

Chapter 6

Cleft sentences reduce information density in discourse

 Swantje Tönnis

University of Stuttgart

This paper develops a novel theoretical analysis of clefts as a discourse structuring device in written German (following Destruel & Velleman 2014, Tönnis 2021). The analysis is based on the assumption of an expectation-driven QUESTION UNDER DISCUSSION (QUD) model where addressees form a probability distribution over questions that an (ensuing) utterance is likely to answer (cf. Kehler & Rohde 2017). Tönnis (2021) argued that clefts are used to address relatively less expected QUDs, in contrast to canonical sentences, which address relatively expected QUDs. In this paper, I propose an information-theoretic take on the expectedness approach and combine it with the cleft's function to disambiguate focus, following the UNIFORM INFORMATION DENSITY (UID) hypothesis (Levy & Jaeger 2007). I hypothesize the following: The cleft in written German is used to reduce information density in order to achieve UID at the discourse level. The additional material in the cleft, compared to the canonical sentence, explicitly marks the addressed QUD (i.e., it disambiguates focus). This way it takes over information that is otherwise carried by the words of the canonical sentence and, thus, reduces information density. I argue that this reduction is only necessary when a less predictable QUD is addressed. Following Asr & Demberg (2015) and Demberg & Keller (2008), I define QUD SURPRISAL in order to integrate the expectedness of the addressed QUD into a model that predicts the choice of the conveyed message (cleft vs. canonical sentence). For the discussed example, and similar ones, the model makes correct predictions in contrast to previous analyses. Furthermore, aiming for UID in discourse provides a reason for why clefts tend to address relatively less expected QUDs, which was missing from Destruel & Velleman (2014) and Tönnis (2021).



1 Introduction

When deciding how to realize the next discourse move in a written text, a German speaker has the choice between a canonical sentence, as in (1), and an *es*-cleft, as in (2), among other options.¹

- (1) *Bo hat die Kekse gegessen.*
Bo has the cookies eaten
'Bo ate the cookies.'
- (2) *Es war Bo, der die Kekse gegessen hat.*
it was Bo who the cookies eaten has
'It was Bo who ate the cookies.'

This paper aims to predict and explain the choice of an author between a cleft and its canonical variant while taking the discourse context into account. The analysis is motivated by an example inspired by Tönnis (2021) which illustrates that the preference between the cleft and its canonical variant varies with the discourse context. In a discourse context such as (3), the cleft in (3b) is preferred over the canonical sentence in (3a).²

¹This paper is only concerned with written German. In the following, I will, thus, refer to the *author* of a text instead of the *speaker* of an utterance. In principle, the analysis to be presented should also apply to spoken German, but it would have to be adapted, which I leave as a topic for further research. For instance, one would have to incorporate the effect of overt focus marking by intonation in spoken German, which is more flexible than implicit prosody in written German. More details on implicit prosody are presented in Section 2.2.

²A further continuation which is frequently judged as equally acceptable as the cleft in (3) is the canonical sentence with a past perfect verb form instead of the present perfect, as in (i). These judgments have not yet been systematically tested in an experimental setting, though.

- (i) *Bo hatte die Kekse gegessen.*
Bo had the cookies eaten
'Bo had eaten the cookies.'

Tönnis & Tonhauser (2022: 678), however, provided the example in (ii) to show that clefts are not only acceptable in anteriority contexts. In this context, they judged the past perfect canonical sentence as equally dispreferred as the present perfect canonical sentence.

- (ii) *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 hatte mit Peter geflirtet.*
'Lena flirted with Peter.' / 'Lena had flirted with Peter.'

- (3) *Als Lena in die Kaffeepause kam, war der Keksteller schon leer. Sie fand auch keinen weiteren Keksteller. Also entschied sie sich zum Bäcker zu gehen.*

‘When Lena joined the coffee break, the plate of cookies was already empty. She couldn’t find any other cookies, either. So she decided to go to the bakery.’

- a. ? Bo hat die Kekse gegessen.
‘Bo ate the cookies.’
- b. Es war Bo, der die Kekse gegessen hat.
‘It was Bo who ate the cookies.’

In contexts such as (4), there is no clear preference between the cleft and the canonical sentence, or possibly a slight tendency towards the canonical sentence.

- (4) *Als Lena in die Kaffeepause kam, war der Keksteller schon leer.*

‘When Lena joined the coffee break, the plate of cookies was already empty.’

- a. Bo hat die Kekse gegessen.
‘Bo ate the cookies.’
- b. (?) Es war Bo, der die Kekse gegessen hat.
‘It was Bo who ate the cookies.’

Based on my theoretical model (Tönnis 2021), I predicted the contrast for the cleft and the canonical sentence in (3) and (4) by referring to the expectedness of the question under discussion (QUD) (Roberts 2012) addressed by the cleft/canonical sentence (here *Who ate the cookies?*). She argued that clefts are used in German to address relatively less expected QUDs in discourse while canonical sentences can only be used to address relatively expected QUDs.

Tönnis & Tonhauser (2022) provided empirical evidence for this contrast for 16 context pairs like (3) and (4) depending on question expectedness. In a norming study, they measured the expectedness rating for the question Q (here *Who ate the cookies?*) to be addressed next by the author in two conditions: 1-sentence contexts, like (4), which gave rise to question Q, and 3-sentence contexts, like (3), which contained two interfering sentences after the Q-raising sentence, giving rise to new, more prominent questions. They found that Q was significantly more expected to be addressed next in the 1-sentence contexts compared to the 3-sentence contexts. Furthermore, they collected relative preference ratings between the cleft and the canonical sentence in the two context conditions. They found that there was a significantly stronger preference towards the cleft in the

3-sentence condition, i.e., when the addressed question was less expected, compared to the 1-sentence condition, where they found no clear preference between the cleft and its canonical variant. The analysis presented in this paper will be illustrated by examples (3) and (4), but it is assumed to apply equally well to other (similar) examples, given Tönnis & Tonhauser's empirical evidence.

The analysis in Tönnis (2021) correctly predicts the judgments as found by Tönnis & Tonhauser (2022). However, it does not provide an explanation for why the cleft fulfills the function of addressing a less expected QUD. Hence, I propose to extend her analysis by referring to information density (e.g., Shannon 1948, Levy & Jaeger 2007) relative to the predictability of the addressed QUD. Employing UNIFORM INFORMATION DENSITY (UID) (e.g., Jaeger 2010), I hypothesize that a cleft is used to reduce information density in discourse in the case of addressing a less predictable/expected QUD. Furthermore, the cleft's function of disambiguating focus (as claimed by, e.g., De Vegaugh-Geiss et al. 2015, Tönnis et al. 2018) plays an important role in predicting the author's choice between the canonical sentence and its cleft variant. Assuming that the additional words of the cleft, compared to the canonical sentence, explicitly mark the QUD and, thus, take over the information of focus marking, clefting contributes to distributing information more uniformly and reduces information density per discourse move. Moreover, I argue that the information of focus marking only needs to be distributed further if the QUD that is addressed is relatively surprising. I implement this approach by defining QUD SURPRISAL, which measures the surprisal of a certain QUD to be addressed.

The information-theoretic approach to clefts can account for the (dis)preference of clefts in contexts such as (3) and (4). Importantly, it does not require any other effects of the cleft, such as exhaustivity or contrast, but still allows for such effects to occur. At the same time, the approach explains why the cleft is a suitable candidate to mark the reduced expectedness of the addressed QUD.

