QUD-Based Annotation of Discourse Structure and Information Structure: Tool and Evaluation

Kordula De Kuthy*, **Nils Reiter**†, **Arndt Riester**†
*Department of Linguistics, Wilhelmstr. 19, 72074 Tübingen, Germany
†Institute for Natural Language Processing, Pfaffenwaldring 5B, 70569 Stuttgart, Germany
kdk@sfs.uni-tuebingen.de,{nils.reiter,arndt.riester}@ims.uni-stuttgart.de

Abstract

We discuss and evaluate a new annotation scheme and discourse-analytic method, the *QUD-tree* framework. We present an annotation study, in which the framework, based on the concept of *Questions under Discussion,* is applied to English and German interview data, using *TreeAnno,* an annotation tool specially developed for this new kind of discourse annotation. The results of an inter-annotator agreement study show that the new annotation method allows for reasonable agreement with regard to discourse structure and good agreement with regard to the annotation of information structure, which covers *focus, background, contrastive topic* and *non-at-issue material*.

Keywords: discourse structure, information structure, annotation, inter-annotator agreement, Question under Discussion

1. Introduction

In this paper, we evaluate a new annotation scheme and discourse-analytic method, the *QUD-tree* framework, developed in Reyle and Riester (2016), Riester (to appear) and Riester et al. (to appear). Its purpose is the cross-linguistic analysis of *information structure* and *discourse structure* of textual data. We furthermore introduce a new tool, *TreeAnno*, which enables the analyst to semi-automatically segment texts, systematically enhance them with implicit *Questions under Discussion (QUDs)*, and transform the data into a new kind of discourse tree called *QUD tree*.

2. QUD trees

For several decades scholars have been claiming that implicit questions (so-called *Questions under Discussion*, or QUDs) are constitutive of the internal structure of texts, e.g. Polanyi (1988), Stutterheim and Klein (1989), van Kuppevelt (1995) or, recently, e.g. Onea (2016), Velleman and Beaver (2016), Riester (to appear). This means that every statement contained in a text is seen as the immediate answer to precisely one implicit or explicit QUD, and potentially also as an indirect answer to one or several more general QUDs. QUDs can be thought of as fine-grained, silent headlines of sections, subsections etc., down to the bottom-level of atomic assertions. The content of the latter bottom-level QUDs is, at the same time, the *background* (or *topic*) of the assertions dominated by it, while the constituent of the assertion that answers the question is called the *focus*. Therefore, QUDs also determine the information structure of the assertions contained in the text (Roberts, 2012), which is then reflected in specific choices of constituent order or the use of cleft constructions and, as to spoken discourse, in a characteristic prosodic realization.

To our knowledge, the QUD-tree method (Riester et al., to appear), briefly sketched in the following, is the first corpus annotation framework which actually *implements* the reconstruction of QUDs, which have, so far, mainly been discussed from a theoretical or experimental perspective. It is at the same time one of the few frameworks that deals with discourse structure and information structure simultaneously; but compare e.g. Kruijff-Korbayová and Steedman (2003), Polanyi et al. (2003), Poláková et al. (2013).

In our framework, we opt for a particular compact representation format for QUD trees, in which the textual assertions (A) represent the terminal nodes of a discourse tree (preserving the linear order of the text from left to right) while (implicit or explicit) QUDs (Q) form the non-terminal nodes. An abstract QUD tree is shown in Figure 1.

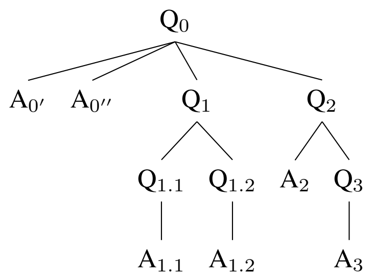


Figure 1: QUD tree

QUD trees combine properties of two types of structures from the literature. On the one hand, they comprise *d-trees* (Büring, 2003), which were designed to capture question-subquestion relations (e.g. $Q_1 > Q_{1.1}$) that guide the occurrence of so called *contrastive topics*. On the other hand, QUD trees can be systematically mapped onto discourse graphs from Segmented Discourse Representation Theory (SDRT; Asher and Lascarides, 2003). In a QUD tree, two assertions (e.g. $A_{0'}$, $A_{0''}$) count as coordinated whenever they are siblings under a joint QUD. An assertion (e.g. A_3) is subordinate to another assertion (A_2) whenever A_3 is the answer to a QUD Q_3 that is a sibling to A_2, which also means that Q_3 contentwise depends on A_2. The crucial difference between SDRT graphs and QUD trees is, of course, that the former contain discourse relations whereas the latter contain question nodes instead. For details and other tree formats representing the same information see Riester (to appear).

