

Information structure in the Potsdam Commentary Corpus: Topics

Manfred Stede and Sara Mamprin

Applied Computational Linguistics / FSP Cognitive Science

Universität Potsdam (Germany)

stede@uni-potsdam.de, mamprin@uni-potsdam.de

Abstract

The Potsdam Commentary Corpus is a collection of 175 German newspaper commentaries annotated on a variety of different layers. This paper introduces a new layer that covers the linguistic notion of information-structural *topic* (not to be confused with ‘topic’ as applied to documents in information retrieval). To our knowledge, this is the first larger topic-annotated resource for German (and one of the first for any language). We describe the annotation guidelines and the annotation process, and the results of an inter-annotator agreement study, which compare favourably to the related work. The annotated corpus is freely available for research.

Keywords: Information Structure, Annotation, Multi-layer text corpus

1. Introduction

The linguistic notion of *information structure* (henceforth: IS) was conceived to explain certain phenomena of relative prominence in linguistic utterances, both in prosody (intonation, phrasing) and in syntax (word order, embedding). It also can have effects on truth-conditional semantics. While a number of competing theories and terminologies persist in the literature, there seems to be a minimal consensus that there are three central notions to be distinguished. Below we illustrate them with examples taken from (Krifka and Musan, 2012), who provide a thorough overview of the matter. (Upper case letters denote stress on the syllable.)

- The *information status* of discourse referents (often *given* versus *new*; sometimes using much more detailed scales)

(1) Ten years after John inherited an old farm, he
SOLD [the shed]_{given}

To mark the givenness in speech, *the shed* needs to be deaccentuated. (Otherwise, the shed would be understood as just part of the farm, which renders it non-given.)

- The *focus/background* partitioning of utterances to reflect prominence. This is typically illustrated with a question representing prior context.

(2) A: What did John show Mary?
B: John showed Mary [the PICTures]_F.
A: Who did John show the pictures?
B: John showed [MAry]_F the pictures.

- The *topic* as the entity that an utterance or a sentence “is about”, often in a dichotomy with *comment*, i.e., the state of affairs that is being predicated of the topic.

(3) (a) [Aristotle Onassis]_{topic} married Jacqueline Kennedy.
(b) [Jacqueline Kennedy]_{topic} married Aristotle Onassis.

Another way of framing these issues is the notion of “information packaging” (Chafe, 1976), which emphasizes the cognitive motivations for making the variety of choices that a speaker faces when formulating linguistic units – it should be done in such a way that the utterance is easy to process by the addressee, who seeks to mentally reconstruct the message conveyed by the speaker, and its underlying intention.

Arguably, at least some aspects of information structure are most fruitfully applied to spoken language, since in many languages prosody is heavily influenced by it, and certain morphosyntactic markers of information-structural aspects occur more often in spoken than in written language. Nonetheless, especially information status and topic are also relevant for producing and understanding written text, which will be our concern in this paper.

Theoretical linguistics has traditionally invoked laboratory examples to explain IS phenomena, which is not unreasonable for a start, because in authentic discourse, IS is but one of a variety of factors that collectively contribute to the form of utterances, thus making it hard to examine the precise role of IS. “Crisp” examples like (1-3) above facilitate delimiting the phenomena in question. Ultimately, though, the notions need to be robust enough so that they can be analyzed also in natural speech and text. Our particular concern here is with *topic* and its application to written text: The goal is, roughly speaking, to identify a linguistically-motivated topic in every sentence. So far, our project uses only German text, but the approach will be applicable to typologically-similar languages (certainly to English).

In terms of computational application, the primary relevance of topic annotation is for coreference analysis. Topics are locally-prominent entities, and therefore they concern any work revolving around centering theory (Grosz et al., 1995) or entity-based coherence (Barzilay and Lapata, 2008). This research has usually equated prominence with subjecthood, which works to a large extent for English, but not so well for other languages; see, e.g., (Strube and Hahn, 1996). – Other than that, there is a potential connection between sentence topics (in the IS sense) and the (more intuitive) topics of larger stretches of text or of complete documents. Thus, research on breaking down texts into a

sequence (or hierarchy) of topic-homogeneous units, which had started with (Hearst, 1994), could benefit from a thorough analysis of topics on the sentence level.

