

German *es*-Clefts in Discourse

A Question-Based Analysis Involving Expectedness

Dissertation

zur Erlangung des Doktorgrades der Philosophie
an der Karl-Franzens-Universität Graz

eingereicht von Swantje Tönnis

am

Institut für Germanistik

Erstbegutachter: Univ.-Prof. Dr. Victor Edgar Onea Gáspár
Zweitbegutachter: Univ.-Prof. Dr. Malte Zimmermann

2021

Acknowledgments

First of all, I would like to thank my supervisor Edgar Onea. He is just an amazing supervisor and person. I thank him for sharing his wisdom with me, e.g. for teaching me that it is good to be wrong. He always encouraged me to do my best, he listened to me and respected me. The discussions with him made this thesis so much better. I could not have written it without him. Furthermore, I would like to thank my second supervisor, Malte Zimmermann. I thank him for various fruitful discussions and valuable feedback on my work. He did not only help me with his expertise, but he also gave me a good advice when I needed it.

Moreover, I would like to thank my colleagues at the University of Graz. Especially, I thank Lea Fricke, who is a wonderful office mate and friend. Working with Lea next to me made everything seem easier. I thank her for various insightful discussions of data, theories, experiments, and everything else that there was to discuss. Her being so well-organized helped me being more structured myself. Moreover, I want to thank the members of the research group *Information Structure* in Graz, particularly Boban Arsenijevic, Madeleine Butschety, Steffen Heidinger, and Dina El Zarka. They gave me valuable feedback at various stages of my work.

Additionally, I thank the audience of *The thirteenth International Tbilisi Symposium on Language, Logic and Computation*, especially Markus Egg, Daniel Hole, Berit Gehrke, Todd Snider, and Carla Umbach, who gave me first feedback on my idea on how to analyze German *es*-clefts in discourse. This conference and the scientific exchange there gave me the motivation to finally start writing up the thesis. Moreover, I thank Stefan Baumann, Caroline Féry, Kriszta Szendrői, Michael Wagner, and Ede Zimmermann, who taught a wonderful summer school in Graz, which helped me to form my theoretical hypotheses. Other linguists who helped me at different stages of my PhD are Anton Benz, Jun Chen, Joseph P. De Veaugh-Geiss, and Maribel Romero, which I very much appreciate.

My work was partly funded by the German Research Foundation (DFG) as part of the project “Exhaustivity in *it*-Clefts” in the priority program XPrag.de. I thankfully appreciate their support.

Moreover, I thank my proofreaders Jessica, Laura, Madeleine, Maria, Maya, and especially Lea, who did a great job even though I did not give them much time. Furthermore, I thank our student assistants Alex, Anna, Daniel and Simon for supporting me, being extremely reliable and clever. I thank Jonas for discussing the logic of my model from a mathematics point of view. I also want to thank Eric Berg for writing amazing crime novels, which contain beautiful *es*-clefts.

I would like to thank Maya, who supported me as a friend, office mate, and colleague. She kept on being excited to talk about clefts with me, which I appreciate very much. She asked just the right questions, and she was an amazing companion for writing a thesis. Maybe most importantly, she made me laugh when the thesis got me down.

I thank my friends for supporting me, feeling with me, and for not forgetting me even though there were times in which I was doing nothing else than writing the thesis. I especially want to thank my wonderful flatmate Lena for always being there for me, giving millions of judgments on weird data, having fun with clefts, and sharing my PhD journey with all its ups and downs.

Now, I would like to express how grateful I am to Arne. He was always there for me even though he was so far away. He always listened to me, came up with the best strategies for how to write a PhD thesis, how to deal with frustration, or how to have fun at every stage of the PhD. He knew when to take my mind off things and gave me endless energy. I thank him for being as wonderful as he is.

Finally, I thank my family for always supporting me and believing in me. My parents always let me make my own decisions even if that was switching from studying physics to studying linguistics. The best decision ever! And my parents celebrated every little step of my PhD journey, which I appreciate very much. I especially want to thank my mother for always listening to me.