3. QUDs and information structure

In line with e.g. Rooth (1992), Büring (2003), Beaver and Clark (2008), Krifka (2008) or Roberts (2012), we assume a question-based definition of information-structural categories, shown in Table 1, which is also used in the examples below.[1]

Category (Label)	Definition
Focus domain (∼)	Piece of discourse that has the same background as the current QUD and that contains a focus
Focus (F)	Constituent that answers the current QUD
Background (BG)	Material mentioned in the current QUD
Contrastive topic (CT)	Material backgrounded w.r.t. the current QUD and focal w.r.t. a super-question
Non-at-issue material (NAI)	Optional material w.r.t. the current QUD

Table 1: Information structure: Label inventory

(1) Q_{15}: {What did the President want to do?}
 $>$ A_{15}: [It was clear from the President's speech that]$_{\text{NAI}}$ [[he wanted to]$_{\text{BG}}$ [make minor changes]$_{\text{F}}$]∼

(1) is an example demonstrating the assignment of information-structure labels in the context of a QUD (in curly brackets). Note that the indentation ($>$) of A_{15} in the textual representation marks subordination in the discourse tree, as shown in Figure 2.

Figure 2: Question-answer pair as part of a discourse tree

4. QUDs and discourse structure

The QUD-tree framework can be applied to any kind of written or spoken discourse or conversation. It is not language-specific and can, in principle, be used in order to investigate data from any language the analyst is able to understand. Since the analysis procedure is described at great length within a separate guidelines document (Riester et al., to appear), we will limit the account to a minimum here.

4.1. Segmentation

Raw texts are segmented into atomic assertions. Apart from orthographic sentence boundaries, segmentation also applies at (2) (information-structurally relevant) coordinations and (3) before (optional) syntactic adjuncts. (Obligatory) sentential arguments (4) are not split off.

(2) A_1: You were working until last summer for the NSA
 A_2: and during this time you secretly collected thousands of confidential documents.

(3) A_{27}: So there is a sort of a trading dynamic there
 A_{28}: but it's not…it's not open.

(4) A_{30}: What they are saying is **that they will not then target people within that data**.

4.2. QUD principles for given information

The actual identification of a QUD for each assertion is guided by a number of well-established principles adapted from the formal literature on information structure (Rooth, 1992; Schwarzschild, 1999; Büring, 2008; Büring, 2016), cf. Riester et al. (to appear):

Q-A-CONGRUENCE: QUDs must be answerable by the assertion(s) that they immediately dominate.

Q-GIVENNESS: Implicit QUDs can only consist of given (or, at least, highly salient) material.

MAXIMIZE-Q-ANAPHORICITY: Implicit QUDs should contain as much given (or salient) material as possible.

Example (5) shows that from these principles we can derive QUD Q_{32} for assertion A_{32} in the context of A_{31}, whereas any of the questions in (6), used in place of Q_{32}, would violate at least one of the QUD constraints in the same context.

(5) A_{31}: So, if they want to spy on a British citizen, they can spy on a British citizen.
 Q_{32}: {What can they do with that data?}
 $>$ A_{32}: and [[then they can even]$_{\text{BG}}$ [share]$_{\text{F}}$ [that data]$_{\text{BG}}$ [with the British government]$_{\text{F}}$]∼

(6) a. {What about spying?} (#Q-A-CONGRUENCE)
 b. {What about the British government?} (#Q-GIVENNESS)
 c. {What can they do next?} (#MAXIMIZE-Q-ANAPHORICITY)

The tree corresponding to (5) is shown in Figure 3. As a rule, the fact that an assertion contains given material leads to the subordination of that assertion under the assertion containing the antecedent. (Recall our definition of *subordination* at the end of Section 2.)

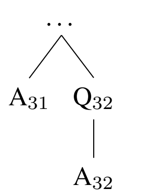


Figure 3: Subordinated assertion A_{32}, containing given material

4.3. QUD principle for parallel information

Two or more assertions are defined as parallel if and only if they share some semantically identical content and represent partial answers to the same QUD, see Example (7).