The paper is structured as follows: Section 2 presents prior analyses of clefts. Section 3 introduces a first take on an information-theoretic analysis of clefts in discourse. In Section 4, advantages, implications, and some possibilities of extensions of the proposed information-theoretic analysis are discussed. Section 5 concludes this chapter.

2 Prior analyses of clefts

Different purposes for choosing a cleft over a canonical sentence have been proposed in the literature, ranging from expressing exhaustivity (e.g., Horn 1981,

Büring & Križ 2013), contrast (e.g., Rochemont 1986), or a violation of expectations (e.g., Destruel & Velleman 2014, Tönnis 2021) to information structural functions, such as disambiguating focus (e.g., De Veugh-Geiss et al. 2015, Tönnis et al. 2018). The main focus of this paper lies on the latter two approaches, given that the information-theoretic take on clefts, presented in Section 3, is based on those. Furthermore, Tönnis & Tonhauser (2022) argued that the other approaches to clefts cannot account for the contrast in contexts such as (3) and (4).

2.1 Inferences of clefts

I will first present those analyses that are concerned with the different kinds of inferences that are conveyed by a cleft. The cleft is commonly claimed to have the meaning components in (5) (e.g., Velleman et al. 2012, Krifka & Musan 2012, De Veugh-Geiss et al. 2018, Destruel et al. 2019).

- (5) Es war Bo, der die Kekse gegessen hat. ('It was Bo who ate the cookies.')
- a. Prejacent: Bo ate the cookies.
- b. Indication of question: Who ate the cookies?
- c. Existential inference: Somebody ate the cookies.
- d. Exhaustivity inference: Nobody other than Bo ate the cookies.

The prejacent in (5a) is the at-issue content of the cleft that is assumed to be identical to the at-issue content of its canonical variant (see De Veugh-Geiss et al. 2018). Furthermore, several approaches, though for different reasons, have argued for the cleft indicating an implicit question (5b). According to Velleman et al. (2012), the cleft structure involves a cleft operator which requires a question to be present in the discourse. In the case of focus-background clefts, in which the clefted element, *Bo* in (5), is focused, focus marking also indicates the same implicit question (Krifka & Musan 2012). More details on the issue of focus marking in the cleft are given in Section 2.2. The existential inference, as in (5c), is commonly analyzed as a presupposition (e.g., Halvorsen 1978, Prince 1978, Percus 1997). The role of the exhaustivity inference of clefts, exemplified in (5d), is still debated. Some approaches (e.g., Szabolcsi 1981, Percus 1997, Büring & Križ 2013, Pollard & Yasavul 2015) analyzed the exhaustivity inference of clefts as a presupposition, i.e., a semantic inference. However, there are also approaches which analyzed it as a pragmatic inference (e.g., Horn 1981, De Veugh-Geiss et al. 2015, Onea 2019). De Veugh-Geiss et al. (2018) provided empirical evidence for the exhaustivity inference of German clefts being stronger than exhaustivity in canonical sentences with a focus on the constituent that forms the cleft pivot

in the cleft, such as (6). A stronger inference meant that violations of exhaustivity were less frequently accepted, and the truth of the exhaustivity inference was more frequently required for judging the respective sentence as true. Furthermore, the exhaustivity inference of clefts was found to be weaker than for exclusives, as in (7).

- (6) *BO_F hat die Kekse gegessen.*
BO_F has the cookies eaten
'BO_F ate the cookies.'
- (7) *Nur Bo hat die Kekse gegessen.*
only Bo has the cookies eaten
'Only Bo ate the cookies.'

De Vaugh-Geiss et al. (2018) concluded that exhaustivity in clefts is a not-at-issue pragmatic inference. Following Tönnis & Tonhauser (2022: 663), I argue that exhaustivity of clefts does not fully account for the preference between the cleft and its canonical variant in contexts such as (3) and (4). Example (8) represents a violation of exhaustivity which, nevertheless, is acceptable in contexts (3) and (4). More importantly, the cleft in (8) is still preferred over the canonical sentence (with or without exhaustivity violation) in context (3).

- (8) *Es war Bo, der die Kekse gegessen hat, und Lou auch.*
'It was Bo who ate the cookies and Lou as well.'

A further function which is frequently assigned to clefts is marking contrast. Rochemont (1986), for instance, argued that a cleft necessarily expresses contrastive focus while its canonical variant can express both contrastive and informational focus. According to Tönnis & Tonhauser (2022), however, the preference between the cleft and the canonical sentence in the above contexts (3) and (4) cannot be accounted for by referring to contrastivity. In particular, they claim that in the cleft in context (3) (English version repeated in (9)) there is no explicit alternative provided by the context (e.g., *Lou ate the cookies* or *Lou didn't eat the cookies*) to establish a contrast with Bo eating the cookies. Accordingly, the cleft would be predicted to be dispreferred.

- (9) *When Lena joined the coffee break, the plate of cookies was already empty. She couldn't find any other cookies, either. So she decided to go to the bakery.*
- a. It was Bo who ate the cookies.
- b. ? Bo ate the cookies.

Anticipating the information-theoretic approach to be presented in Section 3, I argue that the fact that Bo ate the cookies does not have to be particularly surprising for the cleft to be acceptable/preferred in (9), which is in line with the cleft not being used contrastively. Even if the author and the reader knew that Bo frequently finishes the cookies, the cleft (9a) would still be preferred over the canonical sentence (9b) in this context.

2.2 Clefts as a focus-disambiguating device

Some approaches to clefts proposed that clefts are used to disambiguate focus (e.g., De Vaughn-Geiss et al. 2015, Tönnis et al. 2018). In written German, where intonation cannot be used to mark focus, a canonical sentence is ambiguous with respect to focus. Example (10) illustrates some of the possible focus assignments for the canonical sentence (focus is marked by [...]F and the main accent is marked with capital letters).

- (10) a. Bo hat [die KEKse gegessen]F.
 b. Bo hat [die KEKse]F gegessen.
 c. [BO]F hat die Kekse gegessen.
 d. [Bo hat die KEKse gegessen]F.
 Bo has the cookies eaten
 ‘Bo ate the cookies.’

Fodor (2002), among others, pointed out that, even though intonation cannot be marked in written language, there is still evidence for implicit prosody during silent reading (for a comprehensive overview of implicit prosody, see Féry 2017: ch. 9.4). According to her implicit prosody hypothesis, the reader would assume the default prosody, which is “identical to the overt prosody for that sentence in a comparable context (i.e., same illocutionary force, focus structure, etc.)” (Fodor 2002: 115). Hence, when the sentences in (10) are interpreted in their discourse context the ambiguity is usually resolved, as in Krifka & Musan’s (2012: 11) examples, given in (11a) and (11b).

- (11) a. And then something strange happened. [A MEteorite fell down.]F
 b. Mary sat down at her desk. She [took out a pile of NOTES]F.

The discourse context in these examples (represented by the respective first sentence) makes the focus assignments and focus markings, which are indicated in the second sentence, the only reasonable ones.

The discourse context has often been claimed to affect focus assignment (e.g., Beaver & Clark 2008, Krifka & Musan 2012, Simons et al. 2017, Tönnis 2021). Krifka & Musan (2012), for instance, analyzed focus as marking (implicit) questions on the basis of the context. In other words, focus marking in an utterance *U* helps the reader to identify the question which is addressed by *U*. If focus is marked in an ambiguous way, as in the case of the canonical sentence in written German, this implicit question needs to be accommodated by the reader. This process and, thereby, also the implicit prosody strongly depend on contextual cues.