In the following, Section 2. discusses related work on topic annotation in written language. Then, Section 3. provides some background on the data we used for our annotation project, and Section 4. explains our annotation guidelines and the result of an IAA study. Section 5. concludes and sketches directions for future work.

2. Related work

Among the three subfields of IS mentioned at the beginning of the paper, information status has received most attention in computational linguistics. Also, the spoken dialogue community has addressed topic and focus to a good extent. (For larger overviews, see (Kruijff-Korbayová and Steedman, 2003) or (Stede, 2012)).

Concerning topics in written language, probably the most developed annotation effort on written language is found in the Prague Dependency Treebank 2.0¹, a multi-layer corpus of Czech text, which has also been extended to Arabic and English. The PDT was developed in accordance with the theory of the Prague School of functional and structural linguistics (Sgall et al., 1986). As Czech is a free-word-order language, information structure plays a prominent role in the approach, and the central notion for our purposes here is the topic-focus articulation (TFA). Topics have been annotated on top of the syntactic dependency structures, which means that they can easily be queried and analyzed in their syntactic context, bearing in mind that their selection was constrained by the syntactic considerations (which is in contrast to our own project). In particular, (Buráňová et al., 2000) proposed explicit rules for identifying topic and focus on dependency trees.

One of the first corpus-based studies of IS in German was done by (Ritz et al., 2008), who had information status, focus and topic annotated in a variety of genres. Topic was assigned within sentences, with the rule that for complex sentences, first the matrix clause was to be handled and then every finite clause, with the exception of restrictive relative clauses, which did not receive a topic. The markables (units that can be annotated) were various types of NPs, which the guidelines (Götze et al., 2007) differentiated in terms of their syntactic and semantic properties. For deciding on a topic candidate X, annotators were to use diagnostics: Can the phrase “Let me tell you about X” be preposed; is the sentence an answer to the question “What about X?”; can the sentence be paraphrased as “Concerning X, ...”. On texts from the newspaper commentary corpus that we are also using in our work presented here, the annotator agreement for topic was $\kappa=0.44$.

More recently, (Cook and Bildhauer, 2013) used a revised version of the Götze et al. guidelines and conducted a new study on newspaper texts. Four annotators labeled 56 sentences and achieved a Fleiss- κ of 0.447 on topic assignment – essentially the same value as that of the Ritz et al. study. For deciding whether a sentence had a topic or was topic-less (thetic; see Section 4.2.), the agreement was only

$\kappa=0.225$. Cook and Bildhauer provided a very informative error analysis of problematic cases.

3. Our corpus: newspaper commentary

The Potsdam Commentary Corpus 2.0 (Stede and Neumann, 2014) is a collection of 175 newspaper commentaries from a regional German newspaper. It was deliberately collected as an “unbalanced” single-genre corpus, so that research questions on subjectivity and argumentation can be addressed. The PCC has been annotated with sentential syntax, nominal coreference, discourse connectives in the spirit of the Penn Discourse Treebank (Prasad et al., 2008), and rhetorical structure in terms of RST (Mann and Thompson, 1988). These annotations and the primary data have been made publicly available. Various other annotation layers have been developed for parts of the corpus but are not yet released; all annotation guidelines are published in (Stede, 2016).

The layers are being annotated largely independently (i.e., directly on the source text), the exceptions being that (i) a layer of nominal referring expressions is the basis for both coreference and topic annotation, and (ii) a layer of discourse segments is the basis for rhetorical structure, argumentation structure, and the domains for topics, as explained in the next section.

The topic annotation of the 175 texts along with a few corrections on the earlier annotation layers is available from our website² as version 2.1 of the corpus.

