

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

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

*Leibniz-Center General Linguistics Berlin, †Ruhr-University Bochum
{hesse, benz}@leibniz-zas.de, {maurice.langner, ralf.klabunde}@rub.de



INTRODUCTION

Goal of the project: Test Question-under-Discussion (QUD) frameworks by means of a QUD-annotated corpus

QUD-theories view texts as complex answers to implicit questions (QUDs)

Text genre of car & motorcycle reviews poses unique challenges with their blend of **propositional content** (technical specs) & **expressive content** (evaluative statements)

THE QUDGEN CORPUS

- Novel, XML-based corpus of German car & motorcycle reviews (🔗/christoph-hesse/question-under-discussion) annotated with Questions-under-Discussion discourse trees
- First attempt to systematically annotate at-issue/non-at-issue content
- Annotation guidelines extend on previous work (Kuthy et al., 2018; Riester et al., 2018; Riester, 2019; Büring, 2003), new: left- & right-branching discourse trees

Corpus example:

(1) *The massive Bentley Flying Spur is not a monument to perseverance, but the fastest limousine in the world.*

```
<QUD string="What about the Bentley Flying Spur?" >
<CON>
  <SEG>The</SEG>
  <NAI> <SEG>massive</SEG> </NAI>
  <SEG>Bentley Flying Spur</SEG>
</CON>
<F> <SEG>is not a monument
      to perseverance,</SEG> </F>
<F> <SEG>but the fastest limousine
      in the world.</SEG> </F> </QUD>
```

IAA (Fleiss' κ and Krippendorff's α):

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

IAA coefficients for focus & topic comparable to other studies (DeKuthy et al., 2015; Ritz et al., 2008) when text complexity is taken into account; no studies with IAA coefficients for non-at-issue

QUADA ANNOTATION TOOL

- Simultaneous annotation of QUDs, focus, topic, and at-issueness
- Validates discourse trees & calculates inter-annotator agreement
- Web app (🔗/MMLangner/QUADA)
- Graphic UI, XML import/export
- Functions: create QUDs, segmentation, information tagging, left- and right-branching, move subtrees, split-window editing

CASE STUDIES: FOCUS AND NON-AT-ISSUE ANNOTATION

We show the usefulness of the QUDGen Corpus by testing focus and non-at-issue frameworks

The following 2 case studies center around the blend of propositional and expressive content found in the genre of vehicle reviews:

- Focus structure as a proxy for propositional content (one QUD per proposition)
- Expressive content is typically analysed as non-at-issue content

FOCUS

Theory: one QUD per focus constituent

Our corpus: no one-to-one relation

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

Case study: the German focus marker überhaupt: 2 classifications (König, 1983 & Rojas-Esponda, 2014)

4 uses of *überhaupt* (König, 1983): (i) generalisation (cf. 2a-c), (ii) existential statement in downward-entailing context (cf. 2b), (iii) presuppositional with negation (cf. 2d), (iv) sudden inspiration (cf. English *by the way*)

5 uses of *überhaupt* (Rojas-Esponda, 2014): 2 for unfocused uses, 3 for focused uses (2a-c)

- überhaupt geht es im Ioniq progressiv zu* ('generally the Ioniq is very progressive')
- überhaupt wirkt der Raptor im Innenraum wie ein SUV* ('generally the Raptor looks like an SUV on the inside')
- 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.')
- Vorne gibt es überhaupt keine Beschwerden* ('in the front there are no complaints at all')

Our corpus: *überhaupt* belongs to focus + signals an answer to a super-QUD → in-line with Rojas-Esponda (2014)

NON-AT-ISSUE

Information status: distinction between at-issue ('the main point') and non-at-issue content (in some sense, 'secondary')

Case study: restrictive (RRC) and non-restrictive relative clauses (NRRC)

Potts (2005): RRC are at-issue, NRRC at not at-issue

Koev (2013): sentence-final NRRC can be at-issue, sentence-medial NRRC are **always** not at-issue

Roberts et al. (2009), Simons et al. (2010): NRRC at-issue when relevant to QUD of their matrix clause

(3) *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 glued to the asphalt, the acceleration hits abruptly like a prize fighter's fists.')

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

Our corpus: NRRC can be at-issue (contra Potts, 2005), sentence-medial NRRC can be at issue (contra Koev, 2013) → in-line with Roberts et al. (2009) & Simons et al. (2010)

CONCLUSION

- Novel, XML-based corpus for QUD analysis called QUDGen
- Hierarchical QUD trees to represent (1) discourse goals, (2) discourse relations, and (3) information-structural units
- QUADA annotation tool
- Accuracy of annotation ensured through assessment of focus structure
- QUDGen is the first corpus to annotate at-issue and non-at-issue content

REFERENCES

Büring, D. (2003). On d-trees, beans, and b-acents. *Linguistics & Philosophy*, 26:511–545. Koev, T. K. (2013). *Apposition and the Structure of Discourse*. Ph.d. dissertation, Rutgers University. König, E. (1983). Polysemie, Polaritätskontexte und überhaupt. In H. et al. Weydt, editor, *Partikeln und Interaktion*, pages 160–171. Niemeyer, Tübingen. 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)*. 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. 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, ESSLI 2009*. Rojas-Esponda, T. (2014). A discourse model for überhaupt. *Semantics and Pragmatics*, 7(1):1–45, March. Simons, M., Tonhauser, J., Beaver, D., and Roberts, C. (2010). What projects and why. *Semantics and Linguistic Theory*, 20:309–327.