Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 1061–1066
Marseille, 11–16 May 2020

The Potsdam Commentary Corpus 2.2: Extending Annotations for Shallow Discourse Parsing

Peter Bourgonje, Manfred Stede
Universität Potsdam, Applied Computational Linguistics
Potsdam, Germany
{bourgonje, stede}@uni-potsdam.de

Abstract

We present the Potsdam Commentary Corpus 2.2, a German corpus of news editorials annotated on several different levels. New in the 2.2 version of the corpus are two additional annotation layers for coherence relations following the Penn Discourse TreeBank framework. Specifically, we add relation senses to an already existing layer of discourse connectives and their arguments, and we introduce a new layer with additional coherence relation types, resulting in a German corpus that mirrors the annotation scheme of the PDTB (which is a much larger corpus, though). The aim of the extended annotations is to improve usability of the corpus for the task of shallow discourse parsing. In this paper, we provide inter-annotator agreement figures for the new annotations and compare corpus statistics based on the new annotations to the equivalent statistics extracted from the PDTB.
Keywords: Discourse Relations, Discourse Parsing, Corpus Annotation

1. Introduction

The Potsdam Commentary Corpus (PCC) was first introduced by Stede (2004) as a collection of 175 newspaper editorials. It comprises over 34k words in over 2,100 sentences, sourced from a German regional newspaper. The main idea behind the PCC was to provide a single-genre corpus, collected in an "unbalanced" way, to specifically address research questions on subjectivity and argumentation. It was manually annotated on several different levels independently, allowing investigation of different linguistic phenomena related to subjectivity and argumentation and the ways they interact on syntactic, semantic and discourse level. As such, the original version was annotated for sentence syntax, coreference, and rhetorical structure. In Stede and Neumann (2014), an updated version (PCC 2.0) was presented. In addition to revisions on the rhetorical structure and coreference layers, it introduced a new layer consisting of discourse connectives and their arguments. The annotation scheme for this layer was based on that of the Penn Discourse TreeBank (PDTB) (Prasad et al., 2008). In contrast to the PDTB however, the connectives and arguments layer of the PCC 2.0 contained instances of explicit relations only, and also it did not contain the senses for its explicit relations. More recently, Bourgonje and Stede (2018b) introduced the PCC 2.1, in which a new layer containing information-structural aboutness topics and automatically produced dependency parses using the Universal Dependencies scheme were added. As part of the 2.1 release, the entire corpus was made available in the interactive corpus browser ANNIS3 (Krause and Zeldes, 2016).
In this paper, we present the PCC 2.2. In addition to a revision of the connectives and arguments layer, resulting in correction of some minor inconsistencies, and a conversion from inline XML to standoff XML format (more convenient for automatic processing), the key contributions of this version are the addition of:

- Relation senses to the connectives and arguments layer, as introduced in Stede and Neumann (2014), making this layer compliant with the PDTB-guidelines for explicit relation annotations.
- Annotations for implicit relations and three other relation types (see Section 3.).

We provide several corpus statistics based on the additional annotations, compare these to the PDTB, and include inter-annotator agreement scores over the additional annotations. The rest of this paper is structured as follows: Section 2. reviews other corpora annotated for discourse relations as well as provides pointers to applied systems making use of the kind of annotations added in the PCC 2.2. Section 3. describes the structure of the existing and added annotations for discourse relations in more detail. Section 4. provides inter-annotator agreement scores and key statistics for the new annotations. Finally, Section 5. sums up the main contributions and provides pointers to future work.