[1]Aboutness topics (T), defined in the guidelines as "referential entities (terms) in the background", were left out of consideration, in order to avoid categorical embedding.

PARALLELISM: The background of a QUD with two or more parallel answers consists of the (semantically) common material of the answers.

(7) Q_{19}: {What about the programs?}
  > $A_{19'}$: [[they've]$_{BG}$ [never stopped a terrorist attack in the United States]$_{F}$]∼
  > $A_{19''}$: and [[they have]$_{BG}$ [marginal utility at best for other things]$_{F}$]∼.

The resulting tree structure is shown in Figure 4.

Q_{19}

$A_{19'}$ $A_{19''}$

Figure 4: Two coordinated (parallel) assertions

5. TreeAnno

To support the annotation of QUDs according to the above described QUD principles, we developed *TreeAnno*, a web-based tool for the transformation of written text into QUD trees. Conceptually, two distinct steps are involved in the annotation: The segmentation of the text into appropriate units, and their hierarchical organization in the form of QUD trees. The tool imports plain text files, and first adds automatically detected sentence (and token) boundaries (using LanguageTool[2]), as an initial segmentation. The boundaries can later be changed by the annotators. The tree annotation works by indenting segments, similarly to an outliner. This allows for fast annotation of large text segments. Implicit QUDs can be inserted as well. Figure 5 shows a screenshot of the annotation view of *TreeAnno*.

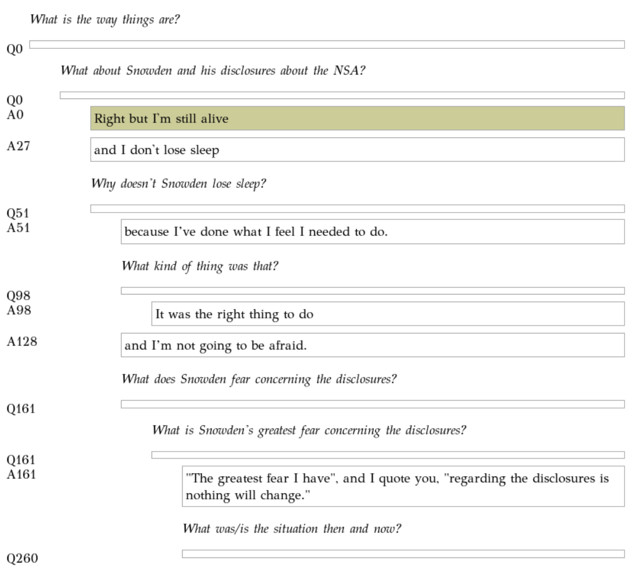


Figure 5: Annotation view of *TreeAnno*: discourse structuring and added QUDs *(italics)*

Behind the scenes, the sentences (and implicit QUDs) are represented as annotations in a UIMA[3] document. The connection to actual textual positions is always kept and allows future integration with other linguistic annotation layers. The tree structure is provided by maintaining a reference to the parent of each annotation.
In addition, TreeAnno supports the export of annotated documents as a simple, tree-oriented XML format, bracket expression (with or without node ids), visually rendered trees (by using GraphViz) or a chart-like matrix.
To sum up, *TreeAnno* allows for an easy transfer of QUD-tree annotations, which have, for instance, initially been carried out by various annotators in a text editor, into a generic XML or bracketing format. It furthermore provides the possibility to visualize, compare and evaluate different annotations, e.g. the ones shown in Figure 6. The tool will be made available as open source software released with this publication.[4]

6. Evaluation: Discourse structure

In a first evaluation of the QUD guidelines, our goal is to show that the above described method of discourse annotation in terms of QUDs can be applied reliably to naturally occurring data. We conducted an empirical study, in which annotators followed the QUD guidelines described in Riester et al. (to appear) to annotate English and German interview data with QUDs, using the above described TreeAnno tool.

