

# Discrepancies Between Database- and Pragmatically Driven NLG: Insights from QUD-Based Annotations

**Christoph Hesse** ✉

Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin, Germany

**Maurice Langner** ✉

Department of Linguistics, Ruhr-University Bochum, Germany

**Anton Benz** ✉

Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin, Germany

**Ralf Klabunde** ✉

Department of Linguistics, Ruhr-University Bochum, Germany

---

## Abstract

We present annotation findings when using an annotated corpus of driving reports as informational texts with an elaborated pragmatics for the automatic generation of corresponding texts. The generation process requires access to a database providing the technical details of the vehicles, as well as an annotated corpus for sophisticated, pragmatically motivated text planning. We focus on the annotation results since they are the basic framework for linking text planning with database queries and microplanning. We show that the annotations point to a variety of linguistic phenomena that have received little or no attention in the literature so far, and they raise corresponding questions regarding the access to information from databases for the generation process.

**2012 ACM Subject Classification** Applied computing → Arts and humanities

**Keywords and phrases** NLG, question-under-discussion analysis, information structure, database retrieval

**Digital Object Identifier** 10.4230/OASICS.LDK.2021.32

**Supplementary Material** *Software (Corpus with annotated QUD-tree structures)*: <https://github.com/christoph-hesse/question-under-discussion>

*Software (Annotation Tool)*: <https://github.com/MMLangner/QUDA>

**Funding** This work has been supported by the Deutsche Forschungsgemeinschaft, project “Propositional and Non-at-issue Content in Text Generation: Exploring the QUD-Perspective”, grants AB 4348/5-1; KL 1109/7-1.

## 1 Generation (NLG) of pragmatically rich texts

Sufficiently precise annotations of linguistic phenomena in corpora are the backbone of almost every task in natural language processing (NLP) and generation (NLG). The interplay between theory-driven assumptions, the creation of annotation guidelines and the actual data analysis during the annotation process determine the quality of the annotation result.

In this paper, we are presenting our findings when annotating pragmatically motivated information structures in texts for the generation of driving reports. The generation process requires access to a database for retrieving information about the corresponding vehicles. For this, we may use the database of the ADAC, Germany’s largest automobile club. However, driving reports do not only inform about technical details, but about subjective impressions and evaluations of the test driver as well, so that these texts are a mix of factual with subjective assertions. The gap between database retrieval and the presentation of subjective, evaluative information has – at least partially – to be closed by a learning approach that is not subject of this paper. Rather we show that even though the annotated information structures provide basic constraints for database retrieval, the annotation results point to a non-trivial match between annotation and retrieval.



© Christoph Hesse, Maurice Langner, Anton Benz, and Ralf Klabunde; licensed under Creative Commons License CC-BY 4.0

3rd Conference on Language, Data and Knowledge (LDK 2021).

Editors: Dagmar Gromann, Gilles Sérasset, Thierry Declerck, John P. McCrae, Jorge Gracia, Julia Bosque-Gil, Fernando Bobillo, and Barbara Heinisch; Article No. 32; pp. 32:1–32:9

OpenAccess Series in Informatics



OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The automatic generation of texts comprises several steps. It starts with providing the information to be verbalized from a data set and continues with the hierarchical ordering of the information (the text plan), the decision which information to realize in single sentences, ending with the language-specific grammatical tasks of determining the lexical items and grammatical encoding [5]. These individual tasks are not independent of each other, which holds even true for the first steps, the retrieval of information from a database and its arrangement in the text plan. If the task is not to generate a purely informational text but one which realizes pragmatically motivated information as, e.g., subjective estimations, attitudes and valuations, the text plan should include these kinds of information as well.

Typically, text plans are based on rhetorical relations for linking text spans in a coherent way [20]. However, rhetorical relations do not trigger constraints w.r.t the information structure in the respective text spans, and, from a theory-oriented point of view, due to their blurry definition, their explanatory power is rather limited.