2. Related Work

Our starting point for this work is the PCC in its 2.1 version (Bourgonje and Stede, 2018b). For the layer of connectives and arguments already present in this 2.1 version, we add sense annotations based on the PDTB 3.0 sense hierarchy (Prasad et al., 2019). For reasons explained in more detail in Section 3. we use the 2.0 guidelines (Prasad et al., 2008) for the annotation of implicit relations and the remaining relation types. With these new annotations added, the nearest neighbour of the PCC 2.2 is the PDTB itself, with the main difference being the size; the PDTB has over 1m words, the PCC just over 34k words. Several other corpora annotated for coherence relations, distributed over different languages and frameworks, are conveniently summarised by Zeldes et al. (2019). Not included in this overview are the Prague Discourse Treebank (Rysová et al., 2016), which contains Czech texts, and the corpus described by Zeyrek et al. (2019), which includes six different languages. Both use the PDTB guidelines as a starting point for annotations. Annotations of this kind are typically exploited for the task of discourse parsing. Some discourse parsers that exploit Rhetorical Structure trees (Mann and Thompson, 1988) are

described by Soricut and Marcu (2003), Hernault et al. (2010) and Joty et al. (2015). Especially more recently though, encouraged by the 2015 and 2016 CoNLL shared tasks on shallow[1] discourse parsing, but also generally reflected by available training data volumes, using PDTB for this task is relatively popular (Lin et al., 2014; Oepen et al., 2016; Li et al., 2016; Wang and Lan, 2015). With the systems cited above working for English and Chinese, for German, to the best of our knowledge, no end-to-end system for discourse parsing is available. Smaller sub-tasks are discussed by Dipper and Stede (2006) and Bourgonje and Stede (2018a) (connective disambiguation) and Bourgonje and Stede (2019) (argument extraction). The main motivation for augmenting the connectives and arguments layer with senses and adding additional relation types to the corpus is to increase usability of the PCC for the task of discourse parsing.

3. Annotation Structure & Method

An example of the current state of play in the PCC 2.1 with regard to connectives and their arguments is given in Listing 1, which translates to: *And FDP-member Jürgen W. Mölleman is eagerly visiting the Near East regions, to prepare himself for this issue and to suggest himself as a new liberal foreign minister.*.

Listing 1: PCC 2.1 discourse annotation excerpt

```
Und FDP-Luftikus Jürgen W. Möllemann
    bereist seinerseits schon jetzt
    eifrig den Nahen Osten , um
<unit type="ext" id="5">
   für diesen Fall gerüstet zu sein
</unit>
<unit type="int" id="5">
   <connective id="5" relation="
       addition">und</connective>
    sich als neuer liberaler Auß
       enminister zu empfehlen .
</unit>
```

The example features one instance of a relation (assigned ID 5), which has an ext(ernal) argument (*Arg1* in PDTB vocabulary), comprising the span *für diesen...zu sein*, an int(ernal) argument (*Arg2* in PDTB vocabulary), comprising the span *sich als...zu empfehlen.*, and a connective (*und* (and)). The shallow nature of the connectives and arguments layer is demonstrated by this example, where there is no annotation for the span *Und FDP-Luftikus...Osten, um*. Additionally, the example illustrates that the connective has no relation sense assigned to it. Thus, in our 2.2 version, we add this relation sense according to the PDTB 3.0 sense hierarchy. The hierarchy has three levels; the first level consists of four coarse sense classes (*Temporal, Contingency, Comparison, Expansion*). On the second level, the hierarchy further distinguishes 22 types, adding more detail to the relation sense. The third level contains 28 sub-types that specify the order of the two arguments of the relation. We refer the reader to Prasad et al. (2019) for more information.

In addition to adding relation senses, we annotate implicit relations and the three remaining relation types of AltLex, EntRel and NoRel. We follow the PDTB 2.0 annotation manual, which instructs the annotator to annotate adjacent sentences within the same paragraph that are not already related through an explicit connective. This deviates from the more recent PDTB 3.0 manual, which also includes intra-sentential implicit relations. Due to the time-consuming nature of the annotation process, we first focused on inter-sentential relations, and leave intra-sentential (non-explicit) relations to future work for now. With regard to senses (for implicit relations), we do follow the PDTB 3.0, to align senses with explicit relations. Thus, for two adjacent sentences in the same paragraph not linked by an explicit relation, the annotator was instructed to assign the relevant sense according to the PDTB 3.0 sense hierarchy, and insert the discourse connective that would fit between the two sentences. If inserting a connective between the two sentences would somehow feel redundant or unnatural, the annotator furthermore had the option to select an AltLex (alternative lexicalisation), indicating that there is a word or phrase that expresses the relation but does not strictly match the connective definition used during the explicit relation annotation phase. An (English) example of an AltLex is *three days later*, being an alternative lexicalisation of the connective *later*. For AltLexes, the annotator also had to assign a relation sense.

