

Notes for the QUD-Anno Challenge

Arndt Riester (Bielefeld University)

1 General issues

Information-structural and QUD annotation largely follows the suggestions made in Riester et al. 2018 and De Kuthy et al. 2018. Sentence-initial and sentence-final adjuncts are treated as separate discourse segments, as suggested in Brunetti et al. 2021. The following list of annotation labels is used (Table 1).

BG	background
CT	contrastive topic
DM	discourse marker / discourse connector (<i>and, as, but, because, or, then, ...</i>)
DP	discourse particle, here: focus-sensitive particle (<i>only, even, ...</i>)
F	focus
NAI	non-at-issue material
T	sentence topic (backgrounded referential expression)

Table 1: Annotation labels

2 Issues to be discussed

2.1 Cataphoric questions

Cataphoric questions (cf. Brunetti et al. 2021, Riester et al. 2021) are questions about *non-at-issue* material at the beginning of an utterance, usually in the form of adjuncts or adjunct clauses. They have the following general structure in (1), where the cataphoric question is Q_1 , because it assumes the givenness of material that follows in A_0 . Like in the case of cataphoric pronouns, the GIVENNESS principle is not actually violated (which would mean that discourse-new material is marked as given), but merely suspended for a short time.

- (1) Q_0 : ...
 > Q_1 : ...
 >> A_1 : ...
 > A_0 : ...

Note that in RST (Mann & Thompson 1988, Taboada & Mann 2006), such structures would typically correspond to mono-nuclear discourse relations in which the Satellite precedes its Nucleus. Examples are shown in (2) and (3).

- (2) [Narrative:]
 Q₆: { *What does Susie do next?* }
 > Q₇: { *Susie rubs [her head] while she does what?* }
 >> A₇: [As]_{DM} [she]_T [checks Mrs Simpson's calendar,]_F (CIRCUMSTANCE)
 > A₆: [Susie]_T [rubs the place where the elastic cap from work scrunched all day]_F
- (3) [Interview:]
 Q₈: { *What is the case, despite the Pro-Brexit vote?* }
 > Q₉: { *Are there people who don't think that was the right way to go?* }
 >> A₉: [And,]_{DM} [yes,]_F (EVALUATION)
 > A₈: [there are many people [. . .] who don't think that was the right way to go]_F.

2.2 Keeping referents accessible/active

One particular difficulty of the narrative text is that certain entities (*the tea, Susie's mother, the factory* etc.) are mentioned briefly at the beginning of the story, then ignored for a long stretch of text, then suddenly re-appear (see Figure 1). For instance, there are 76 discourse units between the first and the second mention of *her mother*. From a discourse-structural perspective, this raises the question when, and according to what criterion, an entity should be considered as *no longer active*, i.e., no longer in the right frontier (Polanyi 1988, Asher & Lascarides 2003) of the current discourse, which would result in the closing of a discourse branch and subsequent attachment of the following discourse segment at a higher node.

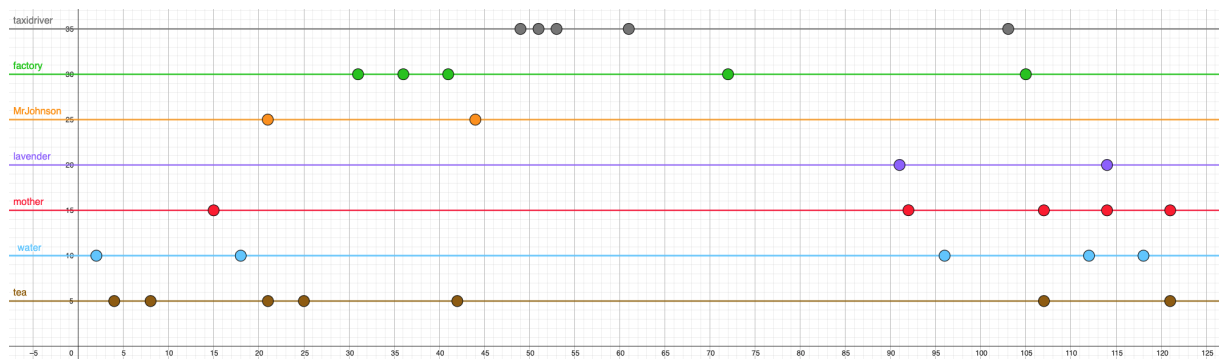


Figure 1: Some entities mentioned in the narrative text, spread over 125 discourse units

In a default, coherent informative text (e.g. an encyclopaedia article), the reader usually has a reasonable sense of when a sub-topic is finished (or, in QUD terms, Roberts 2012 [1996], when a subquestion is sufficiently answered and removed from the stack). In literary text, by contrast, subverting such expectations can be an intentional stylistic device, reminding the reader of the unexpected importance of a certain previous entity. (By contrast, *Mr Johnson*¹ is mentioned twice, in units 21 and 44, then no longer plays a role in the story.)