Contents

1	Introduction	1
1.1	The German <i>es</i> -Cleft	1
1.2	Semantics and Pragmatics of Clefts	7
1.3	The Discourse Function of Clefts	10
1.4	Structure of the Thesis	15
I	Theoretical Background	18
2	Semantics and Pragmatics of Questions	19
2.1	Introduction	20
2.2	Functional Approach	36
2.3	Alternative Semantics Approach	38
2.4	Partition Approach	41
2.5	Inquisitive Approach	45
2.6	Subtypes of Questions	48
2.7	Pragmatic Features of Questions	52
3	Focus	58
3.1	Introduction	58

3.2	Focus Sensitivity	64
3.3	Alternative Semantics	69
3.4	Focus and Questions	72
3.5	Pragmatic Effects	78
4	Discourse Model	82
4.1	Introduction	83
4.2	Context Update	89
4.3	Question under Discussion Framework	101
4.4	Current Question and Discourse Question	104
4.5	Accommodation of the Current Question	116
4.6	Potential Questions	120
4.7	Summary	127
II	Analysis of Cleft Sentences	131
5	Syntax (and Semantics) of Clefts	132
5.1	The <i>es</i> -Cleft and Similar Structures	132
5.2	Challenges	138
5.3	Definite Description Approach	140
5.4	<i>Wh</i> -cleft Approach	149
5.5	Focus Approach	154
5.6	Referential Pronoun Approach	157
5.7	Summary	161
6	Existence Presupposition of Clefts	163

7 Exhaustivity of Clefts	169
7.1 Introduction	169
7.2 Exhaustivity as Entailment	174
7.3 Exhaustivity as Presupposition	175
7.4 Exhaustivity as Scalar Implicature	180
7.5 Exhaustivity as Generalized Implicature or Other Pragmatic Inference . . .	183
7.6 Summary	191
8 Discourse Approaches	193
8.1 Introduction	193
8.2 Clefts and Givenness	195
8.3 Clefts and Discourse Coherence	205
8.4 Clefts and Contrast	209
8.5 Clefts and Focus	215
8.6 Clefts and Questions	218
8.7 Summary	226
9 Empirical Approaches to Clefts	227
9.1 Corpus Studies	227
9.2 Experiments on Exhaustivity	237
9.3 Experiments on Contrastivity	245
9.4 Summary	249
10 Expectedness-based Analysis of German Clefts	252
10.1 More Data and Hypothesis	252
10.2 Expectedness	263

10.2.1 Concept	263
10.2.2 Formalization	271
10.3 Adapted Discourse Model	284
10.4 Applying the Model	291
10.5 Discussion and Outlook	299
11 Conclusion	307

Abbreviations

<i>a</i> -CQ	Current question case of expectedness function for assertive speech act
<i>a</i> -PQ	Potential question case of expectedness function for assertive speech act
CFH	Contrastive Focus Hypothesis
CG	Common Ground
CQ	Current Question
CS	Commitment Space
DQ	Discourse Question
EV	Expectedness Value
FA	Functional Application
<i>i</i> -CQ	Current question case of expectedness function for interrogative speech act
<i>i</i> -PQ	Potential question case of expectedness function for interrogative speech act
<i>i</i> -SUB.Q	Sub-question case of expectedness function for interrogative speech act
LPQ	Likely Potential Question
PL	Predicate Lowering
PPQ	Primary Potential Question
PQ	Potential Question
QUD	Question Under Discussion
SDRT	Segmented Discourse Representation Theory
SPQ	Standard Potential Question

Notations

- * ungrammatical
- # semantically/pragmatically unacceptable
- ? marginally acceptable
- [...]_F .. is focused

List of Glosses

1	first person
3	third person
ACC	accusative case
ASP	aspect
COP	copula
DAT	dative case
DEF	definite
FEM	feminine
FUT	future tense
MSC	masculine
NOM	nominative case
NTR	neuter
ORD	ordinal
PAST	past tense
PL	plural
POSS	possessive
PROG	progressive aspect
PRO	pronoun
PRT	particle
Q	question
REFL	reflexive
SG	singular
V.PRT	verb particle
WH	<i>wh</i> -question

Abstract

This thesis aims to deepen our understanding of cleft structures in German with a focus on how clefts contribute and relate to discourse. The first part of the thesis is concerned with the theoretical background, including the semantics and pragmatics of questions and focus. Furthermore, it presents a new perspective on question-based discourse models, which integrates a current question, a discourse question, and both the speaker's and the addressee's perspective.