If no clear relation sense could be inferred for a given sentence pair, there was the option to select either an EntRel (for cases where the two sentences talk about the same entities, but do not clearly express a particular kind of relation between the propositions expressed therein) or a NoRel (for sentences that were not related in any of the above described ways).

Regarding the actual tools used for the annotation, two particularly relevant ones specifically aimed at discourse relation annotation exist: Connanno (Stede and Heintze, 2004) and the PDTB Annotator (Lee et al., 2016). The explicit relations in the PCC had been annotated using Connanno. However, since Conanno is triggered by (explicit) connectives and skips any text not containing a connective, it is not useful for relations without overtly realised connectives. For explicit connectives, it allows selecting the sense from a dropdown-menu, limiting the annotator to a set of senses pre-defined by the dictionary used by Conanno. Since we wanted to allow for additional (new) senses for given connectives, we did not use Connanno for the sense annotation task either.

The PDTB Annotator does allow the annotator to annotate all the required types of annotations, but is meant to work on plain text. Since we already have explicit relations in our corpus, we would have to adapt it to ignore any material already annotated. This is why we went for a more simplistic option. For the sense annotation task, the annotators were asked to specify the relevant sense directly in the XML attributes. For the implicit (and other) relations, the corpus was exported to tab-separated format, containing one sentence per line. If an explicit relation was already linking

[1] 'Shallow' here means that only individual relations instances are marked, and there is no commitment made to an overarching text structure.

two sentences, this was specified and the annotator could skip the particular sentence pair. For sentence pairs not already linked, in the second, third and fourth column, the annotator was asked to specify relation type (implicit, AltLex, EntRel or NoRel), relation sense (if applicable) and (covert) connective, respectively. The connective only had to be specified for implicit relations and should be the connective that could be inserted between the two sentences, but crucially is not there (otherwise it would have been an explicit relation). The annotators worked with this tab-separated format in their spreadsheet editor of choice and the results were converted back into XML format.

Afterwards, for both annotation tasks, a series of XML syntax checks and general consistency checks were done to catch spelling mistakes and other inconsistencies (upper- vs. lower-case for the sense specification, for example).

Listing 2: PCC 2.2 discourse annotation excerpt

```
<relation relation_id="5"
    pdtb3_sense="Expansion.Conjunction"
    type="explicit">
  <connective_tokens>
    <connective_token id="62"
      token="und"/>
  </connective_tokens>
  <ext_arg_tokens>
    <ext_arg_token id="56"
      token="für"/>
    <ext_arg_token id="57"
      token="diesen"/>
    <ext_arg_token id="58"
      token="Fall"/>
    <ext_arg_token id="59"
      token="gerüstet"/>
    <ext_arg_token id="60"
      token="zu"/>
    <ext_arg_token id="61"
      token="sein"/>
  </ext_arg_tokens>
  <int_arg_tokens>
    <int_arg_token id="62"
      token="und"/>
    <int_arg_token id="63"
      token="sich"/>
    <int_arg_token id="64"
      token="als"/>
    <int_arg_token id="65"
      token="neuer"/>
    <int_arg_token id="66"
      token="liberaler"/>
    <int_arg_token id="67"
      token="Außenminister"/>
    <int_arg_token id="68"
      token="zu"/>
    <int_arg_token id="69"
      token="empfehlen"/>
    <int_arg_token id="70"
      token="."/>
  </int_arg_tokens>
</relation>
```

To facilitate automatic processing of the additional annotations, we converted the inline XML format as illustrated in Listing 1 to stand-off XML format. In this format, each token has a node containing its unique ID and its string value, and each relation has a node containing its unique ID, its relation sense and sub-nodes for connective, Arg1 and Arg2 tokens[2]. The new stand-off version of the relation displayed in Listing 1 is illustrated in Listing 2.