6.1. Evaluation setup

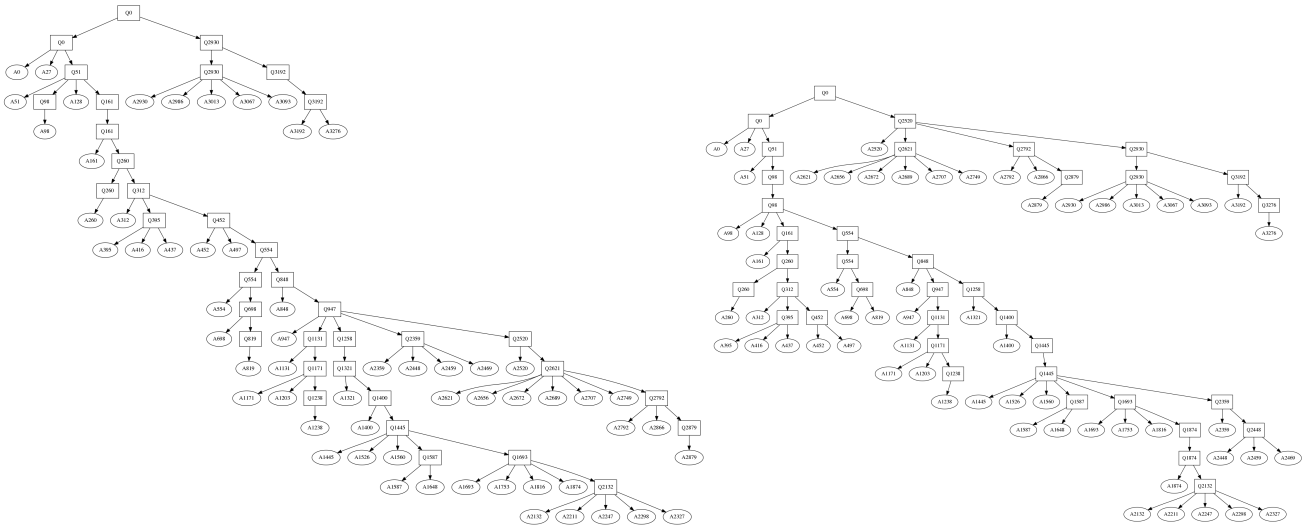
Two trained annotators analyzed two sections from a transcript of an (English) interview with Edward Snowden, broadcast on German ARD TV on Jan. 26, 2014.[5] The first section of the transcript consists of 60 text segments, the second has 69 text segments. The two resulting discourse trees for the first segment are shown in Figure 6.
Two other trained annotators analyzed a German radio interview (SWR2 radio interview with Thomas Oppermann, Social Democratic Party, Sept. 12, 2015), in the form of a single document consisting of 158 segments.[6]

6.2. Method and results

For the comparison of two QUD annotations we need to be able to calculate an inter-annotator agreement score that takes into account, for every segment and every possible span of segments, whether a QUD is present or not. In order to compute a κ statistics based on our QUD annotations, we follow the method described in Marcu et al. (1999), which was developed for measuring agreement in the labeling of rhetorical structure categories in texts. The method is based on the idea of mapping the hierarchical structure of a discourse tree onto a matrix filled with categorical values (in our case whether there exists a (Q)uestion spanning the

[2] `https://languagetool.org`
[3] `http://apache.uima.org`

[4] `http://hdl.handle.net/11022/1007-0000-0007-C634-F`
[5] `https://archive.org/details/snowden_interview_en`
[6] The interview is part of the GRAIN corpus (Schweitzer et al., 2018), which, among other data and annotation layers, comprises twelve more interviews of the same kind, which are all analyzed for QUD trees and information structure.



QUD tree of annotator 1 QUD tree of annotator 2

Figure 6: Two different QUD tree analyses for the same document

respective segments – start to end – or (n)ot). The resulting chart for the sample QUD tree of Figure 1 is shown in Figure 7, in which we observe, for instance, that the root node Q_0 spans over all the tree segments, i.e. from $A_{0'}$ to A_3. Note that the indices on the segments/assertions and questions in this Figure are marked only for the purpose of demonstration, while, in fact, the labels are binary.

	$A_{0'}$	$A_{0''}$	$A_{1.1}$	$A_{1.2}$	A_2	A_3
$A_{0'}$	n	n	n	n	n	Q_0
$A_{0''}$	–	n	n	n	n	n
$A_{1.1}$	–	–	$Q_{1.1}$	Q_1	n	n
$A_{1.2}$	–	–	–	$Q_{1.1}$	n	n
A_2	–	–	–	–	n	Q_2
A_3	–	–	–	–	–	Q_3

Figure 7: A text segment chart representing a QUD tree

A κ statistics can now be computed between two charts that represent two different QUD annotations for the same text; more precisely between the two resulting sets of cells in the upper half of each chart. In the case of Figure 7, this amounts to sets of 21 pairs of cells. Generally, for n segments contained in a document, the number of cells is $\frac{n \times (n+1)}{2}$.