4. Topic annotation

As pointed out above, linguists have mostly applied *topic* to laboratory examples, which usually are relatively simple sentences. When being confronted with authentic text, however, it needs to be clarified what sort of linguistic unit can be assigned a topic. We discuss this below as the segmentation task, and then explain the notion of topic defined in our guidelines, the annotation process, and the results of an agreement study. Our topic guidelines were originally inspired by those of (Götze et al., 2007), but eventually diverged quite a bit from them. The complete document (in German), including also our segmentation guidelines can be found in (Stede, 2016).

4.1. Segmentation

In the PCC, segmentation is a two-staged process. First, an application-neutral *discourse segmentation* employs largely structural criteria to determine segment boundaries and assign syntactic labels to the segments. In this step, there are two sources of complication:

- **Complex sentences** have to be broken into smaller units, which primarily involves identifying different kinds of subordinate clauses. We used the inventory proposed by (Bußmann, 2002) with small modifications, now distinguishing nine types of such clauses, plus an ‘unknown’ category for cases where the annotator is unsure. Full main (and coordinate) clauses

¹<https://ufal.mff.cuni.cz/pdt2.0/>

²<http://angcl.ling.uni-potsdam.de/resources/pcc.html>

receive a corresponding label, as do main clause fragments where some material has been elided. For example:

(4) [Mary left the building.]_{sent} [And lit herself a cigarette.]_{sent-frag}

- **Fragments** are defined as less syntactically complete than main clause fragments; they can introduce sentences or provide add-on information. Annotators are asked to attach them to the right or left neighbouring unit and assign a corresponding type.

(5) [Clear is:]_{frag-init}[The situation is bad.]_{sent}

(6) [The situation is bad.]_{sent}[Very much so.]_{frag-end}

For this step of application-neutral, hierarchical segmentation, we have also built an automatic system that builds discourse segments for German text on the basis of the output of a dependency parser; it is documented in (Sidarenka et al., 2015).

Then, the second stage of the segmentation process maps the generic units from stage 1 to annotation-layer-specific units, such as the 'elementary discourse units' used in RST annotation. Among the stage 1 types, a subset is identified that forms independent units for the target layer; all the units of other types are being added to their embedding units. In this way, for example, non-restrictive relative clauses can be either treated as independent units for a layer of discourse annotation, or they can be kept as part of their host clauses for an argumentation analysis.

For topics, the literature does not offer much insight on these issues, or, as (Matić et al., 2014) put it, it is an "understudied phenomenon". A general assumption is that material in subordinate clauses is presupposed whereas that in main clauses is asserted, but as for instance (Erteschik-Shir, 2007) has shown, this is often too much of a simplification. Matic et al., in their analysis of information structure in complex sentences, distinguish two types of subordination and corresponding kinds of clauses, which we also do in our guidelines:

- *d*-subordination is a predicate-argument relationship that forms one complex proposition:

(7) [Mary thought that Sally would arrive late.]

- *ad*-subordination links an adjunct to the superordinate clause, which corresponds to two separate propositions:

(8) [Mary was sad][because Sally arrived late.]

Further, Erteschik-Shir distinguishes between a clause-internal information structure (in particular, topic) and an external IS that describes the informational contribution made by the subordinate clause to the host clause. Our annotation scheme is generally compatible with that proposal; we regard main clauses, main clause fragments, adverbial subordinate clauses and so-called *weiterführende Nebensätze* ("topic-shifting clauses") as those units that are to receive a topic annotation.

4.2. Defining 'topic'

We indicated earlier that there is no generally-accepted definition of 'topic'; different views result from slightly different explanatory goals, from considerations of an underlying linguistic theory, and from different languages or modes (spoken, written) being studied. A useful analysis was provided by (Jacobs, 2001), who largely followed the conceptions of topic by Strawson and Reinhart, and identified three underlying features of topics (which then can be more or less prominent in different definitions):

1. Informational separation: In the linguistic form, the topic is being presented as separate from the remaining material in the clause;
2. Predication: the remaining material can be construed as a semantic predicate that is applied to (the referent of) the topic;
3. Addressation: the information given in the remaining material is (mentally) stored under the 'address' of the topic.

