Testing Focus and Non-at-issue Frameworks with a Question-under-Discussion-Annotated Corpus

Christoph Hesse[*]**, Maurice Langner**[†]**, Ralf Klabunde**[†]**, Anton Benz**[*]

[*]Leibniz-Zentrum Allgemeine Sprachwissenschaft Berlin, [†]Ruhr-Universität Bochum

{hesse, benz}@leibniz-zas.de, {maurice.langner, ralf.klabunde}@rub.de

Abstract

We present an annotated corpus of German driving reports for the analysis of Question-under-Discussion (QUD) based information structural distinctions. Since QUDs can hardly be defined in advance for providing a corresponding tagset, several theoretical issues arise concerning the scope and quality of the corpus and the development of an appropriate annotation tool for creating the corpus. We developed the corpus for testing the adequacy of QUD-based pragmatic frameworks of information structure. First analyses of the annotated information structures show that focus-related meaning aspects are essentially confirmed, indicating a suffcent accuracy of the annotations. Assumptions on non-at-issueness expressed by non-restrictive relative clauses made in the literature seem to be too strong, given the corpus data.

Keywords: Question-under-discussion (QUD), information structure, QUD-based corpus annotations

1. Introduction

This paper aims at testing pragmatic question-under-discussion (QUD) frameworks by means of a QUD-related annotated corpus.

QUD-based theories assume that texts should be analyzed as complex answers to implicit questions, the Questions-under-Discussions (QUDs). A QUD model describes the conversational goals a speaker wants to pursue. A QUD is a linguistic formulation of informational demands of the intended addressee. Form and content of the respective QUD, as well as the hierarchical organisation of QUDs and sub-QUDs, resulting in QUD trees, provide explanations for the realization of highly diverse pragmatic phenomena, for example referential movement (von Stutterheim and Klein, 1989), the distinction between main and side structures in texts (Carroll et al., 2003), negation effects (Anand and Matell, 2012), dialogue moves (Ginzburg, 2012), the interpretation of VP ellipsis (Kehler, 2015), projection behavior (Beaver et al., 2017), at-issueness (Koev, 2018), focus (Roberts, 2012), topic (van Kuppevelt, 1995), and others.

However, with some exceptions (e.g., (Kuthy et al., 2018; Riester et al., 2018; Anand and Matell, 2012)) these analyses are confined to constructed sample texts, showing the respective phenomenon at hand in a lucid way, but possibly at the prize of underestimating contextual influences and simplifying data.

For example, QUDs that have been tailored to the specific phenomenon to be studied tend to ignore the complexity of the content the QUD asks about. QUDs are linguistic reconstructions of assumed informational demands. These demands are context-sensitive and often linked to previous information expressed in the text. A corpus analysis of the resulting QUD representations makes these dependencies transparent. Moreover, interactions between different information structural levels are barely considered, and the complexity of sentence structures and their corresponding semantics are often reduced to the necessary scope.

By analyzing the QUD structures in our corpus, the focus annotations, and the annotation of non-at-issue content we show ways of extracting information from the corpus and point out possible theory-based analyses. QUDs should be considered from discourse-related and a propositional perspective, reflecting different aspects of textual unfolding. In essence, the focus annotations correspond to standard assumptions on focus, thus certifying sufficient quality of the annotations. This in turn supports our findings on the relation between non-at-issue content and its realization by non-restrictive relative clauses. Contrary to assumptions made in the literature, non-restrictive relative clauses can convey at-issue content, depending on the type of the formulated QUD.

In what follows, we first go into corpus development and annotation in more detail, including the annotation guidelines. Then we introduce the focus annotations and an analysis of the use of the focus-sensitive adverb *überhaupt* ('at all'). Finally, the analysis of non-at-issue content in non-restrictive relative clauses shows the advantages of the corpus-based analysis.

2. The QUDGen Corpus

The corpus (Hesse, C. and Klabunde, R. and Benz, A., 2021) comprises 30 German driving reports from online journals (welt.de and faz.net) with 922 sentences overall and 17.581 token. Its normalized, annotated version is publicly available at https://github.com/christoph-hesse/question-under-discussion, with 2034 QUDs, 2718 annotated focus and 232 non-at-issue segments. We chose driving reports since these texts are combinations of factual information with subjective estimations of the author concerning the driving experience and the quality and functionality of the equipment. As a result, these texts contain

syntactically complex sentences and are highly pragmatically loaded, exhibiting complex, interleaved information structures. Although the number of texts is rather small, their annotation was an extensive action. The annotators, employed as student assistants and paid accordingly, have intensively been trained and they were familiarized with the annotation tool we developed for this purpose. Each text was annotated twice as a whole independently.