For our three annotated documents we calculated κ (Cohen, 1960), based on the the described method. For the text *Snowden 1,* consisting of 60 segments, we calculated the κ statistics based on 1,830 items, for *Snowden 2* with 69 segments based on 2,415 items. And for the German *Oppermann* text, the κ is based on 12,535 items resulting from 158 segments. The results are shown in Table 2

The values show moderate agreement between the annotator pairs. It is entirely clear, though, that the basis of computation in this new task is rather different than, for instance, in a word-based classification task. It is, therefore, perhaps still too early to interpret the results, due to the overall complexity of the task and the lack of a reasonable baseline. However, the results can provide a point of reference for future developments in this area.

Text	Segments	Cells	κ
Snowden 1 (ENG)	60	1,830	.50
Snowden 2 (ENG)	69	2,415	.53
Oppermann (GER)	158	12,561	.45

Table 2: Kappa values for QUD-annotated spoken dialogue

7. Evaluation: Information structure

The second major issue we are interested in is to evaluate the reliability of information-structure annotation based on the previous identification of QUDs.

7.1. Evaluation setup

For the evaluation of the information structure mark-up, the same two documents from the English Snowden interview as well as the transcript of the German Opperman interview were annotated. The same pairs of trained annotators now performed an information-structure annotation of the text segments, still in keeping with the guidelines of Riester et al. (to appear). To keep matters simple, we concen-

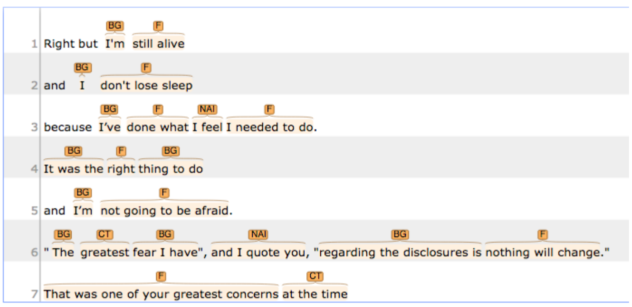


Figure 8: Annotation of information structure in WebAnno

trated on the four categories focus (F), background (BG), non-at-issue material (NAI) and contrastive topic (CT). Focus domain labels (∼) were not annotated, since each text segment (assertion) already corresponds to one focus domain. The annotators based their annotations on the previously performed QUD analysis in the TreeAnno tool described in Section 6.2. As an annotation tool for the token-based information-structure annotation, WebAnno (Yimam et al., 2013) was chosen. Figure 8 shows a screenshot of the information-structure annotation of the beginning part of *Snowden 1*.

7.2. Method and results

Following previous work on the evaluation of information structure annotation (Ritz et al., 2008; Calhoun et al., 2010), we calculated κ values on the annotated data based on tokens. In addition to the specifications in Riester et al. (to appear), in particular the QUD-to-information-structure mapping from Table 1, we defined a number of heuristic (but potentially debatable) rules in order to prevent disagreement due to theoretically unclear issues:

- Discourse connectors (*but, and, although, because, therefore* etc.) at the beginning of discourse segments are not annotated.
- All pronouns (including possessive pronouns), unless contrastive, receive the label BG.
- Function words like auxiliaries, prepositions, discourse particles, articles or complementizers are either labeled as BG or as F, depending on what they adjoin to. In case a function word occurs between a focus and a background, it is backgrounded, on the assumption that it represents salient information.

  (8) Q: {What about John?}
  A: [He **has been**]$_{\text{BG}}$ [lucky]$_{\text{F}}$.

  Exceptions:

  – If the wh-word provides that a function word is part of the focus, or if two function words are explicitly contrasted against each other, then they receive the label F.

    (9) Q: {Whom did she meet?}
    A: [She met]$_{\text{BG}}$ [**the** Pope]$_{\text{F}}$.

  – If a function word is adjoint to an overtly contrastive word, it stays in the background.

