

German clefts address unexpected questions*

Swantje Tönnis
University of Stuttgart

Judith Tonhauser
University of Stuttgart

Abstract In this paper, we provide empirical evidence for Tönnis' (2021) hypothesis that German cleft sentences address relatively unexpected questions in discourse while their canonical variants address relatively expected questions. We present an experiment that measures the relative preference between the German cleft and its canonical variant in contexts that differ with respect to how expected the question is that they answer. The expectedness of the question was measured separately in a norming study. The result of the experiment supports analyses of German clefts that take discourse expectations into account when analyzing the acceptability of clefts in contrast to canonical sentences. Approaches that primarily focus on differences in exhaustivity (e.g., De Veugh-Geiss, Tönnis, Onea & Zimmermann 2018) or contrast (e.g., Rochemont 1986) need to be adapted in order to account for the results.

Keywords: German clefts, discourse expectations, question expectedness, questions in discourse

1 Introduction

The German *es*-cleft, exemplified in (2b), is a rather infrequent construction in German, compared to clefts in languages like French. Contexts in which the cleft is preferred over its canonical variant, exemplified in (2a), are rare. The context in (2) represents such a case (inspired by Tönnis 2021): The cleft is preferred over the canonical sentence in (2). In many other contexts, e.g., in (1), the preference is reversed.

- (1) Als Benni in den Schuppen kam, war sein Fahrrad zugestellt.
'When Benni came into the shed his bicycle was blocked.'
- (2) Als Benni in den Schuppen kam, war sein Fahrrad zugestellt. Er konnte es so schnell nicht frei bekommen. Also fuhr er mit dem Tretroller los.
'When Benni came into the shed his bicycle was blocked. He couldn't get it out quickly enough. So he set off on the scooter.'

* We thank the SALT32 reviewers and audience, the ELM2 reviewers and audience, as well as audiences at Potsdam University and the University of Stuttgart for valuable feedback and fruitful discussions.

- a. Lilly hat vor Bennis Fahrrad geparkt. ✓ in (1), ? in (2)
'Lilly parked in front of Benni's bicycle.'
- b. Es war Lilly, die vor Bennis Fahrrad geparkt hat. ? in (1), ✓ in (2)
'It was Lilly who parked in front of Benni's bicycle.'

Tönnis (2021) argued that previous approaches to clefts which primarily focused on exhaustivity, uniqueness, or contrastivity cannot account for this contrast (e.g., Prince 1978; Horn 1981; Rochemont 1986; Velleman, Beaver, Destruel, Bumford, Onea & Coppock 2012). Instead, Tönnis (2021) hypothesized that the acceptability of German clefts is sensitive to the expectedness of the question addressed: while clefts address questions that are relatively unexpected in discourse, their canonical variants address relatively expected questions. The question addressed by (2a) and (2b) in both contexts above is *Who parked in front of Benni's bicycle?* Following Tönnis (2021), we argue that this question is relatively more expected in context (1) than in context (2) because the story has moved on to a different topic in the latter context. Consequently, the cleft is more acceptable in contexts such as (2), and the canonical variant in contexts such as (1), as predicted by Tönnis (2021).

In this paper, we provide empirical evidence for Tönnis' (2021) hypothesis. We developed a novel experiment paradigm designed to measure question expectedness (norming study). In the main experiment, we investigated the hypothesis by collecting relative preference ratings for the cleft and its canonical variant in contexts which differ with respect to how expected the question addressed by the two variants is.

The paper is structured as follows: In subsection 2.1, we summarize prior approaches to clefts, and show that they cannot capture the contrast in (1) and (2). In subsection 2.2, the analysis developed in Tönnis (2021) is presented. In section 3, we present the experiment. In section 4, we discuss theoretical and methodological implications of our results and specific details of our stimuli. Section 5 concludes.

2 The discourse-sensitivity of clefts

2.1 Previous approaches to clefts

On some prior approaches, a cleft contributes an inference that is either weaker or not present for its canonical variant. On Percus' (1997) analysis, for instance, clefts trigger a uniqueness presupposition that is not triggered by their canonical variants. The presupposition is contributed by a definite description from which the cleft is derived: The cleft in (3a) triggers the presupposition that there is a unique individual that parked in front of Benni's bicycle, derived from the definite description in (3b).

- (3) a. It was Lilly who parked in front of Benni's bicycle.
 b. The one who parked in front of Benni's bicycle was Lilly.

This analysis cannot account for the contrast in (1) and (2) because the two contexts do not differ with respect to the relevant presupposition; specifically, neither context satisfies the presupposition. Consequently, Percus' (1997) analysis predicts that the cleft is either unacceptable in both contexts (because the presupposition is not satisfied) or acceptable in both (if the presupposition can be accommodated). Either way, Percus (1997) does not predict the contrast observed in (1) and (2).

Other analyses considered the cleft to have an exhaustivity inference (e.g., Büring & Križ 2013; Horn 1981). The cleft in (3a) has the exhaustivity inference in (4).

(4) Nobody other than Lilly parked in front of Benni's bicycle.

De Veugh-Geiss et al. (2018) found in a picture-verification-task that the exhaustivity inference of German clefts was stronger than for canonical sentences. De Veugh-Geiss, Zimmermann, Onea & Boell (2015) found that the German cleft was judged as less acceptable than the canonical sentence when followed by an exhaustivity violation, as indicated in the English translation in (5).

- (5) a. ? It was Lilly who parked in front of Benni's bicycle, and Anna as well.
b. Lilly parked in front of Benni's bicycle, and Anna as well.

By contrast, the corpus study by Pavlović (2019) suggested that German clefts do not necessarily contribute an exhaustivity inference, as there were many occurrences of clefts with additive particles in the pivot (i.e., *It was also Lilly who...*). Similarly, we observe that the cleft in (2), repeated in (6), does not appear to contribute exhaustivity: The cleft with the exhaustivity violation in (6b) is acceptable and is still preferred over the canonical variant (6a). This data suggests that analyses on which clefts mainly differ from their canonical variants by an exhaustivity inference cannot account for the contrast in (1) and (2).

- (6) *When Benni came into the shed his bicycle was blocked. He couldn't get it out quickly enough. So he set off on the scooter.*
a. ? Lilly hat vor Bennis Fahrrad geparkt und Anna auch.
'Lilly parked in front of Benni's bicycle, and Anna as well.'
b. Es war Lilly, die vor Bennis Fahrrad geparkt hat, und Anna auch.
'It was Lilly who parked in front of Benni's bicycle, and Anna as well.'