We assume that question-under-discussion (QUD) approaches are more suitable for a theory-based construction of text plans. QUDs are the central concept in analyses that explain linguistic regularities as a consequence of the assumption that the sentences and text segments with which the regularities are associated are answers to an explicitly or implicitly asked question. QUDs figured prominently in theories explaining sequences of possible dialogue moves [3, 6], contextual relevance [16], information structural concepts (e.g. the *topic/focus* distinction, [16, 18, 19]), temporal progression in narration [9], and the analysis of coherence relations and subordination in text and dialogue [9, 18]. Although QUDs have been firmly established in theoretical linguistics, including the theory of discourse and dialogue, there has been, to our knowledge, only one attempt at developing guidelines and tools for text annotation [12, 15, 14], and no attempt at applying them in NLG systems. By generating driving reports, we aim at closing this gap.

In order to illustrate the annotation problem, let us have a look at a section on technical details in a German driving report about a motor bike. Driving reports are characterized by the fact that they combine factual information with subjective driving impressions and quality estimations. The constituents in that section that express information from a corresponding database are given in boldface.

1. **Eine voll ausgestattete Africa Twin mit elektronischem Fahrwerk kostet 18.665 Euro. Inklusive Liefer-Nebenkosten und Gepäcksystem** ist sie dann im Klub der **20.000-Euro-Reiseenduros** angekommen. Erste Probefahrten zeigen: Die Sache mit dem Fahrwerk ist kompliziert. Denn schon **das neue konventionelle Fahrwerk, in der Basis-Twin ausnahmslos verbaut, in der Sports Adventure regulär**, macht seine Sache ausgesprochen gut. Die für Letztere erhältliche **Option des elektronischen Fahrwerks (1600 Euro extra)** geht noch feinfühlig zur Sache, aber der Unterschied wird nicht für jedermann erfahrbar sein. Wir wetten: Die Topversion wird sich gut verkaufen.

**A fully equipped Africa Twin with electronic suspension costs 18,665 euros. Including ancillary delivery costs and luggage system**, it then joins the club of **20,000-euro touring enduros**. First test rides show: The matter with the chassis is complicated. **The new conventional suspension, which is fitted without exception in the basic Twin, and as standard in the Sports Adventure**, does its job extremely well. **The electronic suspension option available for the latter (1600 euros extra)** is even more sensitive, but the difference will not be noticeable for everyone. We bet: The top version will sell well.

The constituents that are realizing these database-related information are unequally linked to different information structural levels. For example, *The new conventional suspension, which is fitted without exception in the basic Twin* expresses at-issue and non-at-issue information (the latter by the apposition), as a whole it could be a focus, and it contains a topic (*the suspension*). The annotations will shed light on the relation between facts and levels of information structure.

## 2 QUD-based annotation

In order to elaborate the relation between textual expansion and the realization of different kinds of information structures, we annotated 30 German driving reports based on the Question-under-discussion approach. Each QUD triggers a number of information structural decisions that together determine the realization of database-related information and subjective meaning components.

These QUD-dependent information structural distinctions are focus/background, topic/comment (with a further distinction between discourse topic and contrastive topic) and at-issue/non-at-issue information. The focus/background distinction as well as topic/comment received much attention in the literature, and their discourse-related function seems, at first, to be worked out in sufficient detail so that their annotation should be largely trouble-free. However, our annotation efforts point to a number of intricate problems in applying these categories to the respective clauses and text segments.

In what follows, we will explain the annotation of QUDs, focus/background, topic/comment and at-issue/non-at-issue information according to the assumptions made in the literature and compare these guidelines with our insights from the annotation work. The consequences w.r.t. database retrieval for the generation process will be outlined.

### 2.1 The annotation process

Initially we were considering using the only existing annotation guidelines [15, 14] and tool [12] for this task, but we found two limitations that prompted us to develop our own: (1) The authors used QUD-tree-structures primarily for focus/background analysis, where focus constituents are taken to be answers to immediate super-QUDs, and background constituents are presupposed in the QUD. However, we were also interested in annotating topic/comment and evaluations through expressive, non-at-issue content. (2) The guidelines and the annotation tool allow for right-branching QUD-trees only.