An example illustrates the combination of factual with evaluative information:

(1) *Der wuchtige Bentley Flying Spur ist kein Monument der Beharrung, sondern die schnellste Limousine der Welt. Zum 100. Geburtstag spendiert sich Bentley die Neuauflage einer Ikone.* ('The massive Bentley Flying Spur is not a monument to perseverance, but the fastest sedan in the world. For its 100th birthday, Bentley is giving itself a new edition of an icon.')

The annotation comprises a statement of the respective QUD, the focused constituent, the topic, and non-at-issue content. One of the annotations for our example is as follows:

<QUD string="What about the Bentley Flying Spur?" >
  <CON>
    <SEGMENT>Der</SEGMENT>
    <NAI>
      <SEGMENT>wuchtige</SEGMENT>
    </NAI>
    <SEGMENT>Bentley Flying Spur</SEGMENT>
  </CON>
  <F> <SEGMENT>ist kein Monument
        der Beharrung,</SEGMENT> </F>
  <F> <SEGMENT>sondern die schnellste Limousine
        der Welt.</SEGMENT> </F>
</QUD>
<QUD string="What about Bentley's 100th birthday?"> …

The text is separated into segments that will be tagged as focus (F), aboutness topic (CON), and non-at-issue (NAI). The QUD string must be freely formulated by the annotator, since formulating a predefined tagset of QUDs is not possible.

Our annotation guideline follows the guideline of (Riester et al., 2018), but with the important modification of allowing left-branching nodes. Although new text can only attach at the right frontier, the trees that result can nevertheless be branching left or right. Left branching can happen if new text segments attach high up on the right frontier. The car reviews contain text structures which would be reflected by left branching structures in the QUD tree, which was taken into account in our guidelines.

Riester et al. (2018) formulate a number of constraints which allow the annotator to derive QUDs from the previous or upcoming discourse context in a bottom-up way. First, the single sentences will be disassembled into constituents that express a single proposition.

Then, three constraints must be satisfied when formulating the QUD:

- QUDs must be answerable by the proposition(s) that they immediately dominate.

- QUDs make reference to the immediate preceding discourse, i.e. they consist of given/salient material.

- Therefore, QUDs must be located as high in the QUD tree as possible.

Contrary to this guideline, we permitted QUD annotations in a top-down and bottom-up manner in order to consider discourse-oriented and propositional QUDs. Whether the annotators started with the propositions as leaf notes, formulated corresponding QUDs for them and proceeded with more complex sections until they arrived at the top and most general QUD, or whether they started with the topmost QUD and tried to refine it recursively through sub-QUDs, had interesting consequences for form and content of these formulated QUDs. Annotators – in their attempt to bridge the gap between top-down and bottom-up approaches to discourse structure – naturally arrived at intermediate QUDs capturing authors' line of argument. Even though we gave annotators full freedom in formulating QUDs, an analysis of *wh*-words in these argumentative QUDs (see Table 1) shows that annotators gravitated towards a limited set of *wh*-words which mirror established discourse relations (for instance, used in RST). RST-relations are not our main interest in this paper. Nevertheless, Table 1 suggests that they are implicitly given by the formulation of QUDs.

Explanation/Reason	#
Warum/Wieso (Why)	55
Was ist der Grund (What is the reason)	12
Aus welchem Grund (For what reason)	19
Woran liegt das (Why is that)	1
Purpose/Goal	
Was …aus diesem Grund (What came of …)	1
Wozu (To what end)	3
Elaboration/Clarification	
Was/Worum handelt es sich (What/is it about)	1
Manner	
Wie (How)	1

Table 1: Wh-questions annotated in argumentative QUDs in the QUDGen Corpus and their matching discourse relations.

