

A universal analysis procedure for information structure, QUDs and discourse structure

Arndt Riester, University of Stuttgart

Matić & Wedgwood (2013) question the possibility of universal categories of information structure. Despite this, I will present a meaning-based, language-independent analysis procedure for basic information-structural notions like *focus*, *background*, *focus domain* and (*aboutness/contrastive*) *topic* that makes use of *Questions under Discussion* (Roberts, 1996). The goal of the procedure is to turn a text or transcript of spoken discourse into a discourse tree (Figure 1) whose terminal nodes consist of (implicit, reconstructed) questions and whose terminal nodes represent the assertions of the text, in linear order. The benefit of such a tree is that it mirrors both the discourse structure and the information structure of a text. We can identify the focus of each assertion as the answer to its immediately dominating QUD. The crucial question is, of course, how to determine the QUDs. If we only consider an isolated assertion – say A_2 – then there are at least as many potential QUDs as there are syntactic constituents. However, if we take more context into account, the search space for the right question is drastically reduced.

A first step is to identify partial answers to the same QUD. For instance, A_1' and A_1'' have been identified as parallel/partial answers to QUD_1 . The latter can now be clearly formulated, since it must comprise the “lowest common denominator” between A_1' and A_1'' , i.e. the material they have in common. This material also represents the background of both assertions (sometimes including an aboutness topic), while the wh-element of QUD_1 corresponds to the non-overlapping parts of the assertions, i.e. their respective foci. The combination of a focus and a background is called a *focus domain* (Büring 2008). Note that parallel assertions in a text need not be immediately adjacent but can also be separated by intervening material. For instance, A_0' and A_0'' both answer the main question QUD_0 , but are separated by material which does not (or at least not directly) answer QUD_0 . This brings us to the second constraint: as long as we do not expect the writer or speaker to have produced an incoherent discourse (by suddenly switching to a completely unrelated topic), we can safely assume that the intervening material, e.g. QUD_1 and its answers, relate to what was immediately said before. This means that any sub-QUD to QUD_0' must contain some *given* material (i.e. relate to and ask *about* what was just said in A_0'). In my talk, I will show, using real corpus data (essays, narratives, spoken interviews, radio news) from English, German, French and Balinese, that with these two constraints we are already able to structure a large proportion of the discourse. Two further phenomena I will talk about concern the identification of *contrastive topics* (Büring 2003) and *not-at-issue content* (Simons et al. 2010), i.e. appositives, parentheticals, evidentials etc. A constructed discourse that corresponds to the abstract tree representation in Figure 1 is given in (1) (tree structure represented by means of indentations and the > symbol).

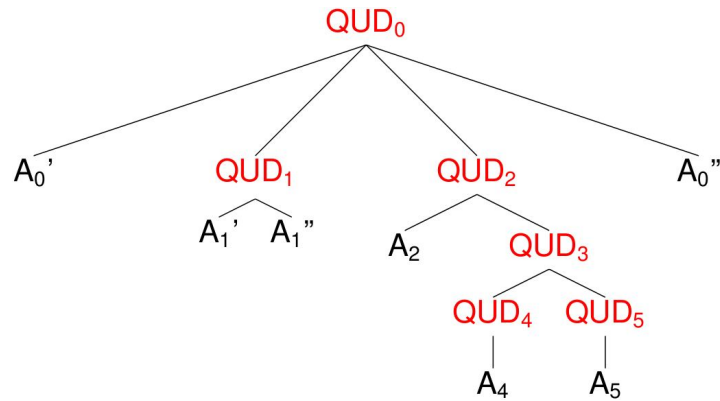


Figure 1: Discourse tree with Questions under Discussion

(1)

QUD₀: {What did Peter do on Saturday?}

> A₀' : [On Saturday, [Peter,]_{topic} [my cousin,]_{not-at-issue} [decided to go shopping]_{focus}]_{focus domain}.

> **QUD₁: {What did he buy?}**

>> A₁' : [[He]_{topic} bought [a fishing rod]_{focus}]_{focus domain}.

>> A₁'' : Moreover, [[he]_{topic} got [some 50 meters of boulder]_{focus}]_{focus domain}.

> **QUD₂: {Why did he do this?}**

>> A₂: [[He]_{topic} [wanted to surprise his friends]_{focus} with this]_{focus domain}.

>> **QUD₃: {Which friends would be surprised and why?}**

>>> **QUD₄: {Why would Mary be surprised?}**

>>> A₄: [[Mary]_{contrastive topic} [had been talking about going fishing for years]_{focus}]_{focus domain}.

>>> **QUD₅: {Why would Harry be surprised?}**

>>> A₅: And [[Harry]_{contrastive topic} [is great at cooking trouts]_{focus}]_{focus domain}.

> A₀'' : Afterwards, [[Peter]_{topic} [went to the hairdresser]_{focus}]_{focus domain}.

References

- Büring, Daniel (2003). On D-Trees, beans, and B-Accents. *Linguistics & Philosophy* 26(5): 511-545.
- Büring, Daniel (2008). What's new (and what's given) in the theory of focus? In: *Proceedings of the 34th Annual Meeting of the Berkeley Linguistics Society*, pp. 403-424.
- Matić, Dejan and Daniel Wedgwood (2013). The meanings of focus: the significance of an interpretation-based category in cross-linguistic analysis. *Journal of Linguistics* 49(1): 17-163.
- Roberts, Craige (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. *OSU Working Papers in Linguistics* 49. The Ohio State University.
- Simons, Mandy, Judith Tonhauser, David Beaver and Craige Roberts (2010). What projects and why. In: *Proceedings of SALT 20*, pp. 309-327. Vancouver.