Focus-background clefts, in contrast, simplify the accommodation process of the implicit question: De Vaughn-Geiss et al. (2015) proposed that clefts structurally mark focus by backgrounding the content of the relative clause. Furthermore, they assumed that focus cannot project out of the cleft pivot. This leads to the unambiguous narrow focus marking in (12).³

- (12) Es war [Bo]_F, der die Kekse gegessen hat.
'It was [Bo]_F who ate the cookies.'

Tönnis et al. (2018) supported their claim by an extensive corpus study on written German, in which they annotated the grammatical function of the cleft pivot in German clefts. They found that there were relatively more subject clefts, such as (12), than non-subject clefts, such as the object cleft in (13), even when the generally higher subject frequency was taken into account.

- (13) Es waren die Kekse, die Bo gegessen hat.
'It was the cookies that Bo ate.'

Tönnis et al. (2018) argued that the subject preference in the pivot results from the fact that intonation cannot be used in written German to mark focus. Hence, the reader has to use cues from the context to accommodate the focus marking of each sentence. If there is no strong contextual cue, the reader will accommodate default focus marking, namely object focus or wide focus. However, if the author

³In some cases, focus is even ambiguous in clefts. Velleman et al. (2012: 442) presented the example in (i), where only part of the cleft pivot is focused.

(i) It was John's [eldest]_F daughter who liked the movie.

Similar clefts can be found in German. In such cases, clefts only reduce the possible focus readings compared to the canonical sentence in written German. The examples discussed in this paper always contain cleft pivots which consist of one word, which always leads to a clear disambiguation. I leave the investigation of possible effects due to narrow focus inside of the pivot to future research.

wants to express narrow subject focus, she could use a cleft to shift default focus to the subject position (e.g., Reinhart 1995, Szendrői 1999). Since this extra marking is necessary for subjects but not for objects, Tönnis et al.'s (2018) approach correctly predicted a higher frequency of subject clefts compared to object clefts, and concluded that clefts are used to disambiguate focus in written German.⁴

The information-theoretic analysis proposed in this paper adopts the idea that clefts disambiguate focus, i.e., they explicitly indicate the addressed question. The question arises when focus disambiguation is necessary, and accordingly, why there is not a much higher cleft frequency compared to canonical sentences in written German. When is it not enough to assume default implicit prosody? I argue that the need to disambiguate focus depends on the expectedness of the question under discussion (QUD) that is addressed by the respective sentence, which will be the topic of the next subsection.

2.3 Expectation-based analyses to clefts

Destruel & Velleman (2014) claimed that clefts can not only be used to mark a contrast to some content mentioned in the discourse context (as in Rochemont 1986), but that they can also be used to mark that the discourse develops into an unexpected direction. Spelling out this idea, I argued in Tönnis (2021) that German cleft sentences, unlike their canonical variants, are used to address relatively less expected questions under discussion (QUDs) (following Roberts 2012). Canonical sentences, I claimed, address relatively expected QUDs. Empirical support for this claim is given in Tönnis & Tonhauser (2022), also presented in Section 1 of this paper.

The underlying assumption is that, at each point of a discourse, the interlocutors have certain expectations about which QUD is likely to be addressed next. These expectations are modeled as a probability distribution over questions to be addressed by the ensuing utterance/sentence (following Kehler & Rohde 2017), which assigns a probability to each possible question with respect to how likely it is to be addressed next. This probability distribution is affected by each new utterance/sentence of a text. For instance, a sentence containing an implicit causality verb, such as *admire*, would raise the probability mass on certain *why*-questions,

⁴Note that Tönnis et al.'s (2018) analysis is only applicable to written German since speakers can freely use intonation on most words and syntactic positions in German. Using intonation is preferred over the more complex cleft structure to mark focus or prominence in spoken German. The consequence is, as Tönnis et al. (2018) claimed, that clefts are less frequently used in spoken German compared to written German, a claim for which, to my knowledge, there is no thorough empirical evidence yet.

as exemplified in (14). Empirical evidence for this claim was provided by Westera & Rohde (2019). In a question elicitation experiment, they showed that significantly more *why*-questions were elicited for implicit causality contexts compared to other contexts.

(14) Lou admired Bo.

→ *Why did Lou admire Bo?* becomes more likely to be addressed in the ensuing utterance than before (14) was uttered.

In Tönnis (2021), I argued that, based on the distribution of QUDs to be addressed next, the addressee needs to decide whether to accept a sentence as a relevant discourse move (as described by Roberts 2012). Note that the definition of QUD in Tönnis (2021) diverges from Roberts' (2012) with respect to the assumed hierarchy between QUDs. In Roberts' (2012) approach, QUDs are organized on a stack, and in most cases, only addressing the top-most question or a sub-question thereof constitutes a relevant discourse move. In Tönnis (2021), QUDs are not strictly organized hierarchically and a wider variety of QUDs are acceptable to be addressed. Whether addressing a QUD constitutes a relevant discourse move mainly depends on the expectedness/probability of this QUD to be addressed next. Tönnis (2021: 286) introduced a threshold of expectedness for QUDs. A discourse move is only relevant if, among other conditions, the expectedness of the question it addresses exceeds this threshold. What the actual value of this threshold is is still to be determined empirically. For this paper, I assume such a threshold exists and that certain constellations push the expectedness value of a question above or below this threshold.

In the following, I present some examples illustrating the approach in Tönnis (2021). In context (14), for instance, the expectedness value of the question Q1: *Why did Lou admire Bo?* is assumed to exceed the threshold. Accordingly, the second sentence in (15) would be accepted as a relevant discourse move because it addresses Q1.

(15) Lou admired Bo. She loved the way he sung Queen's Bohemian Rhapsody.

Relatively expected questions are assumed to remain above the threshold until answered. However, an intervening sentence, which itself gives rise to a question, can lower the expectedness value of a previously raised, unanswered question, as illustrated in example (16).

(16) Lou admired Bo. But Bo had a secret.

The second sentence in (16) strongly increases the probability mass on the question Q2: *What was Bo's secret?*, which automatically reduces the expectedness of the previous question Q1: *Why did Lou admire Bo?* (because of the probabilities

of all questions adding up to 1). Therefore, the continuation in (17a), which addresses Q2, should be more acceptable than the continuation addressing Q1 in (17b), which is the case in my judgment. For similar examples, Tönnis & Tonhauser (2022) presented empirical evidence showing that the expectedness of an unanswered question that was raised by a sentence decreased after intervening sentences raised new questions (see Section 1 of this paper).

(17) *Lou admired Bo. But Bo had a secret.*

- a. His famous cover version of Queen's Bohemian Rhapsody was fake.
- b. Lou loved the way he sung Queen's Bohemian Rhapsody.

As mentioned above, I argued in Tönnis (2021) that a cleft sentence addresses relatively less expected QUDs in discourse. This means that it requires a lower threshold of question probability, compared to an unclefted sentence, in order to qualify as an acceptable discourse move. In particular, I assumed an expectedness value for the question addressed by a cleft which lies between this lower threshold for clefts and the threshold for canonical sentences.

The contrast between the cleft and the canonical sentence in the two contexts, repeated in (18), is correctly predicted by this approach.

(18) *Als Lena in die Kaffeepause kam, war der Keksteller schon leer. (Sie fand auch keinen weiteren Keksteller. Also entschied sie sich zum Bäcker zu gehen.)*

'When Lena joined the coffee break, the plate of cookies was already empty. (She couldn't find any other cookies, either. So she decided to go to the bakery.)'