Another line of prior research was concerned with contrastivity in clefts. Rochemont (1986), for instance, argued that clefts express contrastive focus while their canonical variants are compatible with both information and contrastive focus. This approach does not capture the contrast in (1) and (2), either, because the two contexts do not differ with respect to contrastivity according to the criteria for contrast in Repp (2010). Specifically, the two contexts do not differ in terms of the set of (explicit) alternatives to Benni (no additional individuals are mentioned in either context).

Hence, the contrast in contexts (2) and (1) cannot be explained by arguing that the conditions for a contrastive statement differ in the two contexts.

Other previous approaches analyzed the cleft in its wider discourse context, and investigated how the cleft, compared to its canonical variant, is integrated into the previously constructed discourse structure. Corpus studies on English clefts (a.o., Prince 1978; Hedberg 1990), for instance, showed that it is not always possible to replace a natural occurrence of a cleft in its context with its canonical variant without losing text coherence. Consider the pair of examples from Delin & Oberlander (1995: 471) in (7): the naturally occurring discourse with the cleft in (7a) is coherent. In contrast, the variant in (7b), in which the cleft is replaced by its canonical variant, is incoherent.

- (7) a. Doubling the selling space to 700 square feet was not to be the greatest expense. **It was the new fixtures and fittings to fill this space that would be costly.**
- b. ? Doubling the selling space to 700 square feet was not to be the greatest expense. **The new fixtures and fittings to fill this space would be costly.**

We take this data as evidence that the discourse structure needs to be considered to fully capture the acceptability of the cleft in comparison to its canonical variant. Velleman et al. (2012) presented such an analysis, which is based on questions in discourse. They treated clefts as inquiry-terminating constructions, as in example (8) (a slightly adapted version of Velleman et al. 2012: 449).

- (8) A: What did Mary eat?
 B: I thought she said she was gonna get a pasta dish, but I might be wrong.
 A: And did she also order a salad?
 C: Guys, I was there and actually paid attention. **It was a pizza that Mary ate.**

Velleman et al. (2012) claimed that speaker C uses the cleft to mark the ongoing inquiry with respect to the question *What did Mary eat?* as terminated. This analysis does not capture the contrast in (1) and (2), either. First, the cleft in (2b) does not even contribute to the ongoing inquiry, which would be about what Benni did or about where Benni went. The question which is actually addressed by the cleft in (2) (*Who parked in front of Benni's bicycle?*), is not one which is being discussed, it is rather a question which had been ignored and is now picked up again. Second, the inquiry about this question might be terminated by uttering the cleft, but not necessarily. The author could also continue discussing this question, e.g., as in (6b).

Finally, Destruel & Velleman (2014) proposed an analysis which involves the discourse structure as well as contrastivity. They argued that the acceptability of the

cleft is affected by violations of expectations on the content level (9a) as well as on the discourse level (9b).

(9) *Conflict with expectations*: Clefts are more felicitous the more they conflict with interlocutors' expressed expectations.

a. *Expectations about the world*: These expectations may involve beliefs about the world, expressed as assertions or presuppositions. More strongly expressed beliefs lead to stronger conflict.

b. *Expectations about the discourse*: These expectations may involve beliefs about the direction in which the discourse is going, expressed, among other ways, by marking content as at-issue or not-at-issue.

(Destruel & Velleman 2014: 199)

In their study, Destruel & Velleman (2014) investigated clefts in the corrective function. They manipulated the discourse context with respect to whether the antecedent of the correction was provided as at-issue or as not-at-issue content. The examples in (10a) and (10b) are taken from their study (Destruel & Velleman 2014: 212–213). They argued that the antecedent *Leah convinced him to buy it* is not-at-issue in (10a) while it is at-issue in (10b). One of their main findings was that the cleft was strongly preferred over the canonical sentence when the antecedent was not-at-issue, as in (10a), as opposed to when it was at-issue, as in (10b).

(10) I can't believe Mark bought that ugly car. It looks like it's about to fall apart, too.

a. I have no idea how Leah convinced him to buy it. [not-at-issue]

– Actually, it was Anna who convinced him. [cleft preferred]

b. And it turns out that Leah convinced him to buy it. [at-issue]

– Actually, Anna convinced him. [canonical sentence preferred]

This difference is predicted by their hypothesis, in particular part (9b): Since the antecedent of the correction is marked as not-at-issue in (10a) the interlocutors' expectations are violated more strongly than in the case of the at-issue marked antecedent in (10b). Hence, the cleft is correctly predicted to be more felicitous in (10a) than in (10b).

Following the hypothesis in (9), the contrast in (1) and (2) can be explained: One could hypothesize that the interlocutors' expectations differ in the two contexts with respect to what the next sentence is about. The expectations that the next sentence is about who parked in front of Benni's bicycle are stronger in (1) than in (2), because the text moved on to a new topic in (2) (Benni choosing the scooter instead of the bicycle). Hence, the cleft violates the expectations more strongly in (2) compared to (1) and is, thus, predicted to be more felicitous in (2) than in (1). In contrast

to [Destruel & Velleman's \(2014\)](#) examples, the contexts (1) and (2) do not differ with respect to at-issueness: In both contexts, the antecedent content that somebody parked in front of Benni's bicycle is implied by the first context sentence, which is identical in the two contexts. Thus, this content is not at-issue in both contexts. Hence, the differing expectations about discourse are not signaled in (1) and (2) by explicitly marking content as either at-issue or not at-issue. Rather, as proposed in [Tönnis \(2021\)](#), the two contexts differ in the status of the question addressed by the cleft and the canonical sentence. This analysis is introduced next.

2.2 German clefts and expectations in discourse

Analyzing the discourse functions of German clefts,¹ [Tönnis \(2021\)](#) argued that most of the cases in which a German cleft is acceptable can be subsumed under one discourse function (a related analysis is presented by [Delin & Oberlander 1995](#)). [Tönnis](#) hypothesized that clefts are used to address a relatively unexpected question, and canonical sentences are used to address a relatively expected question.

One of the assumptions underlying this hypothesis is that every sentence addresses a question, which may be implicit or expressed by an interrogative sentence (following [Roberts 2012](#)). As in [Zimmermann \(2011\)](#), [Tönnis \(2021\)](#) proposed that interlocutors form expectations about the direction in which the discourse is going. According to her model, each possible question that could be asked or answered in the next discourse move is paired with an expectedness value (values between 0 and 1). This value represents the likelihood that the next discourse move is going to address the respective question. In context (2), the question *Where did Benni go with the scooter?* is assigned a rather high value (i.e., close to 1), while the question *How was the weather in Colombia?* is assigned a value close to 0. Diverging from [Roberts' \(2012\)](#) model, any question may be addressed (not just the one at the top of a stack), as long as it exceeds a certain expectedness threshold. This approach is in line with [Kehler & Rohde \(2017\)](#), who analyzed the discourse expectations as a probability distribution over questions the ensuing utterance is likely to answer.