For operationalizing such concepts in the form of annotation guidelines for authentic text, (3) corresponds to the notion of 'aboutness' that is elsewhere often stated as the main feature of topics, and (1) can be seen as an additional criterion for identifying topics. (2), on the other hand, is much less amenable to an annotation process, because it is hardly feasible to have annotators first mentally construct semantic representations of the material before making annotation decisions.

Similar to the (Cook and Bildhauer, 2013), we therefore decided to make 'aboutness' the central feature of 'topic', to explain it with the help of Jacobs' criterion (3), and to use (1) as an additional guideline for the annotators. Furthermore, we give three linguistic tests, which have also been used in related work, to be employed as heuristics:

- "What about X" question: If the unit can be read as an answer to this question, then constituent X is a good candidate for the topic. For example (3a) above, the question is "What about Aristotle Onassis?"
- "Concerning X" paraphrase: If the unit (in its context) can be paraphrased as "Concerning X, ...", then constituent X is a good candidate for the topic. Applied to (3a): "Concerning Aristotle Onassis, he married Jacqueline Kennedy."
- Left-dislocation: If the unit can be paraphrased with X being in left-dislocated position, then X is a good candidate for the topic. (This test often works better for German than it does for English.) Applied to (3a): "Aristotle Onassis, he married Jacqueline Kennedy."

Notice that the 'concerning'-test and the left-dislocation test reflect Jacobs' criterion (1): It should be possible to present the topic separate from the remaining material (the 'comment').

In simple sentences like our examples given so far, it is not difficult to identify the topic. But in authentic text, matters can become more complicated. Here is a sentence taken

from one of our corpus texts; we mark the aboutness topic with an *ab* subscript:

- (9) (Context: Author argues in favor of mandatory vaccination for children, to fight various diseases)
Nur wer impfen lässt, hilft mit, dass [Impfungen]_{ab} eines Tages überflüssig werden.
(‘Only those who have {their kids} vaccinated contribute to [vaccinations]_{ab} becoming superfluous some day.’)

Notice that ‘the kids’ are elided in the German sentence, and the subject is the impersonal *wer* (‘who’), which is not properly referring and cannot be a topic. The sentence provides information about the concept of vaccination, as can be verified with the “what about” and the “concerning” test, whereas left-dislocation does not work for such cases where the topic is deeply embedded in the sentence. Different syntactic configurations can lead to different perceptions of ‘unnaturalness’ in the test sentences, which is why there is not one single, and decisive, test. Choosing the topic can be a matter of weighing various candidates against each other, looking for the one where the tests work best.

As a further complication, not every candidate unit (i.e., a clause or sentence that was determined to be a topic “domain” in the segmentation phase) needs to have an aboutness topic. So-called *thetic* sentences, sometimes characterized as “all-new” (alluding to the stipulation that topics ought to have the information status *given* – which we however decided not to use as an explicit criterion in our guidelines) start a text or a new part of a text, often a paragraph. Metaphorically speaking they place all their referents on an “empty stage”. A test is to check whether the sentence more naturally answers the question “What’s new?” rather than a “What about X?” question. In this way, annotators need to decide for any relevant unit whether it is *thetic*, or should have an aboutness topic.

In addition to the “standard” aboutness topics, our guidelines use two further variants of topics, which are sometimes discussed in the literature.

- *Frame-setting* topics (e.g., (Krifka and Musan, 2012) provide a (temporal, local, conceptual) context for the predication and thereby restrict its validity. They typically co-occur with aboutness topics in the clause. We mark them with an *fs* subscript.

- (10) [In Berlin,]_{fs} [many refugees]_{ab} find shelter in sports halls.

- *Contrastive* topics (e.g., (Roberts, 1996)) are relatively rare: The text first mentions some general term, and subsequent spans of discourse then address more specific variants (often by homonymy). We use *co>n* for contrastive topics, where *n* is an index attached to the preceding general phrase.