- a. Bo hat die Kekse gegessen.
'Bo ate the cookies.'
- b. Es war Bo, der die Kekse gegessen hat.
'It was Bo who ate the cookies.'

The canonical sentence in (18a) is only acceptable in the shorter context, i.e., without the sentences in brackets. This context sentence evokes the question *Who ate the cookies?*, which plausibly raises the probability of this question to be addressed above the threshold. In this case, (18a) is predicted to be an acceptable discourse move and the cleft to be dispreferred.⁵

⁵Empirical evidence by Tönnis & Tonhauser (2022) showed that the cleft was not dispreferred, but that there was no clear preference between the cleft and the canonical sentence in contexts such as the short context in (18). Tönnis & Tonhauser explained this by referring to Tönnis' (2021) extended definition of acceptability of clefts/canonical sentences which specifies an overlapping region of expectedness, where both the cleft and its canonical variant are acceptable. In this case, Tönnis' account would correctly predict that both the cleft and the canonical sentence would be acceptable.

In the longer context (full context in (18)), the cleft in (18b) is preferred since the intervening two sentences give rise to new questions, such as *What did Lena get at the bakery?*. Such questions reduce the expectedness of the question addressed by the cleft (*Who ate the cookies?*), just as in example (17). If we assume that the expectedness is pushed below the threshold for canonical sentences but not below the threshold for clefts, the cleft can correctly be predicted to be acceptable while the canonical sentence is predicted to be dispreferred.

Subsection 2.2 pointed out that focus marking is ambiguous in canonical sentences for written German. The German cleft structure, however, makes focus explicit. I argue that it is not enough to analyze the cleft as a focus-disambiguating device in order to explain the preferences between German clefts and their canonical variants in discourse. What also needs to be explained is when it is necessary to disambiguate focus. I argue that the author's wish to disambiguate focus is only present when she wants to address a relatively less expected QUD. Disambiguating focus should only be necessary if it was not yet obvious in any way which QUD could be addressed next.

3 Towards an information-theoretic analysis of German clefts

Summing up previous insights, an analysis of clefts as addressing a relatively less expected QUD correctly predicts the preference between the cleft and its canonical variant in written German. What is still missing is an explanation for why the expectedness threshold is lower for clefts compared to canonical sentences. In the following, I present a proposal of an information-theoretic approach which aims to provide this explanation by combining the idea of clefts disambiguating focus and the idea of clefts addressing less expected QUDs. Note that this proposal still needs to be tested empirically, which exceeds the scope of this paper.

At first glance, the cleft just seems to be the syntactically more marked structure, which is an indication of an additional or a more complex function on some linguistic level. For instance, a more complex definite description, such as *the neighbor's dog*, is usually used to refer to a less salient antecedent in discourse than a less complex pronoun, such as *it* (e.g., Gundel et al. 1993). In the same way, I argue, that the more complex cleft addresses a less salient QUD than the less complex canonical sentence, and it does so by explicitly marking this QUD (via focus marking).

In the information-theoretic approach to language, language production is assumed to be efficient within the bounds of grammar (Jaeger 2010). The most

efficient way involves (i) distributing information uniformly across the speech signal, and (ii) keeping information density (i.e., the amount of information per unit, e.g., per word) close to the channel capacity (Genzel & Charniak 2002, Levy & Jaeger 2007). Jaeger (2010) spelled out (i) as the UNIFORM INFORMATION DENSITY (UID) hypothesis:

Within the bounds defined by grammar, speakers prefer utterances that distribute information uniformly across the signal (information density). Where speakers have a choice between several variants to encode their message, they prefer the variant with more uniform information density (*ceteris paribus*). (Jaeger 2010: 25)

The two variants to encode the same message that will be relevant for the analysis of this paper are the cleft and the canonical sentence. The channel capacity in (ii) represents the information rate, i.e., a fixed amount of information per unit, that no unit should strongly deviate from (see Genzel & Charniak 2002). In other words, no unit, for instance, a word, should convey much more or much less information than the other units of the same category.

Information is understood in the sense of Shannon information (Shannon 1948), also called surprisal, for a unit of a signal. The information I of a unit, such as a word or a sentence, is defined as in (19).

$$(19) \quad I(\text{unit}) = \log \frac{1}{p(\text{unit})} = -\log p(\text{unit})$$

This means that the higher the probability p of a unit, the lower is the information I (or the surprisal) of that unit. For instance, the more expected a word is the less new information it conveys. The information or surprisal of a unit often involves the conditional probability of the unit, for example, conditioned by the probability of the preceding units (Levy & Jaeger 2007), or the probability of possible syntactic trees (Demberg & Keller 2008) or discourse relations (Asr & Demberg 2015). The reasoning for an information-theoretic approach to clefts follows the reasoning used in Levy & Jaeger (2007) for a case of syntactic reduction, and builds on Asr & Demberg's (2015) approach, who defined discourse relational surprisal.

Levy & Jaeger (2007) predicted the syntactic variation observed for relative clauses with respect to the presence/absence of the relative pronoun. They hypothesized that the relative pronoun *that* is usually omitted when it would otherwise precede a relatively expected word, such as *you* in their example, repeated in (20).

- (20) How big is the family (that) you cook for? (Levy & Jaeger 2007: 851)

They argue that the syntactic reduction is a consequence of UID at the sentence level. Both versions of (20), with and without the relative pronoun, express the same informational content, but the information is distributed differently. When the relative pronoun is not present, the first word of the relative clause, here *you*, fulfills two functions: It conveys its semantic content and it marks the onset of the relative clause. When the relative pronoun is present, these two functions are split up between the relative pronoun and the noun phrase *you*. According to the UID hypothesis, the relative pronoun is predicted to be dropped to avoid a trough in information density in case the surprisal/information of the word *you* is low, while it should be inserted to avoid a peak on *you* in case it is relatively surprising in its context. Levy & Jaeger (2007) found empirical evidence for this claim in a corpus study. I employ the same reasoning for the variation between the cleft and the canonical sentence, and hypothesize that the cleft is used to reduce information density at the discourse level when a relatively less expected/more surprising QUD is addressed.

First of all, note that the cleft and the canonical sentence express the same information when the same constituent is focused. Example (21) illustrates this for a subject cleft and a canonical sentence with subject focus marked by intonation.

- (21) BO ate the cookies./It was Bo who ate the cookies.
- a. At-issue content/prejacent: Bo ate the cookies.
 - b. Indication of question (focus): Who ate the cookies?
 - c. Existential inference: Somebody ate the cookies.
 - d. Exhaustivity inference: Nobody other than Bo ate the cookies.

As mentioned in Section 2.1, the meaning components are weighted differently for the two sentences. De Vaugh-Geiss et al. (2018), for instance, showed that the exhaustivity inference is stronger for clefts than for canonical sentences. For the analysis presented in this paper, these gradual differences will not be taken into account. Instead, it uses the simplification that clefts and their canonical variants express the same informational content. Here, I focus on the semantic content (21a) and the information structural contribution (21b).

As mentioned in Section 2.2, focus, and thereby the implicit QUD, is not overtly marked in many sentences of written German since the author cannot indicate intonation.⁶ The focus is ambiguous and the implicitly indicated question must,

⁶In some types of text, for example, chat messages, the author can use capital letters to mark intonation/emphasis. In such cases, this analysis does not apply. I consider cases of written German where using capital letters for emphasis is not common.

thus, be inferred from contextual cues, as example (11) showed. Clefts, in contrast, disambiguate focus and, thus, explicitly indicate the QUD.