We therefore developed our own XML-based annotation tool which allows for both right-branching and left-branching QUD structures.<sup>1</sup> In addition, our tool incorporates labeling of leaf node constituents for focus/background, topic/comment, at-issue/non-at-issue. It also incorporates indexing of constituents to capture phenomena such as split-focus/split-background where, for instance, a relative clause splits a focus constituent into a fragment before the relative clause and a fragment after it. An XML-based example is given below, together with its linguistic representation. Note that the QUD has been formulated by the annotator as a well justified question the sentence answers, and the information structures are derived from it.

```
<QUD String = "Was ist mit dem Antrieb?/What about the drive?">
  <F id="1"><SEGMENT> In der praktischen Außenhaut </SEGMENT></F>
  <CON>
    <SEGMENT> des 3,60 kurzen</SEGMENT>
    <SEGMENT>Fünftürers</SEGMENT>
  </CON>
  <F id="1"><SEGMENT> war der Antrieb </SEGMENT>
    <NAI><SEGMENT> erstmal </SEGMENT></NAI>
    <SEGMENT> kaum zu erkennen. </SEGMENT></F>
</QUD>
```

<sup>1</sup> The tool is available at <https://github.com/MMLangner/QUDA>.

2. [In der praktischen Außenhaut<sub>F<sub>1</sub></sub>] des 3,60 m kurzen Fünftürers [war der Antrieb [erstmal<sub>NAI</sub>] kaum zu erkennen<sub>F<sub>1</sub></sub>]  
 “In the practical outer skin of the 3.6 m short five-door car, the engine was hardly noticeable at first.”

Our tool also allows the use of indexing for capturing focus/background and topic/comment distinctions where one constituent actually consists of multiple pieces of information (e.g., enumeration of facts), which have to be mapped to individual database entries.

Each driving report was independently annotated by two trained annotators (student assistants with a linguistic background). First, we split driving reports into sections (e.g., teaser, introduction, main sections which usually consists of technical specifications as specified by the manufacturer and test drive results either confirming the manufacturer’s promise or not, available models of a vehicle and extras, and a summary/conclusion). Then, we proceed in a bottom-up fashion, annotating according to [15, 14] for focus/background, but also topic and non-at-issue. Above that we annotated common rhetorical structures such as *Contrast*, and above that structures which are more argumentative, for instance, why an author uses Contrast at this point (e.g., because of technical shortcomings of a vehicle, customer preferences/expectations or ongoing societal debates). Throughout this process, coreferential expressions (e.g., *der Fünftürer* [five-door car], *der kleinste Volkswagen* [the smallest VW], *der Stromer* [the electric], all referring to the same electric VW) are tracked, and indexing adjusted where needed. While QUDs on the bottom leaves are usually very concrete, the further up on the hierarchy QUDs are, the more abstract they become (e.g., QUDs for each section of a text, i.e., why certain technical details – e.g., opting for Diesel – are relevant given an issue raised earlier – e.g. Diesel Gate and the push for electric, thereby prompting *Contrast* comparisons with electric competitors).

## 2.2 QUDs

QUDs have to be formulated with database queries in mind. There are examples in our corpus of driving reports where this approach is easy enough to apply, and there are other examples where it is exceedingly more difficult. An easy example is an assertion about a car’s acceleration, which should have a QUD such as *What’s the car’s acceleration?* with the focus constituent containing some measure of acceleration, e.g. *from 0 to 100 km/h in 5.7 seconds*, a measure phrase which should be generated from a database query about the car’s acceleration. This approach is also easily applied to enumeration of facts, e.g., the different models available, which can be retrieved from the database through a QUD such as *What models are available?* This query would then return a list of size  $n$ , where each list element is a string. What makes QUD annotation more difficult is that authors of driving reports are often trying to maximise information density by speaking to multiple QUDs in the same proposition (e.g., what, what for, when, how, why QUDs), as in example 3:

3. Die Doppelstrategie ist aber gar nicht dumm, weil der Fahrer auf diese Weise sein spezielles Asphalt-Programm und ein ebenso spezielles Offroad-Programm hinterlegen kann.  
 “However, this dual strategy is not dumb at all because it lets the driver choose their special/preferred (on-road) driving program and an equally special offroad program.”

