, Italy.

Ryan McDonald, Fernando Pereira, Kiril Ribarov, and Jan Hajič. 2005. Non-projective dependency parsing using spanning tree algorithms. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, pages 523–530, Vancouver, BC.

Stefan Müller. 1995. Scrambling in german – extraction into the *Mittelfeld*. Research Report RR-97-06, DFKI.

Joakim Nivre, Johan Hall, Jens Nilsson, Atanas Chanev, Gülsen Eryigit, Sandra Kübler, Svetoslav Marinov,

and Erwin Marsi. 2007. MaltParser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2):95–135.

Laura Rimell, Stephen Clark, and Mark Steedman. 2009. Unbounded dependency recovery for parser evaluation. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 813–821, Singapore, August. Association for Computational Linguistics.

Geoffrey Sampson and Anna Babarczy. 2003. A test of the leaf-ancestor metric for parse accuracy. *Journal of Natural Language Engineering*, 9:365–380.

Wolfgang Seeker and Jonas Kuhn. 2012. Making ellipses explicit in dependency conversion for a german treebank. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, pages 3132–3139, Istanbul, Turkey, May. European Language Resources Association (ELRA). ACL Anthology Identifier: L12-1088.

Heike Telljohann, Erhard W. Hinrichs, Sandra Kübler, Heike Zinsmeister, and Kathrin Beck. 2012. Stylebook for the Tübingen Treebank of Written German (TüBa-D/Z). Technical report, Seminar für Sprachwissenschaft, Universität Tübingen, January.

Reut Tsarfaty, Joakim Nivre, and Evelina Andersson. 2012. Cross-framework evaluation for statistical parsing. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 44–54, Avignon, France, April. Association for Computational Linguistics.

Andreas van Cranenburgh. 2012. Efficient parsing with linear context-free rewriting systems. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 460–470, Avignon, France.