Furthermore, [Tönnis \(2021\)](#) argued that the distribution of expectedness values is

¹ The analysis is restricted to focus-background clefts, such as (ia), where the pivot represents the focus. Topic comment clefts, as in (ib), where the cleft pivot is the topic and the relative clause contains new information, are not covered. Main stress is marked in capital letters in the examples.

- (i) *When Benni came into the shed his bicycle was blocked. He couldn't get it out quickly enough. So he set off on the scooter.*
 - a. It was LILLY who parked in front of Benni's bicycle.
 - b. It was Benni who CRASHED the scooter later that day.

updated after each new sentence because a sentence can affect the distribution. Sentence (11), for instance, gives rise to the question Q *Why did Mary scold John?* (e.g., [Onea 2016](#)). In other words, asserting (11) would strongly raise the expectedness value of Q. This, in turn, would reduce the expectedness of other questions because all expectedness values have to sum up to 1 (like in a probability distribution).

(11) Mary scolded John.

[Kehler & Rohde's \(2017\)](#) findings on implicit causality verbs (e.g., *scold*) support this observation: In implicit causality contexts, participants produced significantly more *why*-questions than in non-implicit causality contexts as, for instance, in *Mary saw John*.

A factor which reduces the expectedness value of a yet unanswered question is a greater distance to the sentence which raised that question. In other words, when an author decides to not address a question immediately, the reader will expect less and less that the question will still be addressed. This is also a side effect of new questions becoming more expected, since this automatically reduces the expectedness of other questions ([Tönnis 2021: 273ff.](#)). For illustration, consider again the two contexts in (1) and (2), repeated here:

(1) *When Benni came into the shed his bicycle was blocked.*

(2) *When Benni came into the shed his bicycle was blocked. He couldn't get it out quickly enough. So he set off on the scooter.*

The first sentence in (1) and (2) gives rise to the question Q *Who parked in front of Benni's bicycle?*. The word triggering Q is *zugestellt* 'blocked', which describes a resultative state but leaves open the source of this resultative state. This can be argued to give rise to Q, which is asking for this source. In context (2), there are new questions arising due to the second and the third sentence, such as *What else could Benni use instead of the bicycle?* or *Where did Benni go with the scooter?*. Those questions are then relatively more expected than Q, and would cause the expectedness value of Q to decrease.

These assumptions about discourse being structured by questions, and question expectedness decreasing with intervening sentences suffice to account for the contrast between the cleft and the canonical sentence in (12) in the contexts (1) and (2).

- (12) a. Lilly hat vor Bennis Fahrrad geparkt. ✓ in (1), ? in (2)
 'Lilly parked in front of Benni's bicycle.'
- b. Es war Lilly, die vor Bennis Fahrrad geparkt hat. ? in (1), ✓ in (2)
 'It was Lilly who parked in front of Benni's bicycle.'

Following [Tönnis \(2021\)](#), we hypothesize that the question Q is relatively more expected in the context in (1) than in the context in (2). Given [Tönnis' \(2021\)](#)

hypothesis that German clefts are used to address relatively unexpected questions and canonical sentences address relatively expected questions, it is predicted that the cleft in (12b) is judged to be more acceptable in context (2) than in context (1), and that the canonical sentence in (12a) is judged to be more acceptable in context (1) than in context (2).

This analysis naturally extends to the example in (13), which Velleman et al. (2012) identified as problematic for their own analysis:

- (13) A: What did Mary eat?
 B: ?It was a pizza that Mary ate. (Velleman et al. 2012: 449)

The question addressed by the cleft, namely what Mary ate, is relatively expected, given that it is contributed by the immediately preceding interrogative sentence. Therefore, the analysis developed in Tönnis (2021) correctly predicts that the cleft is dispreferred here over the canonical sentence. By contrast, the same question is less expected in (8), where the conversation has already moved on to a different question (*Did she also order a salad?*). Consequently, the cleft is correctly predicted to be acceptable here (for a more detailed discussion see Tönnis 2022: 133–141).

3 Experiment

We ran an experiment designed to investigate Tönnis' (2021) hypothesis that German cleft sentences address relatively unexpected questions in discourse while their canonical variants address relatively expected questions. Participants rated the relative acceptability of a cleft and its canonical sentence variant in a context in which the question addressed by the cleft and its canonical sentence variant was more or less expected. Question expectedness was normed prior to the experiment.²

3.1 Participants

120 self-declared native speakers of German, recruited and paid (£9.10 per hour) on Prolific, participated in the experiment (53 female, 66 male, 1 non-binary; ages 18–71 years, mean age 31).

3.2 Materials

The target sentences were pairs of a German canonical sentence and its cleft sentence variant, as shown in (14a) and (14b), respectively. The two members of a pair convey

² The materials of the experiment and the norming study, as well as the data and R code for generating the figures and analyses reported on in this article are available at https://github.com/swantje-toennis/expectations_cleft. The norming study and the experiment were preregistered: norming study at <https://osf.io/sb8d7>, main experiment at <https://osf.io/pd29h>. Informed consent was obtained.

the same proposition (here, that Lilly parked in front of Benni's bicycle), but differ in their form. There were 16 pairs of target sentences. The full set of stimuli is provided in Supplement A in the GitHub repository mentioned in footnote 2.

- (14) a. Lilly hat vor Bennis Fahrrad geparkt.
'Lilly parked in front of Benni's bicycle.'
b. Es war Lilly, die vor Bennis Fahrrad geparkt hat.
'It was Lilly who parked in front of Benni's bicycle.'

For each of the pairs of target sentences, we created a German target context sentence that gives rise to a question that is addressed by the respective target sentence. This question is referred to as the target question. The context sentence for the pair in (14) is given in (15): This context sentence gives rise to the target question *Who parked in front of Benni's bicycle?*, which is addressed by both sentences in (14).

- (15) Als Benni in den Schuppen kam, war sein Fahrrad zugestellt.
'When Benni came into the shed his bicycle was blocked.'

Each target stimulus consisted of a pair of target sentences, such as (14), that was preceded by a context in one of two context conditions. In the first context condition, the context consisted only of the target context sentence, as in (15). In the second context condition, the context consisted of the target context sentence as well as two additional sentences that carried the story on, as in (16).

- (16) Als Benni in den Schuppen kam, war sein Fahrrad zugestellt. Er konnte es so schnell nicht frei bekommen. Also fuhr er mit dem Tretroller los.
'When Benni came into the shed his bicycle was blocked. He couldn't get it out quickly enough. So he set off on the scooter.'