The second part discusses the syntax, the existence presupposition, and the exhaustivity inference of clefts in German compared to other languages. Moreover, previous discourse approaches and empirical results are presented. The discussion of the literature shows that various and partly divergent claims have been made, many of which only cover the acceptability and function of clefts for a very limited set of examples. Crucially, so far no approach has accounted for the difference between canonical sentences and *es*-clefts as well as for the various discourse functions a cleft can realize. This thesis shows that most of the observations made about clefts can be traced back to EXPECTEDNESS or consequences of expectedness.

Eventually, the thesis provides a question-based discourse model that predicts when a cleft is acceptable, and when it is not, depending on the discourse context. I propose that the German *es*-cleft addresses a rather unexpected question whereas the canonical sentence addresses more expected questions. Expectedness is formalized as a function that assigns an expectedness value to a question given a previous discourse. This function is incorporated into the discourse model as affecting the conditions for accepting a discourse move. The thesis demonstrates the application of the model to several examples of clefts showing that it makes the correct predictions. Expectedness does not only model the acceptability of clefts, but the acceptability of discourse moves in general.

Chapter 1

Introduction

This thesis aims to deepen our understanding of cleft structures in German with a focus on how clefts contribute and relate to the discourse. Eventually, it provides a formal discourse model that predicts when a cleft is acceptable and when it is not depending on the discourse context.

1.1 The German *es*-Cleft

Let us first discover what a CLEFT SENTENCE¹ looks like. Following Hedberg (2000), the ‘standard’ German *es*-cleft², such as (1), has four components (adapted terminology).

- (1) Es ist Arne, der lächelt.
CLEFT PRONOUN + COPULA + CLEFT PIVOT + CLEFT RELATIVE CLAUSE
CLEFT MATRIX CLAUSE

‘It is Arne who is smiling.’

Two different versions of clefts can be distinguished: the FOCUS BACKGROUND CLEFT, as in (2-a), and the TOPIC COMMENT CLEFT, as in (2-b) (c.f. Prince, 1978).

¹The term *cleft sentence* was established by Jespersen (1949).

²Non-standard clefts are discussed in section 5.1.

- (2) a. Es is ARNE, der lächelt.
 it is ARNE who smiles
'It is Arne who is smiling.'
- b. Es ist ARNE, der LÄCHELT.
 it is ARNE who SMILES
'It is ARNE who is SMILING.'

The focus background cleft in (2-a) has the main accent (or focus accent) on the cleft pivot, indicated in capital letters. The cleft relative clause is deaccented. The topic comment cleft in (2-b) superficially only differs from the focus background cleft with respect to its intonation. As indicated in small capital letters, the pivot of the topic comment cleft receives a rising accent (also called topic accent), and the cleft relative clause carries the main accent, here on the verb *lächelt* ('smiling').

According to Prince (1978), these two variants of the cleft also differ with respect to their function in a conversation. She states that the pivot of a focus background cleft represents new, often contrastive, information while its relative clause "represents known or old information, which is not marked as assumed to be in the hearer's consciousness and which is not the theme" (Prince, 1978:896). The topic comment cleft, in contrast, provides new information in the cleft relative clause, and the cleft pivot is the topic of the sentence in most cases. For most topic comment clefts, it is assumed that the addressee does not know the information conveyed by the cleft relative clause. Prince (1978), however, points out that the cleft still marks that information as a "known fact", at least to the author.

The difference between focus background clefts and topic comment clefts is illustrated in examples (3) and (4).³

- (3) A: Wer lächelt?
'Who is smiling?'
- B: Es ist ARNE, der lächelt. / #Es ist ARNE, der LÄCHELT.
'It is ARNE who is smiling.' / #'It is ARNE who is SMILING.'

³# marks a semantically or pragmatically unacceptable sentence, and ? marks a marginally acceptable sentence.

- (4) A: Was ist mit Arne?

'Who is smiling?'

B: Es ist ARNE, der LÄCHELT. / ?Es ist ARNE, der lächelt.

'It is ARNE who is SMILING.' / ?'It is ARNE who is smiling.'

In the context of (3), only the focus background cleft is acceptable. As we will see later, this kind of cleft indicates exactly the question that is asked in (3). The context in (4) prefers the topic comment cleft, although the focus background cleft is also marginally acceptable, implying that Arne usually does not smile.