From the perspective of QUD-tree construction, there is, so far, no clear solution for the question when entities become inactive. The problem is that, if all entities are generally kept *active* by default, this means that attachment of new segments always has

¹Not Boris Johnson

2.3 Not answering an overt question

(4) [Interview:]
 LK: > A₂₄: [you're]_T [trying to create a situation of them and us,]_F
 > Q₂₅: **are you not?**
 BJ: > Q₂₆: { *What about getting Brexit done?* }
 >> Q_{26.1}: { *There is no way of getting Brexit done without what?* }
 >>> A_{26.1}: [I think that the truth is,]_{NAI} there's [no]_{CT} [way of getting Brexit
 done]_{BG} without, [as it were,]_{NAI} [displeasing people who don't
 want]_F [Brexit to get done]_{BG}

3 Desiderata for future annotation projects

(5) [Interview:]
 BJ: Q₄₈: { *What about this new system for custom checks?* }
 > A_{48'}: Now, [we think]_{EV} [those checks]_T [can be absolutely minimal]_F
 > A_{48''}: [and]_{DM} [non-intrusive]_F

(6) [Narrative:]
 Q₄₆: { *Where does Mrs Simpson's prejudice come from?* }
 > A_{46'}: [Susie suspects]_{EV} [it]_T [has something to do]_{BG} [with the small rooms]_F

3.1.2 (C)TT [(Contrastive) Topic Time]

Topic times (Klein 1992) temporally localize the event or state expressed by the current utterance. If overtly expressed in the form of temporal anaphors or adverbials, they may occur in language-specific syntactic positions, see (7). Like ordinary sentence topics, topic times may be contrastive, as in (8).

- (7) [Interview:]
LK: Q₂: { *What happened to Johnson in the last few weeks?* }
> A₂: [in the last few weeks,]_{TT} [you've]_T [lost major votes in the Commons,]_F
- (8) [Narrative:]
> A_{73''}: [She]_T [takes a deep breath]_F.
> A_{73'''}: [Exhales]_F.
> Q₇₄: { *What will Susie do after the bath?* }
>> A₇₄: [After,]_{CTT} [she]_T [will change Mrs Simpson's sheets,]_F

3.2 Information structural annotation of questions

The current version of the QUDA tool does not allow for the information-structural annotation of questions, which might be an interesting path to pursue in the near future.

3.2.1 Overt questions

Overt questions are utterances whose realization is subject to similar – though not exactly the same – information-structural constraints as assertions. In particular, overt questions can have one or several sentence topics (T) and a focus (F), although the latter cannot itself be defined as the answer to a question. The sentence topic of a question is usually its given material, connecting the question to the previous discourse, while the focus of a question indicates a position that is subject to potential alternation. It may also be the element that corresponds to a CT in the answer. (Note that the focus of a question is not its *wh*-element!)

3.2.2 Annotation of discourse topics

I consider as discourse topics (DT) all expressions mentioned in a QUD (both explicit and implicit).² DTs are annotated only in questions. Each DT needs to be annotated exactly one time and has scope over the entire section dominated by the QUD. In other words, a discourse does not have *one* discourse topic, but a hierarchy of discourse topics; the higher in the QUD-tree, the more pervasive an entity is. In the current narrative text, important discourse topics are *Susie*, *Mrs Simpson* and, perhaps, *the tea*.

References

Asher, N. & A. Lascarides. 2003. *Logics of conversation*. Cambridge University Press.

²cf. forthcoming work with Amalia Canes Nápoles

- Brunetti, L., K. De Kuthy & A. Riester. 2021. The information-structural status of adjuncts: A question-under-discussion-based approach. *Discours* 28. 3–48. <https://doi.org/10.4000/discours.11454>.
- De Kuthy, K., N. Reiter & A. Riester. 2018. QUD-based annotation of discourse structure and information structure: Tool and evaluation. In N. Calzolari et al. (eds.), *Proceedings of the 11th language resources and evaluation conference (LREC)*, 1932–1938. Miyazaki, JP. <https://aclanthology.org/L18-1304/>.
- Hesse, C. et al. 2022. Testing focus and non-at-issue frameworks with a question-under-discussion-annotated corpus. In *Proceedings of the thirteenth language resources and evaluation conference*, 5212–5219. Marseille. <https://aclanthology.org/2022.lrec-1.559>.
- Klein, W. 1992. The present perfect puzzle. *Language*. 525–552.
- Mann, W. & S. Thompson. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text* 8(3). 243–281.
- Polanyi, L. 1988. A formal model of the structure of discourse. *Journal of Pragmatics* 12(5-6). 601–638.
- Potts, C. 2005. *The logic of conventional implicatures*. Oxford University Press.
- Riester, A., L. Brunetti & K. De Kuthy. 2018. Annotation guidelines for Questions under Discussion and information structure. In E. Adamou et al. (eds.), *Information structure in lesser-described languages: Studies in prosody and syntax*, 403–443. Amsterdam: Benjamins.
- Riester, A., A. Canes Nápoles & J. Hoek. 2021. Combined discourse representations: coherence relations and questions under discussion. In *Proceedings of the workshop ‘integrating perspectives on discourse annotation (discann)’*. Tübingen.
- Roberts, C. 2012 [1996]. Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5(6). Earlier version in OSU Working Papers in Linguistics, Vol. 49, 1–69.
- Taboada, M. & W. Mann. 2006. Rhetorical Structure Theory: Looking back and moving ahead. *Discourse Studies* 8. 423–459.