The construction of each context followed the same principle. The target context sentence always consisted of two clauses: a temporal clause, which located an individual's actions in time (e.g., when Benni came into the shed), and a main clause which introduced a problematic state for the individual (such as Benni's bicycle being blocked). The target context sentences all contained a state adjective/participle, such as *zugestellt* 'blocked'. In the condition where the context consisted of three sentences, the second context sentence further described this problematic state (such as it being impossible to free the bicycle quickly). By contrast, the third context sentence always described a new event, temporally located after the event denoted in the temporal clause of the target context sentence (such as Benni leaving on his scooter). This temporal progression was also indicated by a discourse marker, such as *also* 'so' or *schließlich* 'eventually'.

We assumed that the target question was relatively more expected immediately after the target context sentence, that is, when the context only consisted of the

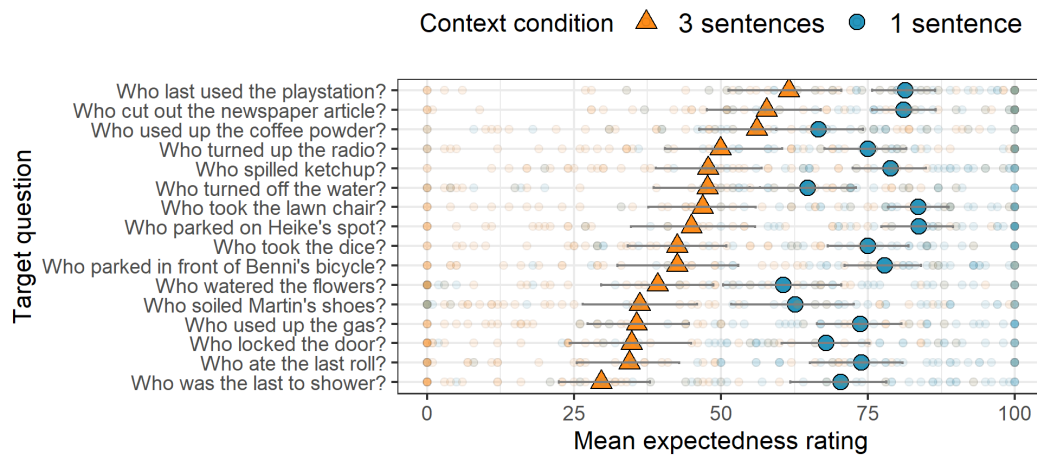


Figure 1 Mean expectedness of the (English translations of the) target questions by context condition in the norming study. Error bars indicate 95% bootstrapped confidence intervals. Transparent dots indicate individual participants' ratings.

target context sentence, as in (15), and relatively less expected when two additional sentences that carried the story on had occurred after the target context sentence, as in (16). We therefore refer to these two context conditions as the ‘higher expectedness’ and the ‘lower expectedness’ conditions, respectively. Our assumption was confirmed by a norming study, in which 80 self-declared native speakers of German gave slider ratings (from 0 to 100) about how expected the target questions (as well as four other questions that varied in their expectedness) were in the two context conditions.³ Figure 1 plots the mean expectedness ratings of the target questions by context condition. As expected, the mean expectedness ratings of the target questions were higher immediately after the target context sentence (i.e., in the higher expectedness context condition) than when the target context sentence was followed by two additional sentences (i.e., in the lower expectedness context condition).

The 32 target stimuli were distributed across eight lists such that each list had four distinct items (where an item is a combination of a target context sentence and its corresponding pair of target sentences): two in the higher expectedness condition and two in the lower expectedness condition. Each list also included the same eight

³ None of the participants of the norming study later participated in the main experiment. Details of the norming study can be found in Appendix A of this paper.

filler stimuli and the same four control stimuli, for a total of 16 stimuli per list. As clefts are very infrequent in German, we only included four target stimuli on each list, to mitigate against participants realizing that the experiment was about clefts. The filler and control stimuli both consisted of a context and a pair of sentences, like the target stimuli. The contexts of the filler stimuli consisted of one, two, or three sentences, and the pairs of sentences differed in their discourse properties, such as the presence of a discourse marker, or whether a determiner phrase was fronted or occurred in situ.

The control stimuli consisted of two one sentence and two three sentence contexts. Each pair of sentences denoted the same proposition but differed in whether a target referent was referred to by a pronoun or a definite description, as illustrated in (17) and (18). Two of the controls were like (17): They involved a non-salient referent (the computer). This referent was introduced in the first sentence of the context and referred to after two intervening sentences. Here, we expected the definite description variant in (17b) to be preferred over the pronoun variant because non-salient referents are referred to by more complex referring expressions, like a definite description (see Gundel, Hedberg & Zacharski 1993). The computer (masculine in German) was the only suitable antecedent for the masculine pronoun *er* ‘he’. Other later introduced entities were excluded as an antecedent because they had a different gender, here the feminine gender of *Freundin* ‘female friend’. The other two controls, such as (18), involved a very salient referent, introduced by a infinite description in the context sentence (here, a colleague from Poland that he is friends with). Here, we expected the pronoun variant in (18a) to be preferred over the definite description because salient antecedents are referred to by a less complex form, such as a pronoun (see Gundel et al. 1993). The full set of filler and control stimuli is given in Supplement A of the GitHub repository mentioned in footnote 2.

- (17) Heike kam in ihr Arbeitszimmer und hat den Computer hochgefahren. Dann hat sie eine Mail an eine Freundin geschrieben. Sie hat eine halbe Stunde dafür gebraucht.
‘Heike came into her office and booted the computer_{masc}. Then, she wrote an email to her friend_{fem}. It took her half an hour.’
- a. Er war sehr langsam.
‘It_{masc} was very slow.’
- b. Der Computer war sehr langsam.
‘The computer was very slow.’
- (18) Martin hat eine befreundete Arbeitskollegin aus Polen nach Hause eingeladen.
‘Martin invited home a colleague_{fem} from Poland that he is friends with.’
- a. Sie hat eine Flasche Rotwein mitgebracht.
‘She_{fem} brought a bottle of wine.’

- b. Die befreundete Arbeitskollegin aus Polen hat eine Flasche Rotwein mitgebracht.
'The colleague_{em} from Poland that he is friends with brought a bottle of wine.'

3.3 Procedure

At the beginning of the experiment, participants were introduced to a family of six that live together: the children Lilly and Benni, their parents Heike and Martin, and their grandparents Irmgard and Wilhelm. The participants were told that they would read short texts about events describing the family's everyday life in and around the house.