- (11) Berlin uses different kinds of [buildings]₁ for hosting refugees. [Sports halls]_{co>1} have been temporarily converted and equipped with beds; [empty warehouses]_{co>1} have been cleaned up and stocked with the necessary equipment.

4.3. Annotation process

Our guideline document is 20 pages long. The second stage of the segmentation process as described above (i.e., the selection of those discourse segments that constitute a topic domain TD) is a part of it, and lays the foundation for the topic annotation proper. Annotators are asked to proceed left-to right through the text. They consider each candidate TD and go through the following procedure (here slightly simplified for presentation purposes):

1. Is there a fragment unit adjacent to the TD? If so, check whether for topic assignment it should be integrated with the TD.
2. Is the TD *thetic*, i.e., an “all new” statement? If yes, mark it as such, and stop. Otherwise, proceed with (3).
3. Apply the aboutness test to all candidate referring expressions in the TD. If there is one such candidate, mark it, and proceed with (5). If there is no candidate, re-consider step (2). If there are multiple candidates, proceed with (4).
4. Decide on a ranking of the candidates and mark them as *ab1*, *ab2*, and so forth. Proceed with (5).
5. Check whether any frame-setting topic should be marked. Then proceed with (6).
6. Check whether there is any configuration of contrastive topics involving this TD, and if so, mark the expressions as such.

As stated in steps 3 and 4, if annotators find it difficult to make a decision among several candidate topics, they have the option to mark more than one, but in this case they have to provide a ranking. (The guidelines do not state a maximum, but in practice there are hardly ever more than two candidates.) The annotation tool we use at the moment is Exmaralda³, but the guidelines are neutral with respect to the tool.

4.4. Agreement study and PCC annotation

4.4.1. Agreement study

We evaluated the discourse segmentation step independently of the topic annotation, which was carried out after creating a gold standard of segments. The data for the IAA study consisted of 10 texts, which amount to 138 discourse segments (in the gold standard). In our evaluation, the segment annotations were produced by two researchers, one being one of the authors of this paper, and the other a graduate student trained with the guidelines. Using the *parseeval* measure (Black et al., 1991), we achieved a labeled F1 score of 89.13 for the hierarchical segmentation with assignment of syntactic types. More details, including a discussion of various other agreement measures, can be found in (Sidarenka et al., 2015).

In the subsequent topic assignment study, two annotators worked on the gold-segmented 10 texts (again one author of this paper and a trained graduate student). There are

³<http://www.exmaralda.org>

different ways to calculate agreement for this type of annotation. To facilitate comparison with the related work of (Cook and Bildhauer, 2013), mentioned at the end of Section 2., we followed their procedure as far as possible.

For theticity, we therefore treated every segment as a data point receiving a non-/thetic label. We have 139 segments, and Cohen κ for the two annotators is 0.60. This compares favourably to the 0.225 reported by Cook and Bildhauer, but their figure is a four-rater Fleiss- κ , and all of the annotators were trained students. Another caveat is segment size; we do not know precisely how they handled complex sentences. Finally, notice that in our case the number of annotated segments is much larger (139 versus 58).

Likewise for aboutness, we treated every referring NP or PP in our texts as a data point receiving a non-/topic label. In the interest of comparability, we ignored our annotations of frame-setting and contrastive topics. Our two-rater κ is 0.71, also considerably higher than the 0.447 given by Cook/Bildhauer; the same provisos apply, though. Here, our number of items is considerably smaller (342 versus 516 markables), which indicates that our study uses smaller segments, which probably makes topic decisions easier (on average, in our data, there are 2.45 markables per segment). In addition, to compare such figures fairly, one would need to investigate the rules used for defining markables: In our study, we did not treat NPs appearing in figurative language as markables, because they do not refer properly. In addition, the handling of embedded NPs/PPs would need to be compared.