From an information-theoretic perspective, all words of the canonical sentence carry both the semantic content as well as the focus.⁷ In the cleft, the same information is distributed onto more words. I argue that, in an information-theoretic sense, the words introduced by clefting (*es* ‘it’, *war* ‘was’, and *der* ‘who’) take over the information of focus (indicating the QUD) since clefting creates a syntactic structure that explicitly separates the focus from the background (see also É. Kiss 1998).⁸

The question arises why authors do not always want to disambiguate focus in written German, which would lead to a much higher frequency of clefts than actually observed in written German. I argue that the reason is efficiency, which information theory is well-suited to capture. For the choice between the cleft and the canonical sentence, I argue that reduction of information density by clefting is only necessary if the focus is difficult to identify, i.e., difficult to accommodate. This is the case if the QUD which the author wants to address is relatively less expected or more surprising. In this case, using a canonical sentence would exceed the channel capacity, i.e., too much information per word. The author is predicted to use a cleft. If the QUD the author intends to address was strongly expected, the words of the canonical sentence would not have to carry much extra information, and no extra marking by clefting would be necessary. Hence, the canonical sentence would be the preferred option.

In order to implement the influence of conveying the implicit QUD on information density, the discourse context needs to be incorporated into the calculation of information/surprisal. In particular, information density must be measured depending on the probability of the addressed QUD. Asr & Demberg (2015) presented a similar approach in their definition of DISCOURSE RELATIONAL SURPRISAL. Discourse relational surprisal describes the effect of a word on the belief distribution of discourse relations by comparing the belief distribution before and after the word. Asr & Demberg (2015) were particularly interested in the relational surprisal of discourse connectives, such as *because* or *therefore*. Relational surprisal is small if the connective did not have a strong effect on the distribution of discourse relations, i.e., the relation marked by the connective was likely

⁷The analysis does not hinge on focus to be conveyed by all of the words in a sentence. If focus is just conveyed by parts of the sentence, the same reasoning applies.

⁸The words *es war* also have a local effect on surprisal within the cleft sentence, in the sense of preparing the reader for what is to come in the cleft pivot (thanks to Lisa Schäfer for this comment). I discuss one such example in (42) in Section 4. For my main analysis, I make the simplification of assuming the words needed for clefting jointly have the function to convey focus.

even before the connective was uttered. In example (22), the connective *because* does not strongly change the distribution over discourse relations because the implicit causality verb *admire* already raised the probability mass attributed to the discourse relation CAUSE.

(22) Lou admired Bo, (because) he was such a good singer.

As indicated by the brackets in (22), the connective can be dropped in such a case. This can be explained by information theory: If the discourse relational surprisal of a connective is small, the connective should be dropped in order to avoid a trough in information density (Levy 2008, Demberg & Keller 2008, Asr & Demberg 2012, 2015).

I propose to adjust Asr & Demberg's (2015) approach and define QUD SURPRISAL, which affects the author's choice of how to encode her next message based on the expectedness of the addressed QUD. QUD surprisal captures this by comparing the two question distributions D_0 and D_w . D_0 is the previous question distribution, which speaker and addressee share given their previous conversation. It is based on the linguistic discourse context, prior probabilities for certain questions to be addressed, and the common ground.⁹ D_w is the question distribution after the first word(s) of the next utterance.

For illustration, assume the simplified question distribution D_0 , given in (23), in a discourse context.

(23) $D_0 =$
 $Q1 \rightarrow 0.1$
 $Q2 \rightarrow 0.2$
 $Q3 \rightarrow 0.7$

Consider (24) as an example discourse context. Then, Q3 could be *Why did Lou admire Bo?*, given that it is a relatively expected question in this context.

(24) Lou admired Bo.

If the author's next word in (24) was *because*, the probability of question Q3 would increase. Accordingly, a possible question distribution D_w after (24) + *because* is given in (25).

(25) $D_w = D_{\text{because}}$
 $Q1 \rightarrow 0.05$

⁹This is compatible with different versions of common ground (management) (e.g., Chafe 1976, Krifka 2008), which affects the probabilities of D_0 . In this paper, I will not be concerned with how exactly D_0 is affected by the common ground.

$$\begin{aligned} Q2 &\rightarrow 0.1 \\ Q3 &\rightarrow 0.85 \end{aligned}$$

If the next word of the author in (24) was *nevertheless* instead, the probability of question Q3 would decrease. In this case, a possible question distribution D_w after (24) + *nevertheless* could be (26).

$$\begin{aligned} (26) \quad D_w &= D_{\text{nevertheless}} \\ Q1 &\rightarrow 0.3 \\ Q2 &\rightarrow 0.6 \\ Q3 &\rightarrow 0.1 \end{aligned}$$

In order to measure the difference between the previous distribution D_0 and the distribution D_w after one or more additional words, I define QUD surprisal S_{QUD} of a unit w as the Kullback-Leibler divergence D_{KL} , also called relative entropy, between the two distributions D_0 and D_w as in (27) (following Demberg & Keller 2008: 195). The set possQ is the set of all possible QUDs that could be addressed, which has no restrictions apart from each question being syntactically well-formed. In our simplified example, possQ is the set containing the questions Q1, Q2, and Q3.

$$(27) \quad S_{\text{QUD}}(w) = D_{KL}(D_0 || D_w) = \sum_{q \in \text{possQ}} D_0(q) \log \frac{D_0(q)}{D_w(q)}$$

$S_{\text{QUD}}(w)$ yields a relatively high value if the previous distribution D_0 differs strongly from the distribution D_w . $S_{\text{QUD}}(w)$ yields a relatively low value if the two distributions are similar. In my example, $S_{\text{QUD}}(\text{nevertheless})$ is higher than $S_{\text{QUD}}(\text{because})$ because the previous distribution D_0 in (23) differs more strongly from the distribution $D_{\text{nevertheless}}$ in (26) than from the distribution D_{because} in (25). This is shown in (28), where possQ is the set $\{Q1, Q2, Q3\}$ and D_0 , D_{because} , and $D_{\text{nevertheless}}$ are the respective distributions presented in (23–26).

$$\begin{aligned} (28) \quad S_{\text{QUD}}(\text{nevertheless}) &= \sum_{q \in \text{possQ}} D_0(q) \log \frac{D_0(q)}{D_{\text{nevertheless}}(q)} \\ &> S_{\text{QUD}}(\text{because}) = \sum_{q \in \text{possQ}} D_0(q) \log \frac{D_0(q)}{D_{\text{because}}(q)} \end{aligned}$$

The consequence of this outcome would be that the connective *nevertheless* should be inserted in our example case since it strongly affects the previous question distribution. The connective *because* could be dropped since it does not have a

strong effect on the previous question distribution. In their corpus study, Asr & Demberg (2015) found that *nevertheless* was more frequently expressed explicitly than *because*.

Coming back to the author's decision between the cleft and the canonical sentence, I assume that the author considers that the addressee has some uncertainty about which question the author wants to address with an utterance. From the addressee's perspective, a canonical sentence in written German always leads to some uncertainty about which question the author wants to address given focus ambiguity (see (10) for some examples of different possible focus assignments for the same sentence). The cleft, in contrast, reduces or, in most cases, eliminates this uncertainty because of a more explicit focus marking, i.e., more explicitly marking the QUD.