On each trial, participants were shown a text and a pair of sentences, consisting of a cleft and its canonical variant: one of the sentences was labeled 'A' and the other one 'B'. Participants were told that the next sentence of the text, marked by 'XXXX', is illegible. Participants were instructed to rate which of the sentences, A or B, they would prefer as the continuation of the text, and to indicate the strength of their preference by adjusting a slider marked '(A viel besser)' (A much better) on the left end, '(B viel besser)' (B much better) on the right end, and '(beide gleich gut)' (both equally good) in the middle. The order of the two members of a pair was randomized on each trial. Slider ratings were coded as a value between -100 (absolute preference of the canonical sentence over the cleft) and 100 (absolute preference of the cleft over the canonical sentence). An English translation of a sample trial in the lower expectedness condition is given in [Figure 2](#). After completing the experiment, participants filled out a short demographic survey (age, gender, education, native language, foreign languages with level, and region where the participant grew up).

3.4 Data exclusion

The data from 27 participants was excluded from the analysis because their preference rating on at least one control stimulus deviated from the group mean by more than two standard deviation in the opposite direction of what we expected. Data from 93 participants entered the analysis (40 female, 52 male, 1 non-binary, age range 18–71 years, mean age 31).

3.5 Results

[Figure 3](#) plots the mean preference rating by context condition, i.e., by the expectedness of the target question. As predicted, the mean preference rating is higher (indicating a stronger preference towards the cleft) in the lower expectedness condi-

Read the text in the box. The next sentence of this text is illegible (marked by XXXX).

*When Benni came into the shed his bicycle was blocked. He couldn't get it out quickly enough.
So he set off on the scooter. XXXX*

- A. Lilly parked in front of Benni's bicycle.
- B. It was Lilly who parked in front of Benni's bicycle.

Which of the sentences A or B would you prefer and how strongly? Adjust the slider accordingly.



Figure 2 English translation of a sample trial in lower expectedness condition of the relative preference rating with A = canonical sentence and B = cleft.

tion than in the higher expectedness condition. This result suggests that participants' preference ratings of the cleft and its canonical variant are sensitive to how expected the target question was. Specifically, the cleft was rated as more preferred over the canonical sentence in the lower expectedness condition, that is, when the target question was relatively less expected, than in the higher expectedness condition, that is, when the target question was relatively more expected.

This result was confirmed by a linear mixed effects model that predicted the preference ratings from a fixed effect of context condition (reference level: higher expectedness) and random by-participant intercepts and slopes and by-item intercepts⁴ (R version 4.1.2, *lme4* package). There was a main effect of context condition ($\beta = 26$, $SE = 5.8$, $t = 4.5$, $p < .001$), suggesting that participants' preference for the cleft over the canonical sentence was stronger when the target question was relatively less expected compared to when it was relatively more expected.

4 Discussion

4.1 Theoretical implications

The experiment revealed that the context and the resulting expectations on how the text could develop significantly affect the relative acceptability of the German cleft compared to its canonical variant. The results confirm that there is a stronger preference for the cleft over the canonical sentence when the target question is relatively less expected than when it is relatively more expected. This partly supports [Tönnis' \(2021\)](#) hypothesis, namely the part that German clefts address relatively

⁴ Random by-item slopes were removed because there was a correlation of -1 with the fixed effect of context condition.

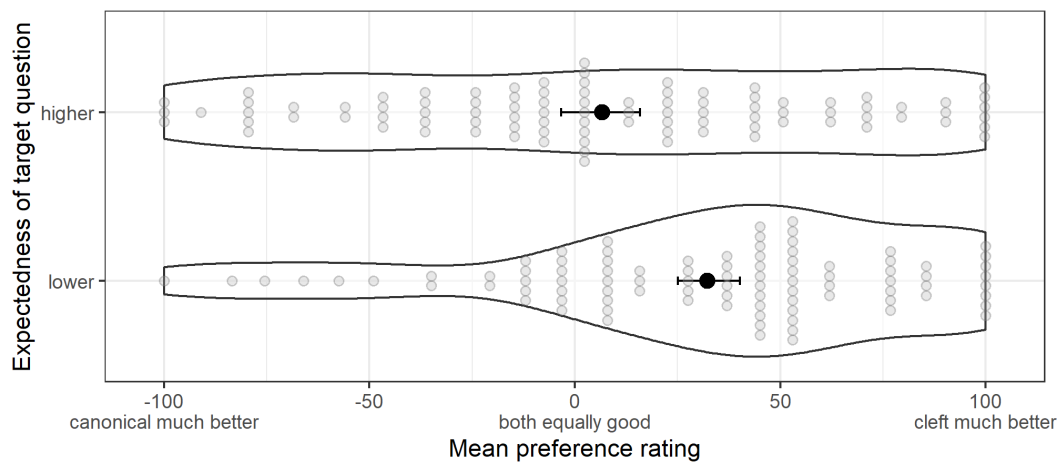


Figure 3 Mean preference rating with 95% bootstrapped confidence intervals by context condition. Kernel probability density of the individual participants' mean ratings, given in transparent dots, overlaid.

unexpected questions. Since the two context conditions in the experiment did not differ with respect to exhaustivity, uniqueness, or contrast, theories of German clefts that primarily focus on such features cannot predict the attested differences. We argue that such theories need to be extended to account for the influence of discourse expectations on the acceptability of German clefts. Tönnis (2021), for instance, explained exhaustivity and contrast as side effects of the requirement for clefts to address a relatively unexpected question (more details in Tönnis 2021: 301–306).

However, in the context condition in which the target question was relatively more expected, such as (1), the canonical sentence was not clearly preferred over the cleft. Hence, the part of the hypothesis that canonical sentences address relatively expected questions cannot be clearly confirmed. This may be due to the target question, which is relatively more expected in the higher expectedness condition compared to the lower expectedness condition, but might not be sufficiently expected in the higher expectedness condition to result in a preference of the canonical sentence over the cleft. This situation was anticipated in an additional hypothesis by Tönnis (2021: 256–261), which says that the cleft and the canonical sentence are equally acceptable when a question is addressed which is neither particularly expected nor particularly unexpected. The one sentence context (1) (*When Benni came into the shed, his bicycle was blocked*), could be such a case, if the target question *Who parked in front of Benni's bicycle?* is expected but not very strongly. The question *How did he free his bicycle?*, for instance, might be relatively more

expected in this context. We leave further exploration of the relative expectedness of questions to future research.

4.2 Methodological implications

Our results imply that experiments designed to investigate factors that modulate the acceptability of German clefts should present the examples in a context to clarify the discourse expectations. If no such context is provided, participants' judgments might be influenced by the expectations created by the discourses that they imagine.