4. Results & Evaluation

In this section we first report on inter-annotator agreement for the two types of new annotations. Subsequently, both types of annotations are merged and corpus statistics including both are presented, and put into perspective by comparing them to the PDTB.

4.1. Inter-annotator Agreement

The entire PCC was annotated for both senses and new relations by the first author of this paper. We calculated agreement for both new annotation types, though the parts of the corpus that were double-annotated differ for both agreement calculations.

For the annotation of senses for the connectives and their arguments, we selected 17 documents, containing 108 relations (~10% of all annotations for connectives and arguments existing at the time, before moving on with implicits). These relations were selected in a balanced way, in order to mirror the distribution of top-level classes in the sense hierarchy over the entire corpus. The 17 documents were annotated by an annotator with extensive experience in annotating PDTB relations. For this annotation task, we reached substantial agreement with Cohen's Kappa at 0.74. Zeyrek et al. (2019) report a score of 0.71 for German, but do not differentiate between explicit and implicit relations. Since the PDTB reports percentages, we calculated percentage-wise agreement, resulting in 70.4%, compared to 80%[3] for the sense annotation task in the PDTB (Prasad et al., 2008) (Table 3). We manually checked instances of disagreement in our annotations to find a cause for this difference, but since almost all cases of disagreement occurred only once or twice, no clear picture emerged.

For the annotation of new relation types, a slightly larger set of 20 documents was selected, but in an unbalanced way (randomly picked instead). This sub-set was annotated by one additional annotator experienced in annotating PDTB relations. For this subset and task, a considerably lower agreement was reached, with Cohen's Kappa at .28 for types (choosing between implicit, AltLex, EntRel and NoRel) and 0.30 for senses (for the overlapping implicit and AltLex annotations). Zeyrek et al. (2019) report a considerably higher type agreement (0.78) for German in their corpus. Prasad et al. (2008) do not report agreement on this subset of types (and consequently the senses relating to this subset only). Upon manual investigation, we found that the most frequent case of type disagreement was between implicit and EntRel cases. A typical example is provided by (1).

[2]In the actual XML, in line with PCC vocabulary, we refer to the latter two as ext_arg and int_arg, respectively.

[3]Note that this number is for explicit and implicit relations combined.

1. Mit Helga Kaden streicht eine der namhaftesten Geschäftsleute der Stadt die Segel. Sie konnte ihr traditionsreiches Geschäft wegen der anhaltenden Kundenflaute nicht mehr über Wasser halten.
   With Helga Kaden calling it a day, one of the most well-known entrepreneurs of the city is gone. Due to a diminishing customer base, she could not keep her business aloft.

Where one annotator interpreted an Expansion.Level-of-detail.Arg2-as-detail relation, where Arg2 (which in turn contains an explicit relation, giving the reason for the bankruptcy) provides more detail, the other annotator interpreted an EntRel relation on the basis of the two sentences being about the same entity (*Helga Kaden* in Arg1, referred to by the pronoun *Sie* (*she*) in Arg2. Another frequent case of disagreement with regard to relation type was between implicit and AltLex, where one annotator consistently interpreted a semi-colon as alternative lexicalisation signal, whereas the other annotator did not. Our annotation guidelines did not explicitly instruct annotators on punctuation marks. Which surface signals should be included in the group of discourse markers is a matter of ongoing debate (for a discussion of such signals from an RST perspective, see Das and Taboada (2018)), and we consider unifying the corpus with regard to this a piece of future work.
Regarding sense disagreement, the most frequent case was Expansion.Level-of-detail.Arg2-as-detail vs. Contingency.Cause.Reason, exemplified by (2).