Driving reports' genre-specific blend of factual information content and evaluative content is reflected in QUD annotation by the fact that there are argumentative QUDs (aQUDs) and fact-based QUDs (fQUDs) in the tree structures (see the following example). We see a general tendency in the corpus for QUDs lower in the trees to be more concrete and fact-oriented (e.g.,

fQUD: 'What is the battery capacity?') and QUDs higher in the trees to be more abstract and argumentative (e.g., aQUD: 'Why will driving it be limited to urban areas?'), but we also find factual QUDs being follow-up by subordinated argumentative QUDs (e.g., the aQUD 'What's advantageous about urban driving for the battery?' subordinated under the fQUD 'What urban range is possible?').

aQUD Where can you drive the motorcycle?
A Das Fahren selbst wird sich vorwiegend auf den urbanen Raum beschränken. ('Driving itself will be limited to urban areas.')
   aQUD Why will driving it be limited to urban areas?
    fQUD What about range?
     fQUD What is the battery capacity?
     A Mit dem 15,5-kWh-Lithium-Ionen-Akku ('With the 15.5 kWh lithium-ion battery')
      fQUD What mixed range is possible?
      A sind laut Harley im gemischten Betrieb Reichweiten von gut 150 Kilometer zu erzielen, ('150 kilometers are possible in mixed driving, according to Harley,')
      fQUD What urban range is possible?
       aQUD What's advantageous about urban driving for the battery?
       A im reinen Stadtbetrieb, wo im Stop-and-Go viel rekuperiert werden kann, ('in urban areas, where stop-and-go allows for lots of recuperation,')
        fQUD What range is possible with recuperation?
        A immerhin zirka 230. ('at least roughly 230 [kilometers].')

QUDs are almost exclusively formulated as *wh*-questions. The type of *wh*-pronoun constrains the material that can be annotated as focus.

Focus is a signal of the existence of alternatives from which the linguistic constituent selects the relevant one. The alternative set coincides with the *wh*-pronoun in the QUD. Given this link to alternatives, a focus often is characterized as that constituent that expresses information that is new to the addressee.

The topic relates to the QUD; it is the referent the QUD is about. Information is not-at-issue if it does not address the QUD.

The inter-annotator agreement concerning focus, topic and non-at-issue content (Fleiss' κ and Krippendorf's α) is given in Table 2.

	κ	α
Focus	0.40	0.70
Topic	0.30	0.65
Non-at-issue	0.25	0.62

Table 2: Inter-annotator agreement for focus structure, topic structure, not-at-issue information.

When comparing these coefficients with other annotation studies on information structure, the syntactic complexity of the single sentences is a relevant factor. For example, De Kuthy et al. (2015) annotated focus in German question-answer pairs of the CREG-1032 learner corpus and the QUIS corpus, which contains Q/A data that have been elicited in a controlled way. Both corpora are characterized by a significantly lower syntactic complexity of the single sentences than the ones in our corpus. The values for focus annotation are $\kappa = 0.75$ (CREG) and $\kappa = 0.87$ (QUIS), respectively. In their annotation study, Ritz et al. (2008) show the variance of annotation results for topic and focus annotations. The data are Q/A pairs, 2 dialogues, and especially texts from the Potsdam Commentary Corpus (PCC) that differ from the other texts in that they are syntactically more complex. The κ values for topic and focus are for the Q/A pairs 0.75 and 0.51, for the dialogues 0.51 and 0.44, and 0.44 and 0.19 for the PCC data.

We are not aware of any annotation studies on (non) at-issue content. The mentioned values for topic annotations do not reflect different topic types. However, in particular the focus values show that annotations do not significantly differ from other annotations of information structures.

Since there is no predefined set of QUDs but the annotators had to formulate them freely, we normalized the resulting QUDs by discussing the results with the annotators and deciding on a version that corresponds to the intention of both annotators.

2.1. The Annotation Tool QUDA

In order to facilitate the annotation process, we developed the QUDA annotation tool (Langner, M. and Klabunde, R. and Benz, A., 2021) that is tailored to the specific demands in annotating QUDs and information structural tiers.

QUDA enables the simultaneous annotation of QUDs, focus, topic, and at-issueness. It validates discourse trees according to the annotation guidelines. As a web application, it is publicly available (see: https://github.com/MMLangner/QUDA). Server-side cookies prevent data loss, a graphical interface supports the annotation process. Annotations can be exported as XML markup, the XML files can be imported again into the viewer. The tool can be adjusted to new requirements and annotation guidelines quite fast.

In detail, the tool supports tree construction methods like adding QUDs and text segments in both, a right-branching and a left-branching manner. It also allows to edit trees, e.g. shifting linear precedence, copy-pasting subtrees to other nodes and deleting elements. Furthermore, the tool offers a split window, where either an annotation and a plain text document can be rendered simultaneously, or two annotations for direct comparison. When comparing two annotations, the

tool also provides the function to calculate the inter-annotator agreement between the two files.