The paradigm used in the norming study (henceforth referred to as the 'question expectedness paradigm'; see Appendix A for a full description) allows researchers to measure the discourse expectations that clefts are sensitive to. The norming study confirmed Tönnis' (2021: 273ff.) assumption that a question that is raised by a sentence becomes less expected when the discourse proceeds. We expect this novel paradigm to be useful for measuring discourse expectations more generally. The question elicitation paradigms used in Kehler & Rohde (2017) and ?, for instance, ask participants to write down a question that comes to their mind after reading a context. Such paradigms, therefore, are very useful to identify the most expected questions in a context. These paradigms do not, however, allow the researcher to investigate questions that are salient but less expected. Thus, one advantage of the question expectedness paradigm is that it can be used to measure the expectedness of any question (independently of its expectedness), which makes it possible to study discourse expectations in more detail.

A second advantage of the question expectedness paradigm over previously used paradigms is that it allows the researcher to measure the expectedness of a question in relation to other questions. This is useful, for instance, if one is interested in understanding how a change in the expectedness of one question affects the expectedness of another. For instance, the expectedness of the target question Q1 *Who parked in front of Benni's bicycle?* in context (2) was measured in relation to four other questions. The questions Q2 *What could Benni have used instead of the bicycle?* and Q3 *Where did Benni go with the scooter?* were raised by the second and third sentence of the context, respectively. The other two questions were questions which we hypothesized to have a very high expectedness throughout the discourse (Q+ *What did Benni do next?*), or a very low one (Q- *How was the weather in Colombia?*). Figure 4 plots the mean expectedness of these five questions by context condition. As predicted by Tönnis' (2021: 273ff.) hypothesis that newly raised questions can lower the expectedness of previously raised questions, the mean expectedness of the target question Q1 appeared to be higher than that of Q3 in the one sentence context condition, but the situation seems reversed in the three sentence context condition, where the third context sentence raises the expectedness of Q3.

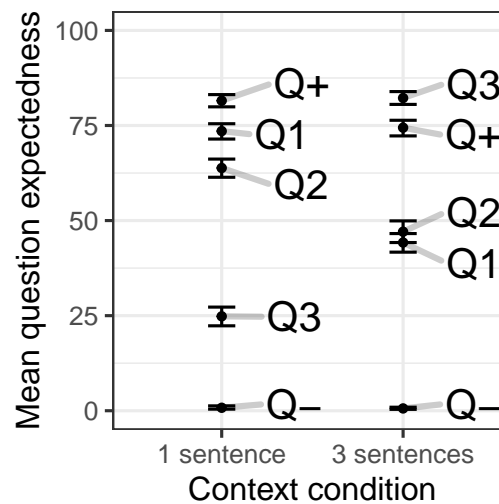


Figure 4 Mean expectedness rating with 95% bootstrapped confidence intervals by context condition and question.

The question expectedness paradigm can, we suggest, be fruitfully employed to empirically investigate a broad range of phenomena that have been analyzed as sensitive to questions and discourse expectations, such as indefinites (e.g., [Onea 2016](#)), contrastive focus (e.g., [Zimmermann 2011](#)), at-issueness (e.g., [Destruel & Velleman 2014](#)), or projection inferences (e.g., [Tonhauser, Beaver & Degen 2018](#)).

4.3 A closer look at the experiment stimuli

As described in [subsection 3.2](#), all the contexts used in this experiment were kept as similar as possible. We consider it necessary to control the contexts of the stimuli in such a strict way since small changes in the discourse structure might have a strong effect on the probability distribution over the questions the next sentence is likely to answer. Consider, for instance, the variant of the three sentence context in (2) given in (19), where the third sentence does not move the story along, but instead continues to describe the same topic. In this context, we would expect the cleft and the canonical sentence to be rated as equally acceptable.

- (19) Als Benni in den Schuppen kam, war sein Fahrrad zugestellt. Der Schuppen war sehr eng und vollgestopft. Jetzt stand auch noch ein pinkes Fahrrad vor seinem.
‘When Benni came into the shed his bicycle was blocked. The shed was very

German clefts address unexpected questions

narrow and packed. On top of that, a pink bicycle now stood in front of his.'

- a. Lilly hat vor Bennis Fahrrad geparkt.
'*Lilly parked in front of Benni's bicycle.*'
- b. Es war Lilly, die vor Bennis Fahrrad geparkt hat.
'*It was Lilly who parked in front of Benni's bicycle.*'

This example suggests that the critical property of the target stimuli in the experiment was not that the three sentence contexts consisted of three sentences but, rather, that the third sentence moved the story along. In other words, this example suggests that question expectedness may not be lowered purely by the number of intervening sentences but, rather, by the temporal properties of the story described. Future research should investigate more closely which linguistic and extralinguistic properties of discourses modulate question expectedness.

Finally, we consider the tense/aspect properties of the target sentences. The cleft sentences in the experiment consisted of a past tense main clause and a present perfect relative clause, while the canonical sentences realized a present perfect main clause. We used the present perfect in the canonical variants because it is the colloquial way of conveying past temporal reference in German (van der Klis, Le Bruyn & de Swart 2017). However, given that our stimuli were designed such that the third context sentence moved the discourse forward in time and the cleft/canonical continuation referred to the first sentence, it perhaps comes as no surprise that, in the three sentence discourse repeated in (20), the past perfect canonical variant in (20a) is relatively more acceptable than the present perfect canonical variant in (20b).

- (20) *When Benni came into the shed his bicycle was blocked. He couldn't get it out quickly enough. So he set off on the scooter.*
- a. Lilly **hatte** vor Bennis Fahrrad **geparkt**.
'*Lilly **had parked** in front of Benni's bicycle.*'
 - b. ? Lilly hat vor Bennis Fahrrad geparkt.
'*Lilly parked in front of Benni's bicycle.*'
 - c. Es war Lilly, die vor Bennis Fahrrad geparkt hat.
'*It was Lilly who parked in front of Benni's bicycle.*'

One might argue that the result of our experiment, that the cleft was preferred over the present perfect variant in the three sentence context, is not due to the target question being less expected, but because the present perfect is less suitable to express anteriority (see Delin & Oberlander (1992) for the claim that clefts can express anteriority). A piece of evidence speaking against this hypothesis is the observation that clefts are preferred over canonical present perfect variants even in contexts that do not require the expression of anteriority. In (21), for instance,

the state of Anna being bored does not temporally follow the state of Lena flirting with some guy, namely Peter, but the cleft is nevertheless preferred over the present perfect canonical variant. The past perfect variant in (21c), which conventionally conveys anteriority, is less acceptable than the cleft in this context, as it suggests that Lena first flirted with some guy and then, later, with Peter (i.e., *Peter* is not coreferential with *some guy*), which leads to a less coherent continuation.