2. Alle Jahre wieder Stress vor dem Fest. Am Sonnabend hetzten nicht wenige noch von Geschäft zu Geschäft, um die letzten Einkäufe zu erledigen.
   The same pre-christmas stress every year. On Saturday many ran from store to store and did last-minute shopping.

Where one annotator interpreted Arg2 as providing more detail for the statement in Arg1 (implicit connective: *genauer gesagt* (*more specifically*)), the other annotator interpreted the clarification in Arg2 as the reason provided for the statement in Arg1 (implicit connective: *denn* (*because/for*)). For an overview of how different senses are distributed over the semantic space (within and across frameworks), we refer to Sanders et al. (2018), and leave further unification of these cases in the PCC 2.2 to future work.

4.2. Corpus Statistics

The distribution of relations in the PCC 2.2 is illustrated in Table 1. The decrease in explicit relations [4] is due to the revision of this layer.
The most relevant corpus to compare our results to is the PDTB 2.0, and Figure 1 illustrates the distribution of the PCC 2.2 alongside the PDTB 2.0.
In terms of relation types, the PCC 2.2 roughly follows the distribution of the PDTB 2.0, with the most striking difference being the smaller number of EntRel relations (2.5% in the PCC vs. 12.8% in the PDTB). Furthermore it has more explicit relations (50.2% vs. 45.5%) and AltLex relations (5.5% vs. 1.5%).

[4]The previous version contained 1,110 explicit relations.

	PCC 2.2
AltLex	122
EntRel	56
Explicit	1,108
Implicit	887
NoRel	35
Total	2,208

Table 1: Distribution of relations in the PCC 2.2

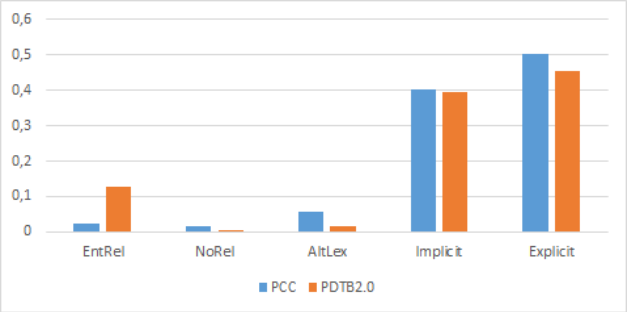

Figure 1: Relative distribution of relations in the PCC 2.2 and PDTB 2.0

This most prominent difference between the distribution of EntRels could be down to genre, or perhaps rather format differences; the average length of one document in the PCC is ~189 words, whereas the average PDTB document has ~455 words. The authors featuring in the PCC may have felt the need to be concise, and thus chose to use fewer EntRel constructions; Knott et al. (2000) discuss the status of entity-based coherence relations as relation instances not compatible with the sense taxonomy and tree-structure, but they do so in the context of Rhetorical Structure Theory, and more research, and above all corpus-based comparison (including other genres and varying text lengths) would be needed to verify the assumption that shorter text lengths lead to less EntRel constructions. Interestingly, this discrepancy between the PCC and the PDTB directly relates to the most frequent case of type disagreement; that between EntRel and implicit (as discussed in Section4.1.). Furthermore, we note that in total, the PCC only has 56 EntRel instances, rendering the EntRel and especially NoRel (with only 35 instances) comparison to its corresponding PDTB amount less significant.
Looking at the four top-level senses[5] over all relations (Figure 2), the PCC resembles the PDTB, with a slightly larger difference for *Temporal* class relations (7.0% in the PCC vs.13.1% in the PDTB).
Because the PDTB features news articles, typically explicitly mentioning when said events took place, we assume that temporal relations between these events are more likely to be expressed in the text. Conversely, with the more argumentative nature of the PCC articles, temporal relations can be considered less relevant, which could explain the lower frequency of *Temporal* class relations in the PCC.