3. Case Studies: Evaluating Focus and Non-At-Issue Annotations

The IAA coefficients indicate the notorious problem of comparing discourse-oriented annotations. Nevertheless, the annotations allow us to gain insight into the adequacy of pragmatic theories. We demonstrate the usefulness of the corpus by an analysis of the focus annotations, especially the focus-sensitive use of the German adverb *überhaupt*, and the relation between non-at-issueness and non-restrictive relative clauses.

3.1. Focus

According to the standard view on focus, a focus indicates the presence of alternatives that are relevant for the interpretation of a linguistic expression (Krifka, 2008; Rooth, 1985). Different types of focus have been proposed in semantic analyses in order to point out the specific alternatives for focus interpretation.

Since the focus in an answer corresponds to the *wh*-element in a constituent question, one would expect that the number of *wh*-elements in a formulated QUD corresponds to the number of foci expressed. However, the annotations show that this assumption does not hold in general. Two randomly chosen driving reports (Audi SQ5 and Ford Range Raptor) show that the relation between QUD and focus is not as straightforward (Table 3).

Report	# QUDs	# Foci	QUD/F	# Split Focus
Audi	72	91	0.79	29
Ford	87	104	0.83	19

Table 3: QUD-structure and focus structure of two driving reports from the QUDGen Corpus.

About every fifth QUD has been associated with more than one focus. In this case, two foci have been assigned with a connective element between them. This seems to contradict the assumption on focus assignment that a question induces one focus domain. However, the items between two split foci are either simple or complex conjunctions with no QUD-sensitive meaning, or information that is given by the QUDs (Table 4). Hence, split foci are perfectly in line with the assumptions made in the literature.

3.1.1. Focus and *überhaupt*

The German adverb *überhaupt* ('at all', 'in general') is not considered a focus particle, but some of its uses have a focus-sensitive meaning (König, 1983). Our corpus contains four occurrences of this adverb so that we checked whether and how the focus annotations reflect the statements on focus-related meaning made in the literature on formal pragmatics, especially (Rojas-Esponda, 2014) and (Anderssen, 2006).

Item	#
und ('and')	23
Strings from QUD	7
Pronoun with reference to item in QUD	4
oder ('or')	2
und auch ('and also')	2
other conjunctive elements	9
NAI between two foci	1

Table 4: Connecting elements between two foci.

Rojas-Esponda (2014) shows that a comprehensive analysis of the discourse functions associated with this adverb can be given by reference to superordinated QUDs in QUD tree structures. Our corpus gives hints on the kind of reference of this adverb to QUDs. According to (König, 1983), *überhaupt* has a focus-related and a non-focus meaning. The former is signaled by bearing the sentential accent. König proposes four usage classes. Class A comprises uses of *überhaupt* in declaratives that express a generalization. In such a context the adverb bears the sentential accent and is, therefore, focused: *Der BMW hat einen luxuriös ausgestatteten Frontbereich. Er ist überhaupt luxuriös ausgestattet.* ('The BMW has a luxuriously equipped front area. It is luxuriously equipped at all.').

Class B comprises uses of focused *überhaupt* that contribute an existential statement: *Ich bin überrascht, dass er überhaupt etwas gesagt hat* (*I'm surprised he said anything at all*). Class-B uses are confined to downward-entailing contexts, i.e. contexts that license an entailment from superordinated to subordinated concepts (the analysis of *überhaupt* by (Anderssen, 2006) is based on the same observation). For example, the verbal form *überrascht sein* ('being surprised') provides such a context, as the following example demonstrates:

(2) a. *Ich bin überrascht, dass er über sein Studium geredet hat*. (I'm surprised he talked about his studies)

   b. *Ich bin überrascht, dass er etwas gesagt hat*. (I'm surprised he said something/anything)

Sentence 2b. entails 2a., but not the other way round. The key point is that *überhaupt* is compatible with negative polarity contexts: *Ich bin überrascht, dass er überhaupt etwas gesagt hat*.

The third class has been labeled as presuppositional by König. It comprises negative sentences where a more general presupposed property has been rejected: (Ed had an accident with his car yesterday) – *Ed besitzt überhaupt kein Auto* ('Ed doesn't have a car at all/doesn't even own a car').