- (21) Lena hat auf der Party mit einem Typen geflirtet. Sie hatte sehr viel Spaß. Anna war eher gelangweilt.
 ‘Lena flirted with some guy at the party. She had a lot of fun. Anna was rather bored.’
- a. Es war Peter, mit dem Lena geflirtet hat.
 ‘It was Peter Lena flirted with.’
 - b. ? Lena hat mit Peter geflirtet.
 ‘Lena flirted with Peter.’
 - c. ? Lena hatte mit Peter geflirtet.
 ‘Lena had flirted with Peter.’

Data like (21) lead us to believe that considerations about which forms can express anteriority relations do not suffice to account for the target contrast we have investigated in this paper. Another hypothesis is that the past perfect may be another strategy to address relatively unexpected questions, given that it is, just as the cleft, a more marked form in German. Either way, interactions between sentence form and tense/aspect are worthy of future investigation.

5 Conclusions

This paper provided empirical evidence for Tönnis’ (2021) hypothesis that German clefts are used to address relatively unexpected questions in discourse. Specifically, the experiment showed that clefts were judged to be more acceptable than their canonical variants in contexts in which they addressed a relatively less expected question. This result implies that analyses of German clefts cannot be limited to uniqueness, exhaustivity, or contrastivity, but should incorporate a sensitivity to question expectedness.

References

- Büring, Daniel & Manuel Križ. 2013. It’s that, and that’s it! Exhaustivity and homogeneity presuppositions in clefts (and definites). *Semantics and Pragmatics* 6(6). 1–29. doi:10.3765/sp.6.6.

- De Veugh-Geiss, Joseph P., Swantje Tönnis, Edgar Onea & Malte Zimmermann. 2018. That's not quite it: An experimental investigation of (non-)exhaustivity in clefts. *Semantics and Pragmatics* 11. doi:10.3765/sp.11.3.
- De Veugh-Geiss, Joseph P., Malte Zimmermann, Edgar Onea & Anna-Christina Boell. 2015. Contradicting (not-)at-issueness in exclusives and clefts: An empirical study. In *Semantics and Linguistic Theory (SALT)*, vol. 25, 373–393. doi:10.3765/salt.v25i0.3054.
- Delin, Judy & Jon Oberlander. 1992. Aspect-Switching and Subordination: the Role of *It*-Clefts in Discourse. In *Fourteenth International Conference on Computational Linguistics (COLING-92)*, 281–287. doi:10.3115/992066.992113.
- Delin, Judy & Jon Oberlander. 1995. Syntactic Constraints on Discourse Structure: the Case of *It*-clefts. *Linguistics* 33(3). 465–500. doi:10.1515/ling.1995.33.3.465.
- Destruel, Emilie & Leah Velleman. 2014. Refining contrast: Empirical evidence from the English *it*-cleft. *Empirical Issues in Syntax and Semantics* 10. 197–214. <http://www.cssp.cnrs.fr/eiss10/>.
- Gundel, Jeanette K, Nancy Hedberg & Ron Zacharski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69(2). 274–307. doi:10.2307/416535.
- Hedberg, Nancy. 1990. *Discourse Pragmatics and Cleft Sentences in English*: University of Minnesota PhD thesis. doi:10.13140/RG.2.2.17499.52009.
- Horn, Laurence. 1981. Exhaustiveness and the semantics of clefts. In *North East Linguistic Society (NELS)*, vol. 11, 125–142. <https://scholarworks.umass.edu/nels/vol11/iss1/10>.
- Kehler, Andrew & Hannah Rohde. 2017. Evaluating an expectation-driven question-under-discussion model of discourse interpretation. *Discourse Processes* 54(3). 219–238. doi:10.1080/0163853X.2016.1169069.
- van der Klis, Martijn, Bert Le Bruyn & Henriëtte de Swart. 2017. Mapping the perfect via translation mining. In *15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, 497–502. Valencia, Spain: Association for Computational Linguistics. <https://aclanthology.org/E17-2080>.
- Onea, Edgar. 2016. *Potential Questions at the Semantics-Pragmatics Interface*, vol. 33 Current Research in the Semantics/Pragmatics Interface. Leiden, Boston: Brill. doi:10.1163/9789004217935.
- Pavlović, Anna-Christina. 2019. *The Interpretation of it-Clefts*: Georg-August-Universität Göttingen PhD thesis. doi:10.53846/goediss-7785.
- Percus, Orin. 1997. Prying Open the Cleft. In *North East Linguistic Society (NELS)*, vol. 27, 337–352. University of Massachusetts. <https://scholarworks.umass.edu/nels/vol27/iss1/25>.
- Prince, Ellen F. 1978. A comparison of *wh*-clefts and *it*-clefts in discourse. *Language*

- 54(4). 883–906. doi:10.2307/413238.
- Repp, Sophie. 2010. Defining ‘contrast’ as an information-structural notion in grammar. *Lingua* 120(6). 1333–1345. doi:10.1016/j.lingua.2009.04.006.
- Roberts, Craige. 2012. Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics. *Semantics and Pragmatics* 5(6). 1–69. doi:10.3765/sp.5.6.
- Rochemont, Michael. 1986. *Focus in generative grammar* Studies in Generative Linguistic Analysis. Amsterdam: John Benjamins. doi:10.1075/sigla.4.
- Tonhauser, Judith, David Beaver & Judith Degen. 2018. How projective is projective content? Gradience in projectivity and at-issueness. *Journal of Semantics* 35. 495–542. doi:10.1093/jos/ffy007.
- Tönnis, Swantje. 2021. *German es-Clefts in Discourse. A Question-Based Analysis Involving Expectedness*: Graz University PhD thesis. http://swantje-toennis.github.io/files/toennis_2021_dissertation.pdf.
- Tönnis, Swantje. 2022. It is not the obvious question that a cleft addresses. In Aybüke Özgün & Yulia Zinova (eds.), *Language, Logic, and Computation*, 128–147. Cham: Springer International Publishing. doi:10.1007/978-3-030-98479-3_7.
- Velleman, Leah, David I. Beaver, Emilie Destruel, Dylan Bumford, Edgar Onea & Liz Coppock. 2012. *It*-clefts are IT (inquiry terminating) constructions. In *Semantics and Linguistic Theory (SALT)*, vol. 22, 441–460. doi:10.3765/salt.v22i0.2640.
- Zimmermann, Malte. 2011. The grammatical expression of focus in West Chadic: Variation and uniformity in and across languages. *Linguistics* 49(5). 1163–1213. doi:10.1515/ling.2011.032.