[5]We only use the top-level senses in all figures, because including all lower level specifications would mean including many very low-frequent senses or even singletons.

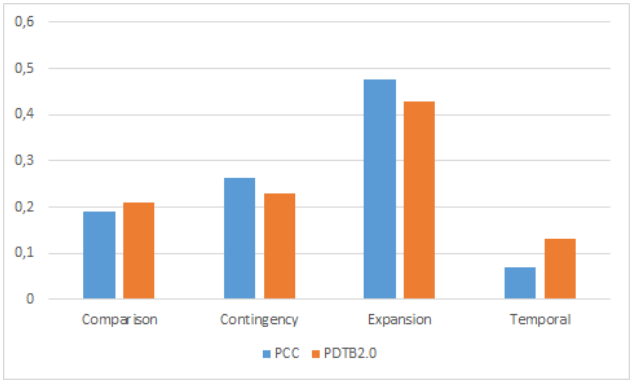

Figure 2: Relative distribution of top-level class senses in the PCC 2.2 and PDTB 2.0

Zooming in on individual relation types (explicits, implicits and AltLex cases[6]), again when considering the four top-level senses, the PCC roughly matches the PDTB for explicits and implicits (Figures 3 and 4), and considerably deviates for AltLex cases (Figure 5). In the latter, especially relations of the *Contingency* class are much less frequent (10.7% vs. 44.3%), mostly at the expense of *Temporal* class relations (29.5% vs. 13.8%). A more detailed error analysis led to no obvious reason for this discrepancy, though we note once more that the AltLex comparison is based on only 122 instances on the PCC side, rendering it less significant than the explicit (1,120 in the PCC) and implicit (887 in the PCC) comparisons.

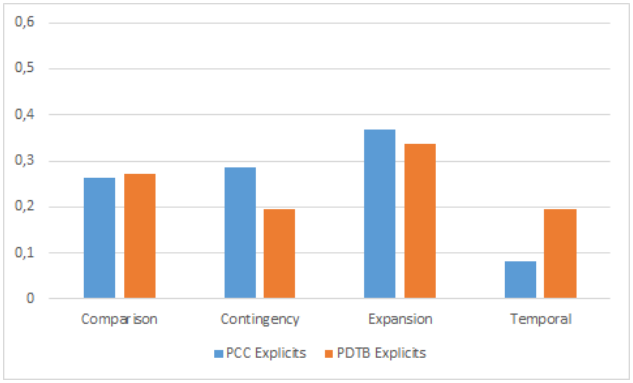

Figure 3: Relative sense distribution of explicit relations in the PCC 2.2 and PDTB 2.0

5. Conclusion

We introduce the Potsdam Commentary Corpus in its 2.2 version, a German corpus of news editorials annotated on several different levels. As its previous versions, it is freely available.[7] The novelties of the 2.2 version are the addition of relation senses for an already existing layer of connectives and arguments, and an all-new layer of implicit relations (including AltLex, EntRel and NoRel instances), following the PDTB annotation scheme. In addition, the new

[6]EntRel and NoRel instances by definition have no sense assigned to them.

[7]`http://angcl.ling.uni-potsdam.de/resources/pcc.html`

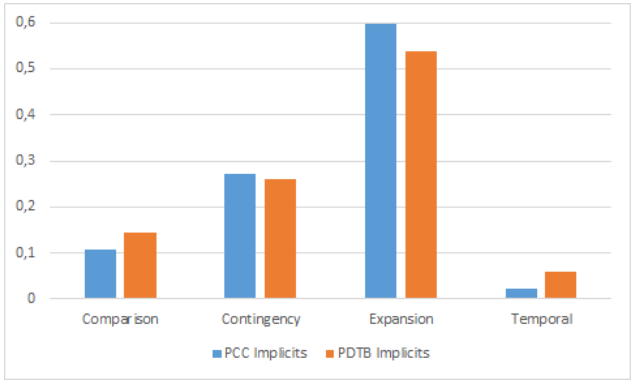

Figure 4: Relative sense distribution of implicit relations in the PCC 2.2 and PDTB 2.0