Analogously to the discourse connectives above, I assume that QUD surprisal of clefting is low when the cleft addresses a relatively expected question. Hence, the words used for clefting should be dropped, in order to avoid a trough in information density. When the cleft addresses a relatively unexpected question, the QUD surprisal of clefting is relatively high and the words used for clefting should not be dropped, in order to distribute surprisal more uniformly.

For calculating the QUD surprisal of clefting, I make the simplification of treating clefting (i.e., the words *it*, *was*, and the relative pronoun) as an operator op_{cleft} that applies to the canonical sentence, following approaches like the one by Velleman et al. (2012). The QUD surprisal of op_{cleft} , as illustrated in (29), compares the distribution D_{can} after having encountered the canonical sentence to the distribution D_{cleft} after adding the cleft operator to the canonical sentence.

$$(29) \quad S_{\text{QUD}}(op_{\text{cleft}}) = D_{KL}(D_{\text{can}} || D_{\text{cleft}}) = \sum_{q \in \text{posQ}} D_{\text{can}}(q) \log \frac{D_{\text{can}}(q)}{D_{\text{cleft}}(q)}$$

The difference between these two distributions is relatively small when the addition of the cleft operator to the canonical sentence in the context does not have a strong effect on the question distribution. This would mean that the question marked by clefting was also a rather likely one in the context. If clefting affected the distribution to a stronger degree, $S_{\text{QUD}}(op_{\text{cleft}})$ would be relatively large.

Using the above example (English translations repeated in (30) and (32)), I will demonstrate how this approach can explain the preference between the cleft and the canonical sentence. Consider first the previous question distributions in the two contexts, which describe the expectedness values of each question before the cleft/canonical sentence is added. I argue that this question expectedness is one of the two crucial aspects one needs to incorporate to explain the choice between

the cleft and the canonical sentence (the other aspect being the cleft's function of focus-disambiguation). In the shorter context, repeated in (30), the QUD intended to be addressed by the author (Q1: *Who ate the cookies?*) is relatively expected, i.e., easy to accommodate for the addressee.

- (30) *When Lena joined the coffee break, the plate of cookies was already empty.*
 a. Bo hat die Kekse gegessen. ('Bo ate the cookies.')
 b. ? Es war Bo, der die Kekse gegessen hat. ('It was Bo who ate the cookies.')

Hence, the previous distribution D_0 , which is based on the non-linguistic context and the context sentence in (30), can be assumed to assign a relatively large amount of the probability mass to Q1. A plausible, but simplified, D_0 for context (30) is provided in (31).

- (31) $D_0 =$
 Q1: Who ate the cookies? $\rightarrow 0.25$
 Q2: What did Bo eat? $\rightarrow 0.05$
 Q3: What happened then? $\rightarrow 0.3$
 Q4: What did Lena eat? $\rightarrow 0.3$
 \vdots
 Qn: ...

The situation looks different in the slightly longer context, repeated in (32).

- (32) *When Lena joined the coffee break, the plate of cookies was already empty. She couldn't find any other cookies, either. So she decided to go to the bakery.*
 a. ? Bo hat die Kekse gegessen. ('Bo ate the cookies.')
 b. Es war Bo, der die Kekse gegessen hat. ('It was Bo who ate the cookies.')

Example (33) illustrates a plausible and simplified question distribution D_0 , given the context in (32).

- (33) $D_0 =$
 Q1: Who ate the cookies? $\rightarrow 0.05$
 Q2: What did Bo eat? $\rightarrow 0.05$
 Q3: What happened then? $\rightarrow 0.4$
 Q4: What did Lena eat? $\rightarrow 0.4$
 \vdots
 Qn: ...

The examples (34–36) illustrate how the question distribution D_0 changes for the addressee when (i) the canonical sentence is added to context (30) (D_{can}), and (ii) the cleft operator is then added to the canonical sentence (D_{cleft}). The examples (37–39) illustrate this for context (32). All the distributions are simplified, but aim to represent reasonable probability ratios between the questions Q1–Q4. Example questions Q1–Q3 are chosen to represent questions that can be addressed by the canonical sentence *Bo ate the cookies* (with the matching focus), Q1 is, furthermore, chosen because it is addressed by the cleft *It was Bo who ate the cookies*. The question Q4 is an example of a question that is likely to be addressed in both contexts, but could not be addressed by the cleft/canonical sentence. In examples (34–39), every step is illustrated in more detail.

Examples (34) and (37) repeat the previous distributions introduced above for the two contexts. After having read the canonical sentence, I assume that only the questions which are associated with one of the possible focus markings of the canonical sentence, here Q1–Q3, receive probability mass. The previous probability ratio between these questions is maintained while the probability of all the other questions drops to 0.¹⁰ This step is illustrated in the pairs (34)/(35) and (37)/(38), which exemplify how the previous distribution D_0 in each context differs from the distribution D_{can} after having read the canonical sentence.

The pairs (35)/(36) and (38)/(39) illustrate how the distributions change after the cleft operator has been applied to the canonical sentence. Since the cleft disambiguates focus, only one question is left to be addressed by it, Q1 in our examples. All the other questions receive a probability of 0 (or close to 0).

(34) Distribution after context (30) (one sentence)

$D_0 =$
 Q1: Who ate the cookies? $\rightarrow 0.25$
 Q2: What did Bo eat? $\rightarrow 0.05$
 Q3: What happened then? $\rightarrow 0.3$
 Q4: What did Lena eat? $\rightarrow 0.3$
 \vdots
 Qn: ...

(35) Distribution after canonical sentence

$D_{\text{can}} =$
 Q1: Who ate the cookies? $\rightarrow 0.42$
 Q2: What did Bo eat? $\rightarrow 0.08$

¹⁰Strictly speaking, the probability would be close to 0, not identical to 0. For reasons of simplicity, we assume it to be 0.

Q3: What happened then? $\rightarrow 0.5$

Q4: What did Lena eat? $\rightarrow 0$

\vdots

Qn: ...

(36) Distribution after clefting

$D_{\text{cleft}} =$

Q1: Who ate the cookies? $\rightarrow 1$

Q2: What did Bo eat? $\rightarrow 0$

Q3: What happened then? $\rightarrow 0$

Q4: What did Lena eat? $\rightarrow 0$

\vdots

Qn: ...

(37) Distribution after context (32) (three sentences)

$D_0 =$

Q1: Who ate the cookies? $\rightarrow 0.05$

Q2: What did Bo eat? $\rightarrow 0.05$

Q3: What happened then? $\rightarrow 0.4$

Q4: What did Lena eat? $\rightarrow 0.4$

\vdots

Qn: ...

(38) Distribution after canonical sentence

$D_{\text{can}} =$

Q1: Who ate the cookies? $\rightarrow 0.1$

Q2: What did Bo eat? $\rightarrow 0.1$

Q3: What happened then? $\rightarrow 0.8$

Q4: What did Lena eat? $\rightarrow 0$

\vdots

Qn: ...

(39) Distribution after clefting

$D_{\text{cleft}} =$

Q1: Who ate the cookies? $\rightarrow 1$

Q2: What did Bo eat? $\rightarrow 0$

Q3: What happened then? $\rightarrow 0$

Q4: What did Lena eat? $\rightarrow 0$

\vdots

Qn: ...