Annotators have to take extra care in deciding what the hierarchical relation between those QUDs is: Can they be put in a clear hierarchy where authors are foregrounding some QUD and incrementally background a proposition’s other QUDs, or are they equally important in light of a bigger argumentative point (e.g., that a release date may be unrealistic given engineering challenges or given societal pressures, or tastes of the target customer, or that expectations are contradicted as in example 3)?

In general, we want to be able to map the leaf nodes of our QUD trees to database entries, which puts constraints on the structure of those QUD trees. However, we find that QUD structures representing the more abstract levels of text interpretation show more variation across annotators given how they interpret the author’s intent. A QUD tree structure which might do full justice to an author’s nuanced argument might result in leaf nodes that are not easily mapped to the database, and vice versa, starting from leaf nodes that map easily to the database might necessitate a compromise higher up in the tree. We prioritize leaf nodes that easily translate to database queries. We made this choice not simply because a text generator would otherwise not be feasible, but because this also means that the resulting QUD structures towards the leaves of the tree can be evaluated in light of [12, 15, 14].

### 2.3 Focus/background

Focus and background are propositional attributes; the focus is that part of a proposition that is “new”, i.e. put into the foreground. Its complementary part is the background. Foci as informational units correspond to specific syntactic constituents, the focus domains. Each focus domain contains a so-called focus exponent – the prosodically most salient element of this focus domain [7]. In this paper, we do not make reference to the phonological properties of foci, however. The relation between the semantic/pragmatic notion of focus and its linguistic counterpart, the focus domain, have been subject to a number of studies, as well as the rules and principles responsible for determining the focus exponent [13, 11, 10].

The default statement in linguistic semantics is to identify focus/background structures with two fundamental functions: They are marking information that is new for the listener, or they contrast information with already realized information (the so-called contrastive focus).

Foci are answers to a QUD; as such they provide new information. Equating focus with new and the background with given information, however, ignores the fact that often a given/new distinction can hardly be drawn, which also leads to corresponding uncertainty in the annotation process. As a consequence, one cannot directly link focus-annotated information with database queries for receiving the required new information in the generation process. Here are some examples for the non-trivial link between focus in our linguistic data and database access:

Split-focus: In our data, some sentences have two focus constituents that express one focus together. For example, one QUD in a driving report is *What about the power unit?* Example (2) provides the answer for this QUD. A plausible assignment of focus is to tag *In der praktischen Außenhaut* (“In the practical outer skin”) and *war der Antrieb erstmal kaum zu erkennen* (“the power unit was hardly noticeable at first”) as being focused, but not “the 3.60 short five-door car” since this constituent doesn’t provide a part for the answer to the QUD.

A further but related phenomenon concerns sentences consisting of two coordinated main clauses, each with its own focus, but answering one QUD:

4. QUD: *How is the Renault Captur?*  
 Der Renault Captur [wächst]<sub>F</sub> und [verändert seinen Charakter]<sub>F</sub>.  
 “The Renault Captur grows and changes its character.”

It is reasonable to assume two separate foci since this coordination refers to two new aspects of the tested car. Both foci are well motivated by the QUD; they demonstrate that one QUD does not necessarily set up one focus only. Ellipses also indicate that the “one QUD – one focus” default can be violated:

5. QUD: *How have the aesthetics changed, compared to the old Captur?*  
 [das sieht scharf trainiert]<sub>F</sub> und [angriffslustig aus]<sub>F</sub>.  
 “that looks sharply trained and ready to attack.”

The non-elliptic sentence in German would be *das sieht scharf trainiert aus und das sieht angriffslustig aus*, with the prefix *aus* separated from the prefix verb *aussehen* and remaining in the base position, and the subject plus verb stem inserted in the second clause. The ellipsis forces an index as well for expressing that both foci belong together; otherwise the ellipsis cannot be handled correctly.

Without doubt, focused/new information is the information that must be retrieved from the database in order to present it to the user. However, as we have shown, the blurry distinction between given and new as well as the various partial mappings in our data between focus domains and foci do not allow for a unique retrieval process.