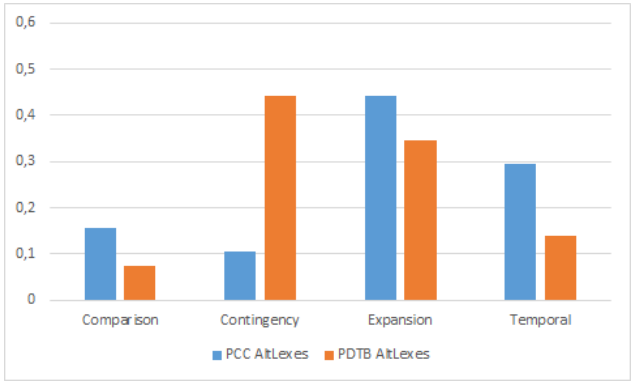

Figure 5: Relative sense distribution of AltLex relations in the PCC 2.2 and PDTB 2.0

version comes with minor modifications to already existing connectives and arguments relations. In its current shape, the PCC 2.2 mimicks the PDTB annotations in its 2.0 version, missing out only on intra-sentential implicit relations that were added in the PDTB 3.0 iteration. The additional annotation layers are published in standoff format, for more convenient automatic processing of the corpus.

We provide inter-annotator agreement numbers for the newly added annotations and compare statistics of the resulting PCC 2.2 corpus to the PDTB 2.0.

The main goal of further developing the corpus is to increase its usability for the task of shallow discourse parsing, and our main focus regarding future work will be on exploiting the new annotations as training data for a German shallow discourse parser, for which some individual components have already been developed (Bourgonje and Stede, 2018a; Bourgonje and Stede, 2019).

With regard to the corpus itself, we consider it an important future improvement to include intra-sentential implicit relations, in line with the PDTB evolution, which saw this addition going from its 2.0 version to its 3.0 version. Finally, manual investigation of disagreement cases revealed issues in type and sense assignment. For this, the literature will have to be consulted before an unambiguous instruction can be presented to the annotators in order to resolve the relevant cases.

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 323949969. We would like to thank the anonymous reviewers for their helpful comments on an earlier version of this paper.

6. Bibliographical References

Bourgonje, P. and Stede, M. (2018a). Identifying Explicit Discourse Connectives in German. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, Melbourne, Australia.

Bourgonje, P. and Stede, M. (2018b). The Potsdam Commentary Corpus 2.1 in ANNIS3. In *Proceedings of the 17th International Workshop on Treebanks and Linguistic Theory*, Oslo, Norway.

Bourgonje, P. and Stede, M. (2019). Explicit Discourse Argument Extraction for German. In *Proceedings of the 21st International Conference on Text, Speech and Dialogue*, Ljubljana, Slovenia.

Das, D. and Taboada, M. (2018). RST Signalling Corpus: a corpus of signals of coherence relations. *Language Resources and Evaluation*, 52(1):149–184.

Dipper, S. and Stede, M. (2006). Disambiguating potential connectives. In *Proceedings of the KONVENS Conference*, Konstanz.

Hernault, H., Prendinger, H., duVerle, D. A., and Ishizuka, M. (2010). HILDA: A Discourse Parser Using Support Vector Machine Classification. *D&D*, 1(3):1–33.

Joty, S., Carenini, G., and Ng, R. T. (2015). CODRA: A Novel Discriminative Framework for Rhetorical Analysis. *Computational Linguistics*, 41(3):385–435, September.