According to the definition in (29), the QUD surprisal of the cleft operator, $S_{\text{QUD}}(\text{op}_{\text{cleft}})$, in contexts (30) and (32) is calculated by comparing (35) to (36) and (38) to (39), respectively. The example values already indicate that the change from (38) to (39) is more drastic than from (35) to (36). In other words, the calculated value $S_{\text{QUD}}(\text{op}_{\text{cleft}})$ is higher in context (32) than in context (30). This means that, given the effect of clefting on the question distribution, it could be dropped in (30) because it is less surprising. The canonical sentence is sufficient. In (32), clefting should be inserted in order to avoid a peak in information density on the words of the canonical sentence and, thus, to distribute information more uniformly. This result is in line with what was observed by Tönnis & Tonhauser (2022) for the choice between the cleft and the canonical sentence in contexts such as (30) and (32).

These examples illustrated how the QUD surprisal of clefting is affected by the cleft's function of explicitly marking the QUD (i.e., disambiguating focus) as well as by previous expectedness values of questions. The respective last step in (36) and (39) shows the effect of focus disambiguation on the question distribution, i.e., assigning probability 1 to question Q1. However, this step only strongly affected the question distribution D_{can} when the question addressed by the cleft was not already relatively likely in D_{can} . Hence, I argue both of these aspects are relevant in order to capture the preference between a canonical sentence and its cleft variant in a context.

4 Discussion

The information-theoretic take on clefts in written German makes the correct predictions for the author's choice between the cleft and its canonical variant in contexts such as (30) and (32), just as previous discourse-dependent analyses did (e.g., Destruel & Velleman 2014, Tönnis 2021). By introducing QUD surprisal, it provides a formal analysis that can account for the discourse context dependency of this choice: A higher QUD surprisal of the clefting operator leads to a cleft while a lower QUD surprisal leads to a canonical sentence.

A huge advantage of the information-theoretic approach presented in this paper is that, besides predicting the choice between the cleft and the canonical sentence, it also provides an explanation for why the cleft is a good candidate to address relatively less expected QUDs: Clefting contributes to establishing uniform information density in discourse. In case of addressing a relatively less expected QUD, the cleft makes this question explicit and, thereby, distributes the information onto more words compared to the canonical sentence.

This explanation was missing from previous discourse-dependent analyses. Those analyses struggled to explain which aspect of the cleft caused it to behave differently in discourse than plain canonical sentences. It is not the pragmatic inferences (exhaustivity inference, existential inference), which were not affected differently in the two discourse contexts in example (30) and (32), and can, therefore, not affect the preferences, as discussed in Section 2.1 above.

One might argue that assuming that clefts are used for the purpose of focus disambiguation is already sufficient to explain the preferences between clefts and canonical sentences in discourse. I argue that a cleft does indeed help the reader to accommodate the QUD. However, it must be explained when exactly focus disambiguation is necessary, and QUD surprisal provides a measure for that: Focus disambiguation is only necessary when the author intends to address a QUD that is still relatively less expected once the canonical sentence is added. This can only occur if it was also relatively less expected in the previous distribution (before the canonical sentence was added). In such a case, focus disambiguation, modeled by assigning probability 1 to the respective question, has a strong effect on the question distribution after the canonical sentence. Therefore, the QUD surprisal of clefting is relatively high, and UID requires a more explicit marking of the QUD in order to avoid a peak in information density on the words of the canonical sentence. If the author wanted to address a relatively expected question, the QUD surprisal of the clefting operator would be relatively low given that focus disambiguation would not strongly affect the question distribution after the canonical sentence was added.

Another benefit of my approach is that it treats clefts on a par with other discourse structuring devices, which I claim could have the same effect of marking relatively less expected QUDs. One such device could be the discourse marker *übrigens* ('by the way'), as illustrated in (40).

- (40) *When Lena joined the coffee break, the plate of cookies was already empty. She couldn't find any other cookies, either. So she decided to go to the bakery.*
Übrigens Bo hat die Kekse gegessen.
 By the way Bo has the cookies eaten
 'By the way, Bo ate the cookies.'

Adding the discourse marker *übrigens* ('by the way') also makes the continuation acceptable while the plain canonical sentence is unacceptable in this context. In this context, the marker *übrigens* ('by the way') makes explicit that a relatively less expected QUD is going to be addressed. This is another way to reduce information density at the discourse level, which should be analyzed parallel to the cleft (only that focus disambiguation does not play a role here). Previous approaches which treated clefts on a par with structurally similar constructions,

such as definite descriptions (Percus 1997), cannot account for this parallel behavior.

Moreover, the information-theoretic approach to clefts can predict why there are less clefts in spoken German than in written German, as claimed by Tönnis et al. (2018) (based on their own and informants' judgments). In spoken German, there is no or less focus ambiguity in the canonical sentence since intonation can freely be used in German to express focus. Instead of clefting, the speaker would, therefore, rather use the canonical sentence with the main accent on the subject, as in (41).

- (41) *BO_F hat die Kekse gegessen.*
BO_F has the cookies eaten
'BO_F ate the cookies.'

Since subject focus is not ambiguous, a probability of 1 would be assigned to the question *Who ate the cookies?* after the canonical (41) already, and the QUD surprisal of clefting would then be very low. Hence, my analysis would frequently predict to drop clefting in spoken German.

So far, the analysis presented in this paper does not make any predictions about how each single word of the cleft or canonical sentence affects QUD surprisal, given that I assumed clefting to be just one operator. For discourse connectives, such as *because* or *übrigens* ('by the way'), QUD surprisal can be calculated equally well as, for example, relational surprisal by Asr & Demberg (2015). More complex discourse structuring devices such as the cleft are more challenging if one intends to calculate QUD surprisal incrementally. Of course, one would eventually want to be able to account for the fact that the cleft is not processed by first reading the canonical sentence and only afterwards encountering the cleft operator. I leave this issue for future research.

An anonymous reviewer pointed out that, instead of treating the addressed QUD as relatively less expected, it could be just the cleft pivot, *Bo* in the above example, that is surprising. As mentioned in Section 2.3, this does not apply to the kind of examples discussed in this paper. However, there might be other uses of clefts where UID does not apply at the discourse level but at the sentence level, as indicated in (42).

- (42) *Gestern war ich in der Kirche. Es waren aber nicht nur die üblichen Verdächtigen da.*
'Yesterday I was at church. But not only the usual suspects were present.'
Es war der Papst, der uns begrüßt hat.
'It was the pope who greeted us.'

In this example, it is not particularly unexpected that the QUD *Who greeted you/us?* is addressed at this point. However, one could assume that the word *pope* is surprising in this context. Therefore, this example could be explained by assuming that the words used for clefting are inserted to reduce information density at the sentence level instead of the discourse level. Hence, the present analysis cannot be generalized to all uses of clefts. At least, it applies when the cleft is used in its discourse structuring function of marking a relatively less expected QUD.

Last but not least, I want to come back to the inferences discussed in Section 2.1. The current approach is not aiming to derive the existential or exhaustive inference of clefts. Nevertheless, it is not in conflict with the existence of such inferences for many occurrences of clefts. I see a potential for future research to investigate what would follow from the information-theoretic and question-based approach for the exhaustivity inference in particular (see Velleman et al. 2012, Pollard & Yasavul 2015 and De Veaugh-Geiss et al. 2018 for approaches to cleft exhaustivity using the QUD framework).

5 Conclusions

In this paper, I presented a new phenomenon, besides connectives, which requires information theory at the discourse level. Building on expectation-based accounts of clefts (e.g., Destruel & Velleman 2014, Tönnis 2021), I analyzed clefts in written German as a device to reduce information density in discourse by relying on its function of disambiguating focus.