Finally, class D is confined to informal conversations where *überhaupt* expresses a 'sudden inspiration': *Ach überhaupt, gestern war ich beim Frisör* ('Oh anyway, yesterday I went to the hairdresser').

Given this classification, our corpus contains three uses that belong to class A and one instance of class C. The question is, how do the formulated QUDs license the use of this adverb?

Vorne gibt es überhaupt keine Beschwerden ('in the front there are no complaints at all') belongs to class C: the adverb is not focused, it appears in a negative context, and it expresses that the property of causing a complaint has been rejected. Hence, we would expect that the adverb should not belong to the annotated focus. However, the QUD formulated by the annotators addresses a general inquiry ('What about the tall passengers?'), which has been answered by a coordinated sentence with the second clause being the *überhaupt* example: *Selbst Zwei-Meter-Mannen sitzen im Fond ganz ordentlich und vorne gibt es überhaupt keine Beschwerden* ('Even two-meter men sit quite neatly in the rear and in the front there are no complaints at all'). Both clauses are annotated as constituting the focus, fulfilling the information needs, but the conjunction *und* ('and') has been excluded. The annotation is perfectly in line with the annotation guidelines, indicating that the constituents that answer the QUD are focused. Since the formulated QUD is quite general, the *überhaupt*-clause with its rejected property fits perfectly to the QUD. The annotation is also in line with the QUD-oriented analysis by (Rojas-Esponda, 2014). Now let us have a look at the three examples belonging to class A:

(3) a. *überhaupt geht es im Ioniq progressiv zu* ('generally the Ioniq is very progressive')

   b. *überhaupt wirkt der Raptor im Innenraum wie ein SUV* ('generally the Raptor looks like an SUV on the inside')

   c. *Trotzdem sind die Preise erreichbarer, soweit sich das in diesen (Preis)Klassen überhaupt behaupten lässt.* ('Nevertheless these prices are achievable, if you can even say that in this price range.')

The third example seems to contain a non-focused *überhaupt*, but the central requirement of expressing a generalization is met.

For these data, it is useful to take the classification of (Rojas-Esponda, 2014) into account. Rojas-Esponda proposes five types of *überhaupt*, two for unfocused uses and three for focused *überhaupt*. An additional type in her classification is the relevant class for these three examples. This type is characterized by the use of focused *überhaupt* with a universal quantifier or a scalar predicate. The first two examples include a scalar predicate (being progressive and appearing like something, respectively), but the adverb has been excluded from the focus by the annotators, so that the annotated data partially fit into the proposed classification.

Rojas-Espondas explains the various uses of *überhaupt*

by the attempt of the speaker to resolve all sub-QUDs of a presumed superordinated QUD by generalizing over the current propositional content. Hence, in our annotated data, we would expect that *überhaupt* belongs to a focus, and its function is to signal the answer to a more general QUD than the current one. The latter is exactly what the annotators realized: they formulated a general QUD (*Was ist mit dem Fahrgefühl/Fahrerlebnis?* and *Was ist mit dem Innenraum?*; 'What about the driving feeling/interior') and a direct subordinated QUD. The *überhaupt* clause seems to address this superordinated QUD. So the annotation is basically in line with the formal analysis of *überhaupt* given in the literature, minus the focus assignment.

The final example seems to fall out of the classification, since *überhaupt* belongs to the focus, but no scalar predicate has been given in the subordinate clause with this adverb. However, the annotators decided to formulate a separate QUD for the subordinate clause that asks for the constraints concerning the statement in the main clause. The subordinate clause provides this constraint as a general statement on the usefulness of the statement made in the main clause. Hence, the generalization function of *überhaupt* is satisfied.

To summarize, the annotations are basically in line with the pragmatic analyses of the discourse functions associated with *überhaupt* given in the literature, implying the annotation guidelines concerning QUD formulation and focus marking are sufficiently precise. The ignorance of focused *überhaupt* in some of the annotations is probably due to the absence of clear phonological cues in the complex sentences concerning sentence accent.

3.2. Non-At-Issueness and Non-Restrictive Relative Clauses

The standard view on relative clauses distinguishes two kinds: (i) restrictive relative clauses such as *Jane reached for the plates which are on the top shelf* where the relative clause restricts which plates we are talking about, and (ii) non-restrictive relative clauses such as *Jack followed Edna, who is a fearless leader*, where the relative clause gives further information about Edna but does not restrict which Edna we are talking about.