## 2.4 Topic/comment

Topics are discourse referents a statement – the comment – is made about. Linguistically, topic candidates are typically introduced by indefinite means and later on, they will be picked up by anaphoric expressions, resulting in a tree-like topic structure that describes what a section is about [8]. Topic structures might involve contrastive topics – discourse referents to whom discourse topics are compared. We are annotating both types of topics.

Topics are addressed by QUDs as well since they are mentioned in them. In order to retrieve the right information in the database, the topic referent must be given in it. As long as topic referents represent vehicles or vehicle parts, their annotations turned out to be easy. However, indexing for stating coreference of different topic expressions is sometimes unclear due to metonymy (see the examples above with reference to an electric VW).

## 2.5 At-issue/non-at-issue

Non-at-issue content is the part of an assertion that is optional in regard to the question under discussion, whereas at-issue is simply all relevant information given in the context. The optionality criteria is defined as the validity of the assertion as answer to the present QUD when the non-at-issue content is omitted. The lack of relevance which is implied by not being “at-issue” is limited to the context of the given QUD and does not entail the irrelevance of the presented information. According to [14], non-at-issue content itself denotes a different assertion including an associated subordinate QUD with a focus-topic distinction of its own, which is irrelevant in the context of the super-QUD in whose scope the constituents are not at-issue. Therefore, the annotation of non-at-issue is made more complex by the fact that depth and detailedness of the annotation decides how well non-at-issue can be distinguished in the respective context, and the identification of what is at-issue greatly depends on the choice of the QUD.

Non-at-issue content ranges from evaluative adverbs on sentence level to less obvious elements like embedding matrix clauses that name the source of a tradicted information, e.g., “[they say that<sub>NAI</sub>][the [car<sub>F</sub>] is [overpriced<sub>C</sub>]]”.

Retrieving information contained in non-at-issue from databases is highly complex. Evaluative adverbs, for example, mirror inferences and subjective impressions the author made on the basis of the propositional content. In example (6) technological understanding of relations between gas consumption and weight triggers evaluating the propositional content and expressing it as non-at-issue content. Therefore, this sort of content cannot simply be queried from a database but must be inferred from domain knowledge.

6. Surprisingly, the new Kawasaki consumes 4 litres fewer gas despite its 5% higher weight in comparison to the previous generation of this model.

The database we are using contains marks for different criteria of the cars, e.g. economic factors or build quality. These marks range from 1.0 (best) to 6.0 (worst) and were either calculated on the basis of sensor inputs (e.g., real gas consumption) or a set of rules. The annotation of non-at-issue in the corpus allows for the association of the evaluative adverbs with the marks given in the database, which provides the opportunity of predicting and probabilistically determining the usage of evaluative adverbs given the dataset from the database. [17] suggests different criteria for identifying non-at-issue content, which base on the observation that their content “survives under negation and projection”.

Annotators need to pay much attention to the applicability of these criteria when formulating QUDs and identifying non-at-issue.

### 3 Evaluation of the annotation results

As [1] point out, sources of representation problems in annotating corpora are ambiguity (several possible tags for one linguistic entity), variation (several variants for one variable exist), uncertainty (no sufficient knowledge for an unambiguous annotation available), error (annotating incorrectly) and bias (using an unbalanced corpus).

QUD-oriented annotating is inherently faced with uncertainty as annotation problem. Only hints for identifying QUDs can be given. Hence, there is no fixed set of theoretically justified QUDs for a certain text type, which results in a wide range of plausible QUDs. Since the information structures we are interested in can directly be derived from the formulated QUD (focus, topic, (non)-at-issue), uncertainty will be propagated to these annotation levels as well. An additional problem that might arise is ambiguity which becomes especially relevant when annotating focus, as our data show.

Furthermore, it is also difficult to compare our results with previous work on annotating information structure due to the varying linguistic complexity of the data, the coefficient used, the segment sizes for the tags used, and the different annotation guidelines controlling the annotation. Some exemplary studies shall illustrate this.

We primarily compared our QUD annotations with the results presented in [4] on QUD-based annotation of information structure, the only comprehensive QUD-based annotations we are aware of. Their data consist of sections of an English and a German interview which have been annotated by two trained annotators. Since the authors used Cohen’s  $\kappa$  as coefficient, based on a flattened representation of QUD trees in a matrix, we adopted this approach in order to achieve comparable results.