German clefts address unexpected questions

Swantje Tönnis

Department of English Linguistics, University of Stuttgart
Keplerstraße 17
70174 Stuttgart, Germany
swantje.toennis@ling.uni-stuttgart.de

Judith Tonhauser

Department of English Linguistics, University of Stuttgart
Keplerstraße 17
70174 Stuttgart, Germany
judith.tonhauser@ling.uni-stuttgart.de

A Norming study

To ensure that the target questions of the main experiment were relatively less expected in the ‘lower expectedness’ context than in the ‘higher expectedness’ context, participants rated the expectedness of the target question as well as four other questions in the two contexts.

A.1 Methods

Participants. 80 self-declared native speakers of German, recruited and paid (£9.12 per hour) on [Prolific](#), participated in the experiment (36 female, 44 male; ages 19–64 years, mean age 35).

Materials. The stimuli consisted of the 32 contexts of the main experiment, that is, 16 contexts that only consisted of the target context sentence, as in (1), and 16 contexts that consisted of the target context sentence followed by the additional sentences, as in (2). For the full set of target stimuli see Supplement B of the Github repository mentioned in footnote 2.

Each pair of contexts that shared the same target context sentence was paired with the same five questions: the target question and four other questions. [Figure A1](#) shows (the English translations of) the five questions for the context in (2). The target question Q1, here *Who parked in front of Benni’s bicycle?*, was hypothesized to be raised by the target context sentence. The questions Q2 *What could Benni have used instead of the bicycle?* and Q3 *Where did Benni go with the scooter?* were hypothesized to be raised by the second and third sentence of the context, respectively. (These sentences were absent when the context only consisted of the target context sentence.) We expected the target question to be relatively more expected immediately after the target context sentence (that is, in the one sentence context condition) than after the third sentence (that is, in the three sentence context condition). By contrast, we expected Q3 to be relatively less expected in the one sentence context condition than in the three sentence context condition. The other two questions served as control questions. We expected Q+ (here, *What did Benni do next?*), which was always a question about what the character did next, to receive high expectedness ratings in both context conditions. In contrast, we expected Q- (here, *How was the weather in Colombia?*), which was always an unrelated question, to receive low expectedness ratings in both context conditions.

The 32 contexts were distributed across two lists, so that each of the 16 target questions occurred only once per list and so that each list had eight stimuli in the one sentence context condition and eight in the three sentence context condition. Each list also included the same 14 filler stimuli: these consisted of two one sentence,

two three sentence, and 10 two sentence contexts followed by five questions each. The fillers were designed to distract from the structure of the target stimuli. Hence, they never used *als* ‘when’, but either used no subjunction or a different one, like *weil* ‘because’ or *bevor* ‘before’. The five questions of the filler stimuli were similar to the five questions of the target stimuli. They were designed to elicit ratings that range over the entire scale from ‘absolutely unexpected’ to ‘very expected’. The filler stimuli of the norming study can be found in Supplement B of the GitHub repository mentioned in footnote 2.

Procedure. Participants were first introduced to the family of six, as in the main experiment. On each trial, they read a context (called ‘text’) and were asked to rate, for each of the five questions, the likelihood that the next sentence of the text would be about that question. As shown in Figure A1, they gave their ratings on sliding scales labeled ‘(absolut unerwartbar)’ (absolutely unexpected) on the left end (coded 0) and ‘(sehr erwartbar)’ (very expected) on the right end (coded 100). Each question was formulated as an embedded question. Question order was randomized on each trial.

Data exclusion. The data from two participants was excluded from the analysis because their mean expectedness ratings were more than two standard deviations higher than the group mean of Q- or more than two standard deviations lower than the group mean on Q+. Data from 78 participants entered the analysis (35 female, 43 male, age range 19–64 years, mean age 35).

A.2 Results

Figure 4 plots the mean expectedness rating of the target question Q1, and the other questions, by context condition. As expected, the mean expectedness rating of the target question was higher immediately after the target context sentence (i.e., in the one sentence context condition) than when the target context sentence was followed by two additional sentences (i.e., in the three sentence context condition). This result was confirmed by a linear mixed effects model that predicted the expectedness ratings for the target question from context condition (reference level: one context sentence) and random by-participant and by-item intercepts and slopes.⁵ There was a main effect of context condition ($\beta = -29.3$, $SE = 2.2$, $t = -13.6$, $p < .001$), such that the target question was less expected in the three sentence context condition than in the one sentence context condition.

⁵ An item consisted of a context, followed by the target question and four other questions.

Read the text in the box. What do you expect that the next sentence will be about? Rate the following proposals and adjust the slider accordingly.

When Benni came into the shed his bicycle was blocked. (He couldn't get it out quickly enough. So he set off on the scooter.) ...

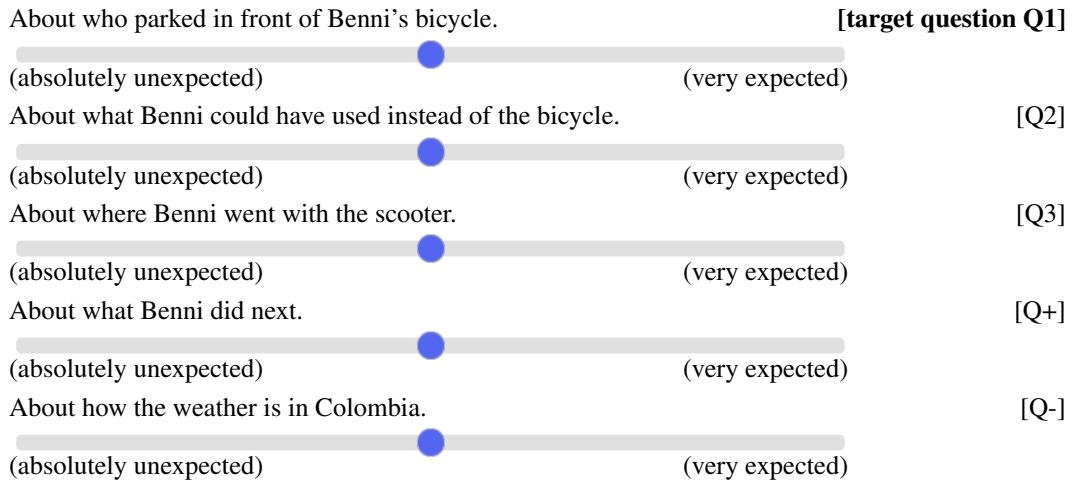


Figure A1 English translation of a sample trial of the question norming study (original in German). The sentences in brackets were only displayed in the three sentence condition. Question labels (target question Q1, Q2, Q3, Q-, Q+) were not presented to the participants.