4.4.2. Corpus annotation: Experiences with problematic instances

Having obtained the encouraging agreement results, we proceeded to annotate the full corpus of 175 texts. One author created annotations, the other checked them, and cases of disagreement were adjudicated. The outcome of our second stage of segmentation yielded about 2800 topic domain units in the corpus; 15% of these were marked as thetic, the others received topic annotation.

In the earlier stages of this annotation project, the majority of problematic cases concerned topic candidates in various types of embedded clauses. This had prompted us to split off the segmentation task and settle the question of domains before starting the topic labeling. This step thus contributed a lot to the efficiency of the overall annotation process; it might be the case, though, that our criteria for selecting possible topic domains need to be refined; this is a question we leave to future work.

The other central source of complication in this annotation project is theticity. A notorious problem is in judging whether a particular discourse referent, which has not been mentioned in the text before, is to be taken as “hearer-old” (Prince, 1992) qua world knowledge and thus makes a good candidate for an aboutness topic. This issue arises in many reference-related annotation projects, and we don’t have a good solution; our guidelines merely appeal to the annotator’s intuition for deciding whether some entity should be “generally known” or not.

Finally, we noticed that when a unit has multiple topic candidates, their relative “weight” can play a role for anno-

tators. Referring expressions like possessive pronouns or pronominal adverbs (e.g., German *damit* = ‘with that’) have additional functions besides reference, and seem “lighter” than full NPs or personal pronouns. In our guidelines we strive to not give purely structural criteria for choosing topics (because we want to correlate our topics with syntactic features *after* the annotation). Another factor that plays a role here is that we ask annotators to consider the previous context when judging a segment, but not to look forward; thus, the “persistence” of a discourse referent in the subsequent text should not affect the decision on topicality. We are generally happy with this rule, but it seems to contribute to the difficulty of judging those “light” referring expressions.

5. Conclusion and outlook

Sentence topics, from the linguistic perspective, play a central role for explaining the management of the common ground in discourse. For computational linguistics, this is clearly relevant for dialogue processing, but also – as argued in this paper – for understanding monologue text. In particular, topic-related questions have been discussed, modeled and partially implemented under the heading of “entity-based coherence”. Text resources with topic annotation can inform those models and contribute to a better understanding of reference-tracking mechanisms, since topics are good candidates for being antecedents of anaphors, and more generally, for many languages they provide a better account of referential prominence than grammatical functions (subjects) do.

We presented the extension of a German corpus with topic annotation, where they supplement a number of other annotation layers: sentence syntax, coreference, PDTB-style connectives, and discourse structure (in terms of RST). The challenge for topic annotation guidelines is to strike a balance between being attentive to linguistic theory on the one hand, and operationalizing the notion of topic for “annotatability” of authentic text. To this end, we focus on the ‘aboutness’ feature and use a characterization of ‘topic’ that combines various linguistic features and diagnostic tests from the literature, but also leaves some features out, such as *givenness*, which in our opinion belongs onto a different layer of annotation and is in principle orthogonal to aboutness (even though very often, aboutness topics turn out to be given).

Our corpus and the guidelines are available for research, and we hope to contribute to stimulating more research on the computation of information structure. Next steps for our work are to explore the correlations of the topic layer with sentence syntax (what grammatical functions correspond to topics, and under what circumstances), and with coreference (what role do topics play for coreference chains).

Acknowledgements

The research reported in this paper was funded by Deutsche Forschungsgemeinschaft as part of the Collaborative Research Center 632 *Information Structure* at Universität Potsdam and Humboldt Universität zu Berlin.