Measuring the agreement of the QUD annotations has been performed by the authors by first mapping the QUD trees to a matrix that represents the segments spanned by the QUDs and then calculating Cohen’s  $\kappa$  based on this matrix. The coefficients range between 0.45 and 0.53. The  $\kappa$  values for the information structural categories are acceptable or even robust (with the non-at-issue annotation having the highest values) with a negative outlier for contrastive topics.

We adopted this approach and achieved values between 0.45 and 0.78 for QUD annotations with a mean of 0.63, but without taking into account pre-theoretic heuristics as [4] did. For calculating  $\kappa$  values for information structural levels, [4] defined heuristic rules to prevent disagreement due to theoretically unclear cases. For example, all pronouns shall be annotated as background, and discourse connectors at the beginning of discourse segments are not annotated at all. We did not specify in the annotation guidelines how to annotate certain lexical items w.r.t. information structures. Instead all information structural decisions have to be derived from the respective QUD.

Since  $\kappa$  is sensitive to the units for which the statistic will be computed [2], we also computed the  $\gamma$  coefficient [21] for QUD annotations, since it especially allows us to measure for long spans of texts the categorization and unitizing as a joint task. In principle, Krippendorff's  $\alpha$  is also a suitable coefficient for categorizing and unitizing, but in some constellations it is sensitive to segment length, while  $\gamma$  treats short and long segments in the same way [21, p. 463]. This feature is especially relevant for our discourse-related annotations. The value  $\gamma = 1$  expresses that all annotators perfectly agree while  $\gamma < 0$  signals that the annotation result is worth less than annotating at random.

The overall  $\gamma$  is 0.13. Non-at-issue annotations result in  $\gamma = -0.075$ , focus results in  $\gamma = 0.115$ , background in  $\gamma = 0.08$ , and the auxiliary tag for every constituent that cannot be assigned to one of the information structural notions results in  $\gamma = 0.32$ . These values show a considerable gap between the frequently used  $\kappa$  statistic and the rarely used  $\gamma$ ; these results require a deeper analysis of the meaningfulness of applying these statistics to discourse phenomena.

However, in our  $\gamma$  statistic, for the structurally less important tags, agreement is three times higher than for topic, focus and non-at-issue. Further insights must be gained whether the combined text spans of the three main tags correspond, which would mean that the low  $\gamma$  agreement is due to confusion in the classification task. In general, QUD annotations are trees, which means that apart from the correct classification of the terminal nodes, the complexity of the tree structure needs to be compared as well. If the tree structures prove to be closely equivalent, this indicates that the low agreement is rooted only in the unique identification of topic, background and non-at-issue.

However, divergent discourse tree structures do not mean that the texts will be understood differently. Rather, what we observe is that the annotated structures express subjective opinions on the levels of information structure within texts, and structural decisions only weakly reflect subjective views.

## 4 Conclusion

The annotations show that retrieving information from a database for generating the text plan for driving reports involves an accurate annotation of focus, topic and non-at-issue information, since focused information should be retrievable from the database, and topic referents must be given as entities in the database as well. Non-at-issue information, often expressed by subjective estimations, require NAI-specific analyses of database entries in order to justify the use of corresponding linguistic means. For example, *überraschenderweise* “surprisingly” requires a comparison of the actual non-at-issue content with defaults in the database.

The annotation results contrasts with the high agreement concerning which facts should be realized. What the facts in the texts are is undisputed and what important facts are as well. We thus can state that the challenge is not to determine which facts to retrieve for their linguistic realisation, but how to do that on discourse level.

---

## References

- 1 Christin Beck, Hannah Booth, Mennatallah El-Assady, and Miriam Butt. Representation problems in linguistic annotations: Ambiguity, variation, uncertainty, error and bias. In *Proceedings of the 14th Linguistic Annotation Workshop*, pages 60–73, Barcelona, Spain, 2020. Association for Computational Linguistics. URL: <https://www.aclweb.org/anthology/2020.law-1.6>.
- 2 Jean Carletta. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22 (2):249–254, 1996.



