Addressing unexpected questions in discourse

Swantje Tönnis and Judith Tonhauser
Stuttgart University

Previous research has assumed a broad range of linguistic phenomena to be sensitive to questions in discourse (e.g., Roberts 1996/2012, Beaver & Clark 2008, Rojas-Esponda 2014, Onea 2016). There have, however, only been few experimental investigations of the question-based structure of discourse (e.g., Kehler & Rohde 2017, Westera & Rohde 2019); in particular, there are no investigations on when unexpected questions can be addressed. In this paper, we contribute to filling this gap by providing evidence for two hypotheses: Exp 1 uses a novel experimental design to show that, in German narrative discourses, questions that are expected to be addressed become more unexpected as the discourse proceeds. Exp 2, a case study on German clefts, shows that relatively unexpected questions can in fact be addressed (in line with Tönnis 2021), not only the most recently-introduced question (as in Roberts 1996/2021).

Data and previous research: Most previous research focused on how to address expected questions. Roberts (2012), for instance, claimed that a discourse move must address the top-most question on the QUD stack, or sub-questions thereof. Extending this, Rojas-Esponda (2014) proposed that it is also possible to address super-questions of the top-most question. Kehler & Rohde (2017) assumed that addressees form a probability distribution over possible questions that the ensuing utterance is going to address. Tönnis (2021) pointed out that this distribution changes when discourse proceeds. For example, the question Q1 is more expected in (1) than in (2).

(1) *When Lilly joined breakfast the rolls were already gone.*
Q1: Who ate the last roll? [relatively expected question]

(2) *When Lilly joined breakfast the rolls were already gone. There weren't any croissants or toast either. So she went to the bakery nearby.*
Q1: Who ate the last roll? [relatively unexpected question]

Roberts (1996/2012) and Rojas-Esponda (2014) predict that Q1 cannot be addressed in the next sentence of (2), given that it is neither the top-most question nor a super-question. Following Onea (2016) and Kehler & Rohde (2017), Tönnis (2021) argued that in German Q1 in (2) can in fact be addressed, namely by a cleft (*It was Benni who ate the last roll*), which she assumed to mark that a relatively unexpected question is addressed. Tönnis' (2021) discourse analysis assumed that an expectedness value is assigned to each possible question at each stage of a discourse. This value represents how strongly the addressee expects the respective question to be addressed in its context. She assumed that the expectedness of a question is higher the smaller the distance of the question to the question-raising sentence is. The question Q1 is raised by the sentence in (1), and is predicted to be more expected in context (1) than in context (2). Exp 1 tests this prediction while the Exp 2 tests whether the expectedness of the addressed question affects the acceptability of German clefts.

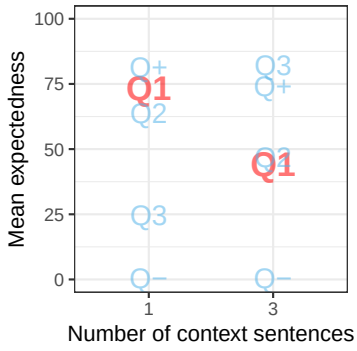
Previous experiments mainly focused on eliciting questions which are evoked in discourse. Kehler & Rohde (2017) used continuation tasks, which showed that linguistic cues affect the identification of the QUD. Westera & Rohde (2019) used an elicitation task to investigate which questions arise to readers in text snippets taken from corpora. However, those methods only covered expected questions. In our paradigm, it is possible to also target unexpected questions, which is necessary to test Tönnis' (2021) hypotheses.

Experiment 1 (n=80): Expectedness was measured for 16 German questions in 2 conditions: after the first sentence of a discourse, as in (1), and after the third sentence of a discourse, as in (2). For each discourse, an array of 5 different questions was presented consisting of a question raised by the first (Q1:*Who ate the last roll*), second (Q2:*What could Lilly have for breakfast instead?*) and third (Q3:*What did Lilly buy at the bakery?*) sentence, a very unexpected control (Q–:*What*

was the weather in Colombia?), and a relatively expected control (Q+:*What did Lilly do next?*). Participants were asked to rate the expectedness of each question to be addressed in the next sentence on a sliding scale from 'absolutely unexpected' (coded as 0) to 'very expected' (coded as 100). The expectedness of Q1 was evaluated while the other questions served as baselines.

Results Exp 1: The mean ratings of Q1 were significantly higher after the first sentence than after the third sentence, see Fig. 1 for the expectedness means of all five questions in both conditions. The result was confirmed by a linear mixed effects model (R, *lme4*) that predicted the expectedness rating of Q1 from a fixed effect of number of context sentences (reference level: one context sentence) with participant and item as random effects and a by-participant slope (β = -29.3, *SE* = 2, *t* = -15, $p < .001$).

Figure 1: Mean expectedness by question and number of context sentences.

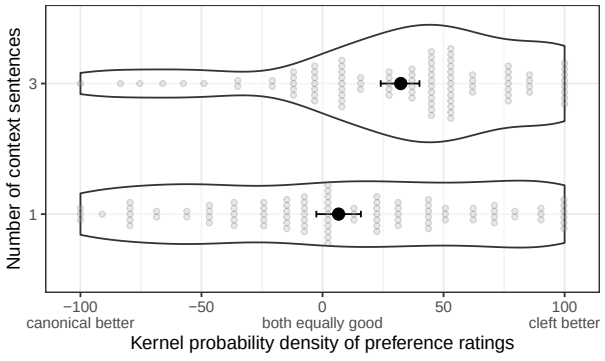


Experiment 2 (n=120): Relative preference ratings for German clefts (e.g., *It was Benni who ate the last role*) compared to their canonical variants (*Benni ate the last role*) were measured for 16 target items both after the first and after the third sentence. The cleft and the canonical sentence both addressed the question raised by the first sentence. Participants were told that the next sentence of the text was illegible, and they were asked to indicate their relative preference between the two alternatives (A and B) on a slider ranging from 'A (canonical) much better' (coded as [-100,0]) to 'B (cleft) much better' (coded as [0,100]), and 'equally good' in the middle.

Results Exp 2: There was a significantly stronger preference for the cleft after the third sentence than after the first sentence, see Fig. 2. This result is supported by a linear mixed effects model that predicted the relative preference rating from a fixed effect of number of context sentences (reference level: one context sentence) with participant and item as random effects and a by-participant slope (β = 26, *SE* = 5.8, *t* = 4.5, $p < .001$). Given the results of Exp 1, this means that clefts are more acceptable when they address a relatively unexpected question.

Discussion: The results support Tönnis' (2021) hypotheses. Exp 1, furthermore, revealed that our method is suitable to attest different levels of expectedness of questions. Crucially, it can also investigate relatively unexpected questions, which cannot be elicited using the designs described above. Exp 2 showed that the QUD may very well be a relatively unexpected question as long as it is addressed with a cleft. This result speaks in favor of more flexible discourse models with respect to which questions can be addressed. The method we introduce could be used as a general paradigm for investigating further phenomena affected by discourse expectations.

Figure 2: Preference ratings by number of context sentences. Black dots represent means with 95% CIs. Light dots represent participants' means.



Selected references • Kehler & Rohde (2017). Evaluating an expectation-driven question-under-discussion model of discourse interpretation. *Discourse Processes*. • Roberts (2012). Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics. *Semantics and Pragmatics*. • Tönnis (2021). *German es-Clefts in Discourse. A Question-Based Analysis Involving Expectedness*. • Westera & Rohde (2019). Asking between the lines: Elicitation of evoked questions in text. *Amsterdam Colloquium*.