The proposed analysis was based on the assumption that the expectations of the author and addressee can be modeled as a probability distribution over questions that could be addressed, which is updated after each new sentence of the text or conversation (following Kehler & Rohde 2017, Tönnis 2021). Accordingly, the proposed analysis incorporated the concept of QUD surprisal (inspired by Demberg & Keller 2008, Asr & Demberg 2015), which measures the difference between the question distribution after having read the canonical sentence in a context and the question distribution after the cleft operator is applied to the canonical sentence. If the QUD surprisal of clefting is high, the extra marking provided by the cleft is required in order to satisfy UID. Ideally, this extra marking could also be achieved by a different means than the cleft, for instance, by adding the discourse marker *by the way*. As a consequence, this analysis treats clefts on a par with other constructions that reduce information density in discourse.

Previous analyses of clefts, such as those focusing on the semantic/pragmatic inferences, were shown to have problems accounting for the examples discussed

in this paper, where the cleft is used to address a relatively less expected QUD. Furthermore, the information-theoretic approach to clefts does not only make the correct predictions for the choice between the cleft and its canonical variant, but it also provides an explanation for why the cleft has the discourse function of marking a relatively less expected QUD.

Acknowledgements

I thank the audience of the DGfS 2022 workshop *Discourse obligates*, the audience of the linguistic colloquium at the University of Stuttgart, Judith Tonhauser, Lisa Schäfer and Edgar Onea for valuable feedback on this work.

References

- Asr, Fatemeh Torabi & Vera Demberg. 2012. Implicitness of discourse relations. In *COLING 2012*, 2669–2684. Mumbai: The COLING 2012 Organizing Committee. <https://aclanthology.org/C12-1163>.
- Asr, Fatemeh Torabi & Vera Demberg. 2015. Uniform Surprisal at the level of discourse relations: Negation markers and discourse connective omission. In *The 11th international conference on computational semantics*, 118–128. London: Association for Computational Linguistics. <https://aclanthology.org/W15-0117>.
- Beaver, David I. & Brady Z. Clark. 2008. *Sense and sensitivity: How focus determines meaning*. Susan Rothstein (ed.), vol. 5 (Explorations in Semantics). Oxford: Wiley-Blackwell. DOI: 10.1002/9781444304176.
- 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.
- Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic*, 27–55. New York: Academic Press.
- De Vaugh-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 Vaugh-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.

- Demberg, Vera & Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109(2). 193–210. DOI: 10.1016/j.cognition.2008.07.008.
- Destruel, Emilie, David I. Beaver & Elizabeth Coppock. 2019. It's not what you expected! the surprising nature of cleft alternatives in French and English. *Frontiers in Psychology* 10. 1–16. DOI: 10.3389/fpsyg.2019.01400.
- 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/>.
- É. Kiss, Katalin. 1998. Identificational focus versus information focus. *Language* 74(2). 245–273. DOI: 10.1353/lan.1998.0211.
- Féry, Caroline. 2017. *Intonation and prosodic structure* (Key Topics in Phonology). Cambridge: Cambridge University Press. DOI: 10.1017/9781139022064.
- Fodor, Janet Dean. 2002. Prosodic disambiguation in silent reading. In *North East Linguistic Society (NELS)*, vol. 32, 113–132. <https://scholarworks.umass.edu/nels/vol32/iss1/8>.
- Genzel, Dmitriy & Eugene Charniak. 2002. Entropy rate constancy in text. In *40th Annual Meeting of the Association for Computational Linguistics (ACL)*, 199–206. DOI: 10.3115/1073083.1073117.
- 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.
- Halvorsen, Per-Kristian. 1978. *The syntax and semantics of cleft constructions*. University of Texas. (PhD thesis).
- 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>.
- Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61(1). 23–62. DOI: 10.1016/j.cogpsych.2010.02.002.
- 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.
- Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55(3–4). 243–276. <https://www.jstor.org/stable/26190905>.
- Krifka, Manfred & Renate Musan. 2012. Information structure: Overview and linguistic issues. In Manfred Krifka & Renate Musan (eds.), *The expression of information structure*, vol. 5, 1–43. Berlin: De Gruyter. DOI: 10.1515/9783110261608.1.

- Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3). 1126–1177. DOI: 10.1016/j.cognition.2007.05.006.
- Levy, Roger & T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. In *Advances in Neural Information Processing Systems 19: The 2006 Conference*, 849–856. <https://ieeexplore.ieee.org/document/6287477>.
- Onea, Edgar. 2019. Exhaustivity in *it*-clefts. In Chris Cummins & Napoleon Katsos (eds.), *The Oxford handbook of experimental semantics and pragmatics*, 401–417. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780198791768.013.17.
- Percus, Orin. 1997. Prying open the cleft. In *North East Linguistic Society (NELS)*, vol. 27, 337–352. <https://scholarworks.umass.edu/nels/vol27/iss1/25>.
- Pollard, Carl & Murat Yasavul. 2015. Anaphoric *it*-clefts: The myth of exhaustivity. In Ross Burkholder, Carlos Cisneros, Emily R. Coppess, Julian Grove, Emily A. Hanink, Hilary McMahan, Cherry Meyer, Natalia Pavlou, Özge Sarıgül, Adam Roth Singerman & Anqi Zhang (eds.), *Proceedings of the fiftieth annual meeting of the Chicago Linguistic Society*, 381–393. Chicago: The Chicago Linguistic Society.
- Prince, Ellen F. 1978. A comparison of *wh*-clefts and *it*-clefts in discourse. *Language* 54(4). 883–906. DOI: 10.2307/413238.
- Reinhart, Tanya. 1995. *Interface strategies*. Utrecht: OTS working papers in linguistics.
- 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. 1–231. DOI: 10.1075/sigla.4.
- Shannon, Claude E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27(3). 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.
- Simons, Mandy, David I. Beaver, Craige Roberts & Judith Tonhauser. 2017. The best question: Explaining the projection behavior of factives. *Discourse Processes* 54(3). 187–206. DOI: 10.1080/0163853X.2016.1150660.
- Szabolcsi, Anna. 1981. The semantics of topic-focus articulation. In J. A. G. Groenendijk, T. M. V. Janssen & M. B. J. Stokhof (eds.), *Formal methods in the study of language*, 513–540. Amsterdam: Mathematisch Centrum.
- Szendrói, Kriszta. 1999. A stress-driven approach to the syntax of focus. *UCL Working Papers in Linguistics* 11. 545–573.
- Tönnis, Swantje. 2021. *German “es”-clefts in discourse: A question-based analysis involving expectedness*. Graz University. (Doctoral dissertation).

- Tönnis, Swantje, Lea M. Fricke & Alexander Schreiber. 2018. Methodological considerations on testing argument asymmetry in German cleft sentences. In Eric Fuß, Marek Konopka, Beata Trawiński & Ulrich H. Waßner (eds.), *Grammar and Corpora 2016*, 231–240. Heidelberg: Heidelberg University Publishing. DOI: 10.17885/heiup.361.c4703.
- Tönnis, Swantje & Judith Tonhauser. 2022. German clefts address unexpected questions. In John R. Starr, Juhya Kim & Burak Öney (eds.), *Semantics and Linguistic Theory (SALT)*, vol. 32, 661–684. DOI: 10.3765/salt.v1i0.5359.
- 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.
- Westera, Matthijs & Hannah Rohde. 2019. Asking between the lines: Elicitation of evoked questions in text. In *The 22nd Amsterdam Colloquium*, 397–406. <https://archive.illc.uva.nl/AC/AC2019/Proceedings/index.html>.