- 3 Lauri Carlson. *Dialogue Games: An Approach to Discourse Analysis*. Reidel, Dordrecht, 1983.
- 4 Kordula De Kuthy, Nils Reiter, and Arndt Riester. QUD-based annotation of discourse structure and information structure: Tool and evaluation. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, 2018. European Language Resources Association (ELRA). URL: <https://www.aclweb.org/anthology/L18-1304>.
- 5 Albert Gatt and Emiel Krahmer. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research*, 61:65–170, 2018.
- 6 Jonathan Ginzburg. Interrogatives: Questions, facts, and dialogue. In Shalom Lappin, editor, *Handbook of Contemporary Semantic Theory*, pages 359–423. Blackwell, Oxford, 1996.
- 7 Daniel Glatz and Ralf Klabunde. Focus as perspectivation. *Linguistics*, 5(41):947–977, 2003.
- 8 Zachary Kimo Stine and Nitin Agarwal. Comparative discourse analysis using topic models: Contrasting perspectives on china from reddit. In *International Conference on Social Media and Society, SMSociety'20*, page 73–84, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3400806.3400816.
- 9 Wolfgang Klein and Christiane von Stutterheim. Quaestio und referentielle Bewegung in Erzählungen. *Linguistische Berichte*, 109:163–183, 1987.
- 10 Angelika Kratzer and Elisabeth Selkirk. New vs. given. In Daniel Altshuler and Jessica Rett, editors, *The Semantics of Plurals, Focus, Degrees and Times*, pages 157–162. Springer, 2019.
- 11 Manfred Krifka. Association with focus phrases. In Valerie Molnár and Susanne Winkler, editors, *The architecture of focus*, pages 105–136. Mouton de Gruyter, 2006.
- 12 Kordula De Kuthy, Nils Reiter, and Arndt Riester. QUD-Based Annotation of Discourse Structure and Information Structure: Tool and Evaluation. In Nicoletta Calzolari et al., editor, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France, May 2018. European Language Resources Association (ELRA).
- 13 Knud Lambrecht. *Information Structure and Sentence Form*. Cambridge University Press, 1994.
- 14 Arndt Riester. Constructing QUD trees. In Malte Zimmermann, Klaus von Heusinger, and Edgar Onea, editors, *Questions in Discourse: Pragmatics*, volume 2, pages 403–443. Brill, Leiden, 2019.
- 15 Arndt Riester, Lisa Brunetti, and Kordula De Kuthy. Annotation guidelines for questions under discussion and information structure. In Katharina Haude Evangelia Adamou and Martine Vanhove, editors, *Information Structure in Lesser-described Languages. Studies in prosody and syntax*, pages 403–443. John Benjamins, Amsterdam, 2018.
- 16 Craige Roberts. Information structure in discourse: Toward an integrated formal theory of pragmatics. In Jar Hak Yoon and Andreas Kathol, editors, *OSU Working Papers in Linguistics*, volume 49, pages 91–136. The Ohio State University, Department of Linguistics, Ohio, 1996.
- 17 Judith Tonhauser. Diagnosing (not-)at-issue content. In E. Bogal-Allbritten, editor, *Proceedings of the Sixth Conference on the Semantics of Under-represented Languages in the Americas and SULA-Bar*, Anherst, 2012. GLSA Publications.
- 18 Jan van Kuppevelt. Discourse structure, topicality, and questioning. *Journal of Linguistics*, 31:109–147, 1995.
- 19 Christiane von Stutterheim. *Einige Prinzipien des Textaufbaus: Empirische Untersuchungen zur Produktion mündlicher Texte*, volume 184 of *Reihe Germanistische Linguistik*. Niemeyer Verlag, Tübingen, 1997.
- 20 S. Williams and Ehud Reiter. Generating basic skills reports for low-skilled readers. *Natural Language Engineering*, 14 (4):495–525, 2008.
- 21 Antoine Widlöcher Yann Mathet and Jean-Philippe Métivier. The unified and holistic method gamma ( $\gamma$ ) for inter-annotator agreement measure and alignment. *Computational Linguistics*, 3(41):437–479, 2015.