More recently a distinction is made between at-issue content (the "main point" of a proposition) and not-at-issue content, which is "secondary", in some sense, to that main point or backgrounded relative to it. There is no commonly accepted definition of at-issueness yet. Instead, approaches to at-issueness generally fall into two camps: (a) grammatical approaches (Murray, 2010; Murray, 2014; AnderBois et al., 2010; Potts, 2005; Koev, 2013; Koev, 2018) which assume that information status is marked in the grammar through discourse markers, lexical choice, and marked syntax, and (b) QUD approaches (Amaral et al., 2007; Roberts et al., 2009; Simons et al., 2010; Beaver et al., 2017),

where information status is assumed to be determined in terms of relevance to the central discourse topic, i.e. the question under discussion (an interrogative phrasing of the "main point").

Important for us here is that it has been proposed by Potts (2005) and others that the at-issueness distinction maps onto the restrictiveness distinction of relative clauses—at least in some respect—with restrictive relative clauses being at-issue and non-restrictive relative clauses being potentially not-at-issue. Potts (2005) assumes that all non-restrictive relative clauses are not-at-issue and classes them with appositions such as *Edna, a fearless leader, started the decent* and interjections as supplementary material with the syntactic status of sentence adjuncts and semantic projection beyond the scope of any operator. More recently, Koev (2013) points out that non-restrictive relative clauses can sometimes be at-issue when they appear sentence-final and are accessible to anaphoric reference. Conversely, Koev (2013) argues that sentence-medial non-restrictives are always not-at-issue.

QUD annotation lends itself naturally to investigating the connection between at-issueness and restrictiveness. In order to annotate which information content of a proposition is at-issue relative to its QUD and which content is not, we introduce a NAI tag to mark text spans containing not-at-issue content and, by definition, any text span not marked by a NAI tag is assumed to be at-issue (see, e.g., the XML code snippet in Section 2). Since our XML annotation uses QUD tree structures following (Kuthy et al., 2018; Riester et al., 2018; Riester, 2019; Büring, 2003), relative clauses can be annotated with their own sub-QUD. For instance, in *Jack follows Edna, who is a fearless leader* the non-restrictive relative clause could have a sub-QUD 'What kind of a leader is Edna?' or 'Why is Jack following Edna?,' which is subordinate to a QUD for the main clause, such as 'What is Jack doing?' In QUD trees, subordination of QUDs is equivalent to sub-QUDs being secondary to super-QUDs, similar to how not-at-issue content is assumed to be secondary to at-issue content. In the QUDGen Corpus, having the NAI tag and subordinated QUDs allows us to look at the information status of non-restrictive relative clauses from the QUD perspective. Annotators applied the at-issueness definition by Simons et al. (2010). According to this definition a proposition is at-issue iff the speaker intends it to be relevant to the QUD. It is relevant to the QUD iff it either contextually entails a partial or complete answer to the QUD or contextually entails another QUD which has a relevant answer. In order to look at the at-issueness of non-restrictives from the grammatical perspective, we introduce a tag for discourse markers, and we record the position of relative clauses within sentences and the syntactic function of the constituent modified by the relative clause.

We analyze the corpus under the hypothesis that both, the grammatical approach and the QUD approach, predict the same information status for non-restrictive relative clauses. Corpus results (Table 5), however, are in

	Info Status		Sent Position	
RC Type	# NAI	# AI	% Mid	% End
Restrictive	1	17	4/17	13/17
Non-restr.	1	52	9/52	43/52

Table 5: Information status (at-issue/not-at-issue) and sentential position (middle/end) of restrictive/non-restrictive relative clauses in the QUDGen Corpus.

stark contrast to theoretical expectations: (i) Contrary to both theoretical views, restrictive relative clauses are not more likely to be at-issue than non-restrictive relative clauses. In fact, we find the opposite: there are more at-issue non-restrictive relative clauses (52 times) than there are at-issue restrictive relative clauses (17 times). (ii) Relative clauses are predominantly at-issue, regardless of whether they are restrictive or non-restrictive (ratio of not-at-issue:at-issue is 1:17 and 1:52 respectively). (iii) While at-issue non-restrictives have a tendency to appear sentence-final (43 out of 52 times), sentence-medial non-restrictives are not—as the grammatical view predicts—always not-at-issue. Indeed, there are a number of corpus examples of sentence-medial at-issue non-restrictives (9 times) such as:

(4) *Zwar fehlt es dem 400 PS starken Antrieb nicht an Ambitionen und mit Reifen, die wie Pattex am Asphalt kleben, erwischt dich der Vortrieb so unvermittelt wie die Faust eines Preisboxers.* ('Though the 400 horsepower engine does not lack ambition and with tyres, which feel like they are clued to the asphalt, the acceleration hits abruptly like a prize fighter's fists.')

where the relative clause is non-restrictive because it does not differentiate this set of tyres with good grip from alternative sets. At the same time, the non-restrictive relative clause is relevant to the QUD of the matrix clause '*Was ist gut am Kickdown?*' (What is good about the kickdown?), which names two advantages of the car, namely (i) the 400 horsepower engine and (ii) the tyres with good grip. At first glance the relative clause seems to give a reason *why* the tyres are good, which could be analyzed through a sub-QUD. However, notice that the author also could have used the adjective *good* to describe the tyres (or even stronger superlatives such as *excellent*) and world knowledge would tell us that *good tyres* in consumer-level cars usually means that the tyres have good grip— which is exactly the idiomatic reading of the expression *wie Pattex am Asphalt kleben* (Pattex, a brand of glue, 'glued to the asphalt') in the relative clause. Example (4) is an example of a non-restrictive supplying an answer to the QUD of the matrix clause which

is contextually entailed. In other cases we find sub-QUDs for the non-restrictives with an answer relevant to the QUD of the matrix clause. We find a clear trend towards the discourse relation between non-restrictive and matrix clause being either Explanation or Evidence; two relations with high relevance, and two relations whose QUDs we characterized as argumentative rather than fact-oriented (see Table 1 for *wh*-words). Contrary to the grammatical view we thus conclude that (i) sentence-medial non-restrictive relative clauses can be at-issue (in the sense of (Simons et al., 2010)), and (ii) when they are at-issue it is because they supply an explanation for an argument made in the main clause or supplies evidence for an evaluation made in the main clause. The grammatical view would identify sentence-medial non-restrictives as giving secondary information to the matrix clause and would thus, contrary to the Simons et al. (2010) definition, identify them as not-at-issue. The corpus can be used to draw out this disagreement between the two theoretical views, and gives new insights into the relevance requirement. At the same time, the corpus contains new and contradicting evidence against both theoretical views in that non-restrictives are more likely to be at-issue than restrictive relative clauses.

To summarize, corpus results as to the at-issueness of non-restrictive relative clauses are more in line with the theoretical QUD view than with the grammatical view in the literature, implying that hierarchical QUD trees with NAI tags and subordinating QUD structures characterize the information status of non-restrictive relative clauses. The corpus is able to draw out disagreement between the two theoretical views on non-restrictive relative clauses on the basis of a number of key factors.

4. Conclusion

We present a novel, XML based, corpus of German test drive reports for the analysis of QUD-oriented information structural distinctions called the QUDGen Corpus, which uses hierarchical QUD trees to represent discourse goals, discourse relations, and information structural units, where text chunks serve as answers to these QUDs. In order to aid the annotation process the corpus comes with an annotation tool called QUDA. We developed the corpus with the goal of testing the adequacy of QUD-based pragmatic frameworks of information structure. Currently in the literature, QUD frameworks are used as an analytical tool.

We look at the annotated focus structures to assess the quality of the QUDGen corpus. First analyses show that with respect to focus-related meaning aspects our corpus is in line with current theory, indicating sufficient accuracy of annotations.

Apart from focus structure – with the use of *überhaupt* as sample analysis –, we show how the QUDGen Corpus can be used to investigate theoretical debates such as the not-at-issue status of non-restrictive rela-

tive clauses. To our knowledge the QUDGen Corpus is the first attempt to annotate at-issue and non-at-issue content. Using information-status-specific tags and the hierarchical QUD tree structure, we show that the corpus is able to draw out disagreement in the literature about the information status of non-restrictive relative clauses, and that the corpus analysis is able to provide new insight into the interaction of various factors including position within the matrix clause and the relative's discourse relation to the matrix clause. We hope the QUDGen Corpus and its accompanying annotation tool QUDA will be useful to researchers interested in discourse structure, information structural units, information status, and Question-under-Discussion.

5. Acknowledgements

6. Bibliographical References

Amaral, P., Roberts, C., and Smith, E. A. (2007). Review of the logic of conventional implicatures by chris potts. *Linguistics & Philosophy*, 30:707–749.