6. References

- Barzilay, R. and Lapata, M. (2008). Modeling local coherence: An entity-based approach. *Computational Linguistics*, 34(1):1–34.
- Black, E., Abney, S. P., Flickinger, D., Gdaniec, C., Grishman, R., Harrison, P., Hindle, D., Ingria, R., Jelinek, F., Klavans, J. L., Liberman, M., Marcus, M. P., Roukos, S., Santorini, B., and Strzalkowski, T. (1991). A procedure for quantitatively comparing the syntactic coverage of English grammars. In *Proceedings of the Workshop on Speech and Natural Language*, Pacific Grove/CA.
- Buráňová, E., Hajičová, E., and Sgall, P. (2000). Tagging of very large corpora: topic-focus articulation. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING)*, pages 139–144, Saarbrücken.
- Bußmann, H. (2002). *Lexikon der Sprachwissenschaft*. Kröner, Stuttgart.
- Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Li, C. N., editor, *Subject and Topic*, pages 27–55. Academic Press, New York.
- Cook, P. and Bildhauer, F. (2013). Annotating information structure. The case of "topic". *Dialogue and Discourse*, 4(2):118–141.
- Erteschik-Shir, N. (2007). *Information structure: The syntax-discourse interface*. Oxford University Press, Oxford.
- Götze, M., Endriss, C., Hinterwimmer, S., Fiedler, I., Petrova, S., Schwarz, A., Skopeteas, S., Stoel, R., and Weskott, T. (2007). Information structure. In *Information structure in cross-linguistic corpora: annotation guidelines for morphology, syntax, semantics, and information structure*, volume 7 of *ISIS Working papers of the SFB 632*, pages 145–187.
- Grosz, B., Joshi, A., and Weinstein, S. (1995). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 21(2):203–226.
- Hearst, M. A. (1994). Multi-paragraph segmentation of expository text. In *Proceedings of the 32nd Meeting of the Association for Computational Linguistics*, Las Cruces/NM.
- Jacobs, J. (2001). The dimensions of Topic–Comment. *Linguistics*, 39(4):641–681.
- Krifka, M. and Musan, R. (2012). Information structure: Overview and linguistic issues. In Krifka, M. and Musan, R., editors, *The Expression of Information Structure*. De Gruyter Mouton, Berlin/Boston.
- Kruijff-Korbayová, I. and Steedman, M. (2003). Discourse and information structure. *Journal of Logic, Language and Information: Special Issue on Discourse and Information Structure*, 12(3):249–259.
- Mann, W. and Thompson, S. (1988). Rhetorical Structure Theory: Towards a functional theory of text organization. *TEXT*, 8:243–281.
- Matić, D., van Gijn, R., and van Valin, R. (2014). Information structure and reference tracking in complex sentences: An overview. In van Gijn, R., Hammond, J., Matić, D., van Putten, S., and Galucio, A. V., editors, *Information structure and reference tracking in complex sentences*. John Benjamins, Amsterdam.
- Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., and Webber, B. (2008). The Penn Discourse TreeBank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC)*.
- Prince, E. F. (1992). The ZPG letter: Subjects, definiteness, and information status. In Mann, W. and Thompson, S., editors, *Discourse description: Diverse linguistic analyses of a fund-raising text*, pages 223–255. John Benjamins, Amsterdam.
- Ritz, J., Dipper, S., and Götze, M. (2008). Annotation of information structure: An evaluation across different types of texts. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC)*, pages 2137–2142, Marrakech, Morocco.
- Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. In Yoon, J. and Kathol, A., editors, *OSU Working Papers in Linguistics 49: Papers in Semantics*, pages 91–136. The Ohio State University, Columbus.
- Sgall, P., Hajičová, E., and Panevová, J. (1986). *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Reidel, Dordrecht.
- Sidarenka, U., Peldszus, A., and Stede, M. (2015). Discourse Segmentation of German Texts. *Journal of Language Technology and Computational Linguistics (JLCL)*, 30(1):71–98.
- 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. (2012). Computation and modeling of information structure. In Krifka, M. and Musan, R., editors, *The Expression of Information Structure*. De Gruyter Mouton, Berlin/Boston.
- Stede, M., editor. (2016). *Handbuch Textannotation: Potsdamer Kommentarkorpus 2.0*. Universitätsverlag, Potsdam.
- Strube, M. and Hahn, U. (1996). Functional centering. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, pages 270–277. Assoc. for Computational Linguistics.