

Notes on QUD Annotation: Short Story & Interview

Oliver Deck Tatjana Scheffler Hannah Seemann

Ruhr-Universität Bochum

1 Notes on the Annotation Process

All three authors carried out annotations individually, following the annotation guidelines described in (Riester et al., 2018). We focused on the narrative text and the interview, we did not annotate the car review yet.

The individual annotation was followed by a joint curation of **only** the segmentation. We then amended our text segmentations so that we now have three annotations per text based on identical segments. This will make further comparisons easier.

In the following section, we enumerate some points that led to discussion when comparing our annotations.

2 Discussed Issues

2.1 Segmentation

The segmentation guidelines are quite sparse; in practice, this leads to many edge cases. In general, clear guidelines for how and when to split conjunctions, short sentences, etc., would be needed.

Verbal complements. We decided not to split off clauses if they are embedded under a verbal head (1). This is in accordance with the annotation guidelines, but we had to discuss some of the cases.

- (1) But I think once we get it done, and once we can begin building a new partnership with our new friends, once we can start thinking about how we can do things differently, how we can interact with the rest of the world, how we can recover on our impetus, our mojos, as a global outward looking.

Conjoined phrases. Conjoined phrases pose problems since it is not always clear if the phrases should be considered (elliptical) clauses (i.e., segments), or not. In (2) we decided to split the segments, in other cases one could decide that conjoined noun phrases or even VPs are not separate assertions.

- (2) Along the shelves are glass jars of coloured salts,
and powders,
and liquids.

Disfluencies. Should the following example be one segment or two? The first sentence is not a complete sentence, but incorporates false starts and repetitions. One might argue that it nevertheless conveys a complete meaning. In our structure the first line could also be the answer to a question one level above where it would not cause problems.

- (3) The extrication, after 45 years of our legal system, from the orbit of European law, which is you know, has become very, very pervasive.
It's a very complicated thing to do.
- (4) A11: because obviously what the UK is going through is a big constitutional change.
Q12: {What big constitutional change is the UK going through?}
A12: The extrication, after 45 years of our legal system, from the orbit of European law, which is you know, has become very, very pervasive.
Q12.1: {How complex is the extrication from the orbit of European law?}
A12.1: It's a very complicated thing to do.

Discourse markers, e.g., ‘Let’s talk about that’. We discussed whether these constitute their own assertions (especially when they have sentence form like here) or not. In this case, we decided to treat the discourse marker as part of the following segment.

- (5) LK: Let’s talk about that. So let’s be completely clear, under the proposals that you were about to take to Brussels, there would be extra checks on the island of Ireland, how and where?

2.2 Labelling and annotation

NAI in narrative (‘A play’). Here, ‘a play’ is clearly the answer to an implicit question like “What does Susie see when she checks the calendar?”. However, in the previous sentence, “as she checks the calendar” is syntactically marked as not-at-issue by being phrased in an embedded as-clause. The at-issue clause is the other one. It seems that NAI content can be used creatively in narrative to advance the action.

- (6) As she checks Mrs Simpsons calendar, Susie rubs the place where the elastic cap from work scrunched all day.
A play.

Explicit vs. implicit questions. We discussed whether Q10 is a question or not (and decided it is not, see annotation in (8))

- (7) Q10: LK: But you're suggesting that people ought to come together, when transparently, you have been trying to create this idea of them and us,
A10': you who want to get Brexit done, which you said every possible opportunity.
A10'': And the people on the other side, which you've just suggested, are only trying to hold you up and stop Brexit.
Q11: And that's transparent, you're trying to create a situation of them and us are you not?
- (8) Q7.1: {What is Boris Johnson saying about the people?}
A7.1: LK: But you're suggesting that people ought to come together,
Q7.2: {What has Boris Johnson actually been doing to the public discourse?}
A7.2: when transparently, you have been trying to create this idea of them and us,
Q7.2.1: {What are the two sides of "them and us"?}
Q7.2.1.1: {Who is "us"?}
A7.2.1.1: you who want to get Brexit done, which you said every possible opportunity.
Q7.2.1.2: {Who is "them"?}
A7.2.1.2: And the people on the other side, which you've just suggested, are only trying to hold you up and stop Brexit.
Q7.3: And that's transparent, you're trying to create a situation of them and us are you not?

But. Assertions starting with 'but' (very frequent in the narrative text) turned out to be difficult for annotation, because there is no question type that may be answered with a 'but'-statement. E.g., one cannot felicitously phrase (as an implicit question): "* What does this contrast with?"

Numbering of questions. We were unclear on how to number questions and answers and did this haphazardly. Guidelines would be helpful.

Side stories/backtracking. Complex narrative structure is difficult if not impossible to capture. One sometimes wishes for a graph structure. Some implicit questions even seemed to repeat themselves, but this cannot be indicated

(it seems however like something that the QUD model may want to be able to model: “reopening” of questions).

Embedded clauses. Some nested embedded clauses are very long and seem to drive the narrative, in cases like “Once X Y Z A B C, things will go really well”.

Initial questions. Initial questions (especially in narrative) are often impossible to phrase without introducing new information.

3 Plans

We are planning to:

- evaluate our segmentation disagreements using standard inter-annotator agreement measures (token-based, probably)
- evaluate our annotation agreement based on the curated (silver) segmentation, using the tools provided in <https://qud-comp.github.io/QUD-comp/> / tree-based F-scores or inter-annotator agreement

References

Riester, Arndt, Lisa Brunetti & Kordula Kuthy. 2018. Annotation guidelines for questions under discussion and information structure. doi:10.1075/slcs.199.14rie.