Anand, P. and Matell, C. (2012). Annotating the focus of negation in terms of questions under discussion. In *Proceedings of the ACL-2012 Workshop on Extra-Propositional Aspects of Meaning in Computational Linguists*, pages 65–69.

AnderBois, S., Brasoveanu, A., and Henderson, R. (2010). Crossing the appositive/at-issue meaning boundary. *Semantics and Linguistic Theory*, 20:328–346.

Anderssen, J. (2006). Generalized domain widening überhaupt. In *West Coast Conference on Formal Linguistics (WCCFL) 25*, pages 58–66.

Beaver, D., Roberts, C., Simons, M., and Tonhauser, J. (2017). Questions under discussion: Where information structure meets projective content. *Annual Review of Linguistics*, 3:265–284.

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

Carroll, M., von Stutterheim, C., and Klein, W. (2003). Two ways of construing complex temporal structures. In F. Lenz, editor, *Deictic conceptualisation of time, space and person*, pages 97–134. Benjamins, Amsterdam.

De Kuthy, K., Ziai, R., and Meurers, D. (2015). Learning what the crowd can do: A case study on focus annotation. 01.

Ginzburg, J. (2012). *The Interactive Stance*. Oxford University Press, Oxford.

Kehler, A. (2015). On qud-based licensing of strict and sloppy ambiguities. In *Proceedings of SALT 25*, pages 512–532.

Koev, T. K. (2013). *Apposition and the Structure of Discourse*. Ph.d. dissertation, Rutgers University.

Koev, T. (2018). Notions of at-issueness. *Language and Linguistics Compass*, pages 1–16.

König, E. (1983). Polysemie, Polaritätskontexte und *überhaupt*. In H. et al. Weydt, editor, *Partikeln und Interaktion*, pages 160–171. Niemeyer, Tübingen.

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

Kuthy, K. D., Reiter, N., and Riester, A. (2018). Qud-based annotation of discourse structure and information structure: Tool and evaluation. In *Proceedings of the 11th Language Resources and Evaluation Conference (LREC)*.

Murray, S. (2010). *Evidentiality and the Structure of Speech Acts*. Ph.d. dissertation, Rutgers University.

Murray, S. (2014). Varieties of update. *Semantics & Pragmatics*, 7:1–53. DOI: 10.3765/sp.7.2.

Potts, C. (2005). *The Logic of Conventional Implicatures*. Oxford University Press, Oxford.

Riester, A., Brunetti, L., and Kuthy, K. D. (2018). Annotation guidelines for questions under discussion and information structure. In Evangelia Adamou, et al., editors, *Information Structure in Lesser-Described Languages: Studies in Prosody and Syntax*, pages 403–443. Benjamins, Amsterdam.

Riester, A. (2019). Constructing qud trees. In Malte Zimmermann, et al., editors, *Questions in Discourse: Pragmatics*, volume 2, pages 403–443. 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 6th International Conference on Language Resources and Evaluation*, pages 2137–2142.

Roberts, C., Simons, M., Beaver, D., and Tonhauser, J. (2009). Pre-supposition, conventional implicature, and beyond: A uniform account of projection. In *Proceedings of the Workshop on New Directions in the Theory of Presupposition, ESSLLI 2009*.

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

Rojas-Esponda, T. (2014). A discourse model for *überhaupt*. *Semantics and Pragmatics*, 7(1):1–45, March.

Rooth, M. (1985). *Association with Focus*. Ph.d. dissertation, University of Massachusetts.

Simons, M., Tonhauser, J., Beaver, D., and Roberts, C. (2010). What projects and why. *Semantics and Linguistic Theory*, 20:309–327.

van Kuppevelt, J. (1995). Discourse structure, topicality and questioning. *Journal of Linguistics*, 31:109–147.

von Stutterheim, C. and Klein, W. (1989). Referential movement in descriptive and narrative discourse. In R. Dietrich et al., editors, *Language Processing in Social Context*, pages 39–76. Elsevier Science Publishers, North-Holland.

7. Language Resource References

Hesse, C. and Klabunde, R. and Benz, A. (2021). *QUDGen Corpus*. QUDGen Project, available at https://github.com/christoph-hesse/question-under-discussion.

Langner, M. and Klabunde, R. and Benz, A. (2021). *QUDA Annotation Tool*. QUDGen Project, available at https://github.com/MMLangner/QUDA.