    (10) Q: {What kind of cars were there?}
    A': [**A**]$_{\text{BG}}$ [red]$_{\text{F}}$ [**one**]$_{\text{BG}}$
    A”: and [**a**]$_{\text{BG}}$ [green]$_{\text{F}}$ [**one**]$_{\text{BG}}$.

- Punctuation: Quotation marks around an expression, commas within and at the right edge of an expression are part of the markable. Periods, colons, semicolons, exclamation marks are not.

Text	Label	Tokens	κ
Snowden 1	all	657	.69
(ENG)	F		.67
	BG		.46
	CT		.55
	NAI		.71
Snowden 2	all	842	.67
(ENG)	F		.65
	BG		.57
	CT		.61
	NAI		.71
Oppermann	all	1646	.67
(GER)	F		.63
	BG		.60
	CT		.14
	NAI		.69

Table 3: Kappa for information structure annotation

Results are shown in Table 3, divided into scores for all labels taken together, and individual scores for each of the four labels.
The results show that the described method supports the successful information-structure annotation with substantial agreement (i.e. $\kappa > .6$) between two annotators: On the *Snowden 1* text, the agreement score for all annotated categories taken together is at .69, for the category F (focus) the score is .67, while on *Snowden 2*, the score for all categories is at .67, for F at .65. As for the German text, the overall agreement score is again .67, for the category F the score is .63. However, the score for CT (contrastive topic) was very low here. Moreover, in general, there was a high agreement for the classification of non-at-issue material (NAI).
These agreement scores are much higher and exhibit more reliable annotation results than, for example, the results reported in (Ritz et al., 2008) for a similar annotation study on naturally occurring data. In their study, the highest κ value calculated for focus on all tokens of a spoken dialogue is at .44. Other studies reporting higher κ values usually did not base their annotation on all tokens of a text or used fewer categories in the annotation. Calhoun et al. (2010), for example, report a κ value of .67 for the binary distinction between kontrast (their terminology for focus) and background. However, in their study not all tokens of a text but only certain words were annotated, i.e. nouns, verbs, adjectives, adverbs and pronouns. Summing up, these high agreement scores show that the successful annotation of information structure in spoken-language data based on explicit QUDs is very promising, despite the fact that there is still some degree of disagreement on the QUD-based discourse structures.

8. The formulation of QUDs

Finally, we would like to address the similarity of the actual QUD formulations chosen by the two annotators. It might seem surprising at first that we are not evaluating this issue, although a substantial amount of variation can be expected here and indeed occurs, as shown, for instance, in Figure 9. We think that it would be rather futile to evaluate the string match between two annotators' free QUD formulations, since language allows for endless possibilities of variation when expressing any statement or question. But this is not an important point, as far as our task is concerned. What counts is that the two QUDs chosen by the annotators have the same denotation, i.e. give rise to the same discourse-structure analysis and information-structure classification, which in the case of Figure 9 is fulfilled.

	Annotator 1:	**Annotator 2:**
Q_9:	{What did the political class do?}	{What exactly did the government do?}
$> A_9$:	[Instead of circling around the public and protecting their rights]$_{\text{NAI}}$ [the political class]$_{\text{BG}}$ [circled around the security state]$_{\text{F}}$	

Figure 9: Two different QUDs by two annotators, which lead to the same discourse and information structure

9. Conclusion

We have presented a novel method for the annotation of information structure which achieves good inter-annotator scores. The method is based on the reconstruction of QUDs, which moreover leads to the definition of a new kind of discourse structure, QUD trees. Although initially earmarked for the annotation of information structure, QUD trees represent an interesting contribution to discourse theory itself, which can be analyzed with a reasonable agreement. Finally, we introduced a new annotation tool for QUD trees, TreeAnno, which is made available as open source software.

10. Acknowledgements

We would like to thank our student annotators Daniel Knaus and Nadja Schäfer for their dedicated work and helpful comments. We would also like to thank Detmar Meurers for his insightful comments and discussion. This work was funded by the German Research Foundation (DFG) via the Tübingen Sonderforschungsbereich 833, Project A4 and the Stuttgart Sonderforschungsbereich 732, Project A6. TreeAnno has been developed within the Center for Reflected Text Analytics (CRETA), funded by the German Ministry of Education and Research.