Knott, A., Oberlander, J., O'Donnell, M., Mellish, C., and Mellish, E. M. O. C. (2000). Beyond elaboration: The interaction of relations and focus in coherent text. In *Text Representation: Linguistic and Psycholinguistic Aspects, chapter 7*, pages 181–196. John Benjamins.

Krause, T. and Zeldes, A. (2016). ANNIS3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities*, 31.

Lee, A., Prasad, R., Webber, B., and Joshi, A. K. (2016). Annotating discourse relations with the PDTB annotator. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, pages 121–125, Osaka, Japan, December. The COLING 2016 Organizing Committee.

Li, Z., Zhao, H., Pang, C., Wang, L., and Wang, H. (2016). A Constituent Syntactic Parse Tree Based Discourse Parser. In *Proceedings of the CONLL 2016 Shared Task*, pages 60–64, Berlin.

Lin, Z., Ng, H. T., and Kan, M.-Y. (2014). A PDTB-Styled End-to-End Discourse Parser. *Natural Language Engineering*, 20:151–184.

Mann, W. and Thompson, S. (1988). Rhetorical structure theory: Towards a functional theory of text organization. *TEXT*, 8:243–281.

Oepen, S., Read, J., Scheffler, T., Sidarenka, U., Stede, M., Velldal, E., and Øvrelid, L. (2016). OPT: Oslo–Potsdam–Teesside—Pipelining Rules, Rankers, and Classifier Ensembles for Shallow Discourse Parsing. In *Proceedings of the CONLL 2016 Shared Task*, Berlin.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., and Webber, B. (2008). The Penn Discourse Treebank 2.0. In *Proc. of the 6th International Conference on Language Resources and Evaluation (LREC)*, Marrakech, Morocco.

Prasad, R., Webber, B., Lee, A., and Joshi, A. (2019). Penn Discourse Treebank Version 3.0. Philadelphia: Linguistic Data Consortium.

Rysová, M., Jínová, P., Mírovský, J., Hajičová, E., Nedoluzhko, A., Ocelák, R., Pergler, J., Poláková, L., Zdeňková, J., Scheller, V., and Zikánová, Š. (2016). Prague discourse treebank 2.0.

Sanders, T., Demberg, V., Hoek, J., Scholman, M., Asr, F., Zufferey, S., and Evers-Vermeul, J. (2018). Unifying dimensions in coherence relations: How various annotation frameworks are related. *Corpus Linguistics and Linguistic Theory*, 05.

Soricut, R. and Marcu, D. (2003). Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*.

Stede, M. and Heintze, S. (2004). Machine-assisted rhetorical structure annotation. In *Proceedings of the 20th International Conference on Computational Linguistics*, COLING '04, Stroudsburg, PA, USA. Association for Computational Linguistics.

Stede, M. and Neumann, A. (2014). Potsdam Commentary Corpus 2.0: Annotation for discourse research. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, pages 925–929, Reikjavik.

Stede, M. (2004). Kontrast im Diskurs. In H. Blühdorn, et al., editors, *Brücken schlagen. Grundlagen der Konnektorensemantik*, pages 255–286. Walter de Gruyter, Berlin.

Wang, J. and Lan, M. (2015). A Refined End-to-End Discourse Parser. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning - Shared Task*, pages 17–24. Association for Computational Linguistics.

Zeldes, A., Das, D., Maziero, E. G., Antonio, J., and Iruskieta, M. (2019). The DISRPT 2019 Shared Task on Elementary Discourse Unit Segmentation and Connective Detection. In *Proceedings of the Workshop on Discourse Relation Parsing and Treebanking 2019*, pages 97–104, Minneapolis, MN, June. Association for Computational Linguistics.

Zeyrek, D., Mendes, A., Grishina, Y., Kurfalı, M., Gibbon, S., and Ogrodniczuk, M. (2019). TED Multilingual Discourse Bank (TED-MDB): a parallel corpus annotated in the PDTB style. *Language Resources and Evaluation*, 04.