For German speaking readers, the cleft sentences in (3) and (4) might sound a bit unnatural, or at least too effortful, given that one could say *Arne lächelt* ('*Arne is smiling*') instead. In order to convince those readers of the relevance of investigating this structure in German, I present two original clefts taken from newspaper articles. Those show a very natural use of clefts in German. The first example is an occurrence of a German focus background cleft, taken from an article about Qatar wanting to host the Olympic games, and about the newest achievements of the country.

- (5) Wer durch die multimediale Schau im Katar-Haus streift, erfährt ein wenig über die Entwicklung Dohas, die Neuentwicklung ganzer Städte, Kulturprojekte und wissenschaftliche Ambitionen. Ein reicher Staat baut sich neu. Dabei helfen renommierte Architekten und berühmte Schauspieler, die man zeigt, wie sie über den roten Teppich des Filmfestivals von Doha schlendern. Doch die Filmstars spielen nur Nebenrollen. **Es ist der Sport, über den sich das Land neu definieren will.**

*'If one walks through the multi-media show in the Qatar house, one gets to know a little about the development of Doha, the creation of entire new cities, cultural projects, and scientific ambitions. A rich state rebuilds itself. Prestigious architects and famous actors are displayed while strolling over the red carpet at the film festival of Doha. But the film stars just play supporting roles. **It is through sport that the state wants to redefine itself.**'*

(Die Tageszeitung, 14.08.2012, p. 03, Scheichs mögen Frauen)

The main accent in the cleft lies on the word *Sport*, which is contrasted with other ways Qatar could redefine itself. The content of the cleft relative clause is given in the context and, thus, deaccented. Hence, this cleft can be categorized as a focus background cleft.

The example in (6) is an occurrence of a German topic comment cleft, taken from an article about an art journal.

- (6) Lehning glückte es auch, den zu jener Zeit in Dessau lebenden Maler und Kunstretheoretiker Wassily Kandinsky für seine Idee zu begeistern. Kandinsky war es, der dann auch den Aufmacher-Beitrag für die erste Ausgabe schrieb.

'Lehning also succeeded in getting the painter and art theorist Wassily Kandinsky, who lived in Dessau at that time, interested in his idea. It was Kandinsky who then also wrote the lead story for the first edition.'

(Frankfurter Rundschau, 23.08.1997, p. 2, ZEIT UND BILD, Vor 70 Jahren wurde die internationale Kulturzeitschrift "i 10" gegründet)

Here, the main accent lies somewhere in the cleft relative clause, probably on the word *Ausgabe* ('edition'). We see that the information conveyed in the relative clause is new information, while the pivot *Kandinsky* was mentioned right before. Hence, this cleft is a topic comment cleft.

This thesis will be mainly concerned with focus background clefts, since these are much more frequent in German than topic comment clauses (Dufter, 2009; Tönnis et al., 2018). In general, cleft sentences are not as frequent in German compared to other languages. A comparative study on translations of English *it*-clefts to different Romance languages and German showed that *es*-clefts in German occurred much less frequently as a translation than the corresponding cleft in any of the Romance languages (Dufter, 2009). Dufter argues that clefts are not used as frequently in German because there are competing and less complex means of achieving the same effect, e.g. topicalization, scrambling, or modal particles, which do not exist in English. This makes it even more interesting to find out why clefts are used in German at all. As the examples in (5) and (6) and many other corpus occurrences of clefts show, there are natural examples of clefts in German, which must have some purpose. This thesis aims to investigate this purpose.

Let us now take a closer look at what structurally defines a cleft. There are several structures which are similar to the cleft, but which are not categorized as a cleft. One structure that needs to be distinguished from the cleft is what I call a presentational statement involving a relative clause, such as in (7).

- (7) Das Museum_i wurde eröffnet. Es_i ist eine Bereicherung, die nicht zu
 the museum.NTR was opened it.NTR is a gain that not to
 unterschätzen ist.
 underestimate is
'The museum was opened. It is a gain not to be underestimated.'

In order to be able to clearly identify clefts, and distinguish them from sentences such as (7), I developed the decision tree for German clefts in Figure 1.1, which I will explain in the following.