11. Bibliographical References

Asher, N. and Lascarides, A. (2003). *Logics of Conversation*. Cambridge University Press.

Beaver, D. and Clark, B. (2008). *Sense and Sensitivity. How Focus Determines Meaning*. Wiley-Blackwell, Oxford.

Büring, D. (2003). On D-trees, beans, and B-accents. *Linguistics and Philosophy*, 26(5):511–545.

Büring, D. (2008). What's new (and what's given) in the theory of focus? In *Proceedings of the 34th Annual Meeting of the Berkeley Linguistics Society*, pages 403–424.

Büring, D. (2016). *Intonation and Meaning*. Oxford University Press.

Calhoun, S., Carletta, J., Brenier, J., Mayo, N., Jurafsky, D., Steedman, M., and Beaver, D. (2010). The NXT-format Switchboard corpus: A rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Language Resources and Evaluation*, 44:387–419.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.

Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, 55(3-4):243–276.

Kruijff-Korbayová, I. and Steedman, M. (2003). Discourse and information structure. *Journal of Logic, Language and Information*, 12:249–259.

Marcu, D., Amorrortu, E., and Romera, M. (1999). Experiments in constructing a corpus of discourse trees. In *Proceedings of the ACL Workshop on Standards and Tools for Discourse Tagging*, pages 48–57, College Park, MD, USA.

Onea, E. (2016). *Potential Questions at the Semantics-Pragmatics Interface*. Brill, Leiden.

Poláková, L., Mírovský, J., Nedoluzhko, A., Jínová, P., Zikánová, Š., and Hajičová, E. (2013). Introducing the Prague Discourse Treebank 1.0. In *Proceedings of the International Joint Conference on Natural Language Processing*, pages 91–99, Nagoya.

Polanyi, L., van den Berg, M., and Ahn, D. (2003). Discourse structure and sentential information structure: An initial proposal. *Journal of Logic, Language and Information*, 12:337–350.

Polanyi, L. (1988). A formal model of the structure of discourse. *Journal of Pragmatics*, 12(5-6):601–638.

Reyle, U. and Riester, A. (2016). Joint information structure and discourse structure analysis in an Underspecified DRT framework. In Julie Hunter, et al., editors, *Proceedings of the 20th Workshop on the Semantics and Pragmatics of Dialogue (JerSem)*, pages 15–24, New Brunswick, NJ, USA.

Riester, A., Brunetti, L., and De Kuthy, K. (to appear). Annotation guidelines for Questions under Discussion and information structure. In Evangelia Adamou, et al., editors, *Information Structure in Lesser-Described Languages: Studies in Syntax and Prosody*. Benjamins, Amsterdam.

Riester, A. (to appear). Constructing QUD trees. In Malte Zimmermann, et al., editors, *Questions in Discourse. Vol. 2: Pragmatics*. Brill, Leiden.

Ritz, J., Dipper, S., and Götze, M. (2008). Annotation of information structure: An evaluation across different types of texts. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC)*, pages 2137–2142, Marrakesh.

Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5(6):1–69.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, 1(1):75–116.

Schwarzschild, R. (1999). GIVENness, AvoidF, and other constraints on the placement of accent. *Natural Language Semantics*, 7(2):141–177.

Schweitzer, K., Eckart, K., Gärtner, M., Faleńska, A., Riester, A., Rösiger, I., Schweitzer, A., Stehwien, S., and Kuhn, J. (2018). German Radio Interviews: The GRAIN release of the SFB732 silver standard collection. In *Proceedings of the Eleventh Language Resources and Evaluation Conference (LREC)*, Miyazaki, JP.

Stutterheim, C. v. and Klein, W. (1989). Referential movement in descriptive and narrative discourse. In Rainer Dietrich et al., editors, *Language Processing in Social Context*, pages 39–76. North Holland, Amsterdam.

van Kuppevelt, J. (1995). Discourse structure, topicality and questioning. *Journal of Linguistics*, 31:109–149.

Velleman, L. and Beaver, D. (2016). Question-based models of information structure. In Caroline Féry et al., editors, *The Oxford Handbook of Information Structure*, pages 86–107. Oxford University Press.

Yimam, S. M., Gurevych, I., Eckart de Castilho, R., and Biemann, C. (2013). WebAnno: A flexible, web-based and visually supported system for distributed annotations. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 1–6, Sofia.