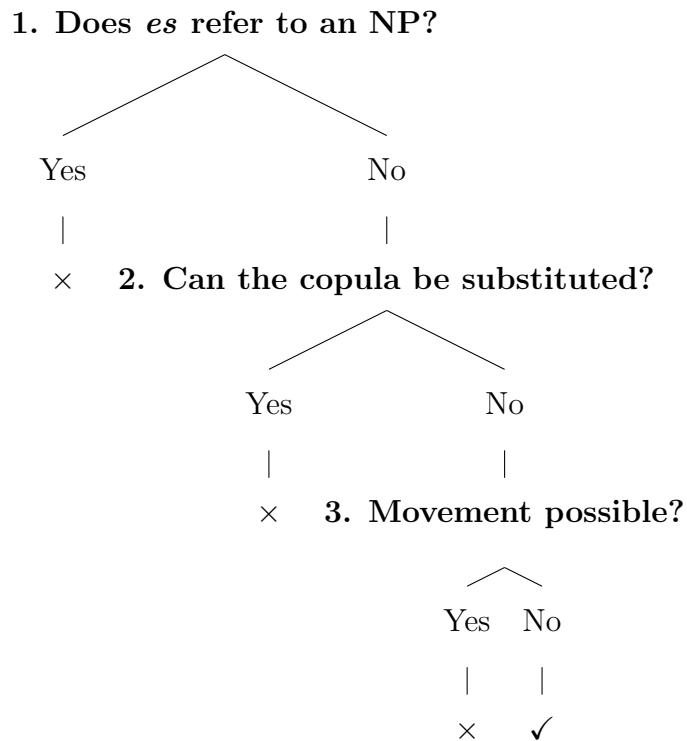


Figure 1.1: Decision tree for identifying a cleft in German

- 1. Does *es* refer to anything?** If the cleft pronoun *es* refers to an antecedent, it cannot be a cleft. This can be tested by replacing the antecedent of the pronoun and checking whether the sentence is still well-formed, as in (8).

- (8) Die Ausstellung_i wurde eröffnet. *Es_i ist eine Bereicherung, die nicht zu
 the exhibition.FEM was opened it.NTR is a gain that not to
 unterschätzen ist.
 underestimate is
'The exhibition was opened. It is a gain not to be underestimated.'

In this example, the neuter noun phrase *das Museum* ('the museum'), as in (7), is replaced with the feminine noun phrase *die Ausstellung* ('the exhibition'). As (8) indicates, this replacement makes the alleged cleft sentence ungrammatical (as marked with *). Hence, the test shows that the sentence in (7) is not a cleft.

2. Can the copula be substituted? If the first test comes out negative (i.e. the pronoun does not refer to anything), there is a second test to execute. For this, consider the following example, an original cleft taken from Wikipedia (<http://de.wikipedia.org/wiki/Aargletscher>).

- (9) Die Aargletscher sind eine Gletschergruppe in den östlichen Berner Alpen am Finsteraarhorn.

'The Aar-Glaciers are a system of glaciers in the East Bernese Alps at the Finsteraarhorn.'

- a. Es **sind** vier gewaltige Eisströme, die aus den Firnmulden der Finsteraarhornguppe in den Hintergrund des Haslitals hinabsteigen.
 'It is four gigantic ice streams which descend from the firn hollows of the Finsteraarhorn mountains into the background of the Halsi Valley.'
- b. Es **handelt sich um** vier gewaltige Eisströme, die...
 it deals with four gigantic ice.streams which...
 'We are dealing with four gigantic ice streams which...'
- c. **Sie bestehen aus** vier gewaltigen Eisströmen, die ...
 They consist of four gigantic ice.streams which...
 'They consist of four gigantic ice streams which...'

The original sentence is the one in (9-a). However, the variants in (9-b) and (9-c) indicate that the copula *sind* ('are') can be replaced by other verbs without changing the content. If this is possible, again the sentence cannot be a cleft. Therefore, (9-a) is not a cleft, either.

3. Is movement possible? If both previous tests come out negative, the movement test can be applied, as illustrated in (10-a) and (10-b).

(10) *I don't think Maria won the game.*

- a. Es war Nina, die das Spiel gewonnen hat.
It was Nina who the game won has
'It was Nina who won the game.'
- b. #Nina, die das Spiel gewonnen hat, war es.
Nina who the game won has was it
'Nina who won the game was it.'

The cleft pivot does not form a constituent together with the cleft relative clause. Hence, it cannot be moved into a sentence initial position without affecting the discourse coherence. The pivot of the cleft in (10-a) cannot be moved together with the cleft relative clause, as the unacceptability of (10-b) indicates. Hence, (10-a) is identified as a cleft.

Furthermore, Hedberg (1990) suggests felicitous modification of the pivot with *even* or *also* as a diagnostic for distinguishing the topic comment cleft from the focus background cleft. A topic comment cleft can be modified by such a focus particle, while the focus background cleft cannot, as indicated in the following example.

(11) a. It is even/also John who climbed Mount EVEREST.
b. #It was even/also JOHN who climbed Mount Everest.

In this test, it is important that the intonation of the original cleft is not changed after modification with the focus particle. If it is a focus background cleft, this will lead to an unacceptable sentence, such as (11-b). If it is a topic comment cleft, it stays felicitous, as in (11-a).

1.2 Semantics and Pragmatics of Clefts

For estimation of the semantic and pragmatic effects, the cleft is frequently compared to the CANONICAL SENTENCE, as in (12-a), or to an EXCLUSIVE, as in (12-b).

- (12) a. Arne smiled./ARNE smiled.
b. Only ARNE smiled.

These structures seem to be similar, although not the same as the cleft structure. A comparison might provide insight into what the purpose of the cleft actually is. The canonical sentence sometimes serves as comparison including a focus accent on the former cleft pivot, as in the second variant in (12-a), or without considering intonation, as in the first variant. The first variant is also referred to as the PREJACENT. Beaver and Clark (2008) explain the term *prejacent* with the example in (13). They argue that there is an underlying proposition (*David wears a bow tie when teaching*) that *only* modifies. This proposition is called the *prejacent*.

- (13) David only wears a bow tie when teaching. (Beaver and Clark, 2008:68)

For the cleft, Beaver and Clark (2008) assume such an underlying prejacent as well, as represented by the first variant in (12-a).

There is substantial literature on cleft sentences in all kind of languages. Accordingly, several semantic and pragmatic features have been assigned to the cleft structure. First of all, clefts are frequently claimed to be contrastive (e.g. Destruel and Velleman, 2014; Seuren, 1985; Prince, 1978), as in (14). The same holds for German clefts. For the purpose of brevity, I will sometimes use English examples which translate to German if not stated otherwise.

- (14) a. Lisa did not look so happy. It was Arne that smiled.
b. Lisa did not look so happy. Arne smiled.

The cleft in (14-a) unambiguously establishes a contrast between Arne and Lisa. The canonical sentence in (14-b) can also establish this contrast, but it is not forced as in (14-a).

Furthermore, the cleft is assumed to have an EXISTENCE PRESUPPOSITION (e.g. Percus, 1997; Halvorsen, 1978; Prince, 1978), as in (15-b), and to give rise to an EXHAUSTIVITY INFERENCE (see, among many others, Horn, 1981; Szabolcsi, 1981), as in (15-c).

- (15) a. It is ARNE who is smiling.
- b. Existence: Somebody is smiling.
- c. Exhaustivity: Nobody other than Arne is smiling.

The existence presupposition is widely agreed upon, while the exhaustivity inference is still hotly debated. Many theoretical approaches treat exhaustivity as semantically encoded in the cleft structure. In contrast, experiments show that the inference is not very robust, which speaks in favor of a pragmatic analysis. The inference can, as Horn (1981) pointed out, easily be canceled, as in (16).

- (16) It is Arne who is smiling, and Nina is smiling, too.

If the inference was part of the semantics of clefts, it should be impossible or at least more complicated to cancel it.

Others have noticed parallels between the cleft and the focused canonical sentence in question–answer pairs (c.f. Seuren, 1985).

- (17) Was hat Arne verkauft?
'What did Peter sell?'
- a. #Er hat wenig / die meisten Postkarten verkauft.
 he has few / the most postcard sold
'He sold few/most postcards.'

- b. Er hat eine Postkarte/ ein paar Postkarten verkauft.
 he has a postcard / a few postcards sold
'He sold a/a few postcard(s).'

- (18) a. #Es waren wenig / die meisten Postkarten, die Arne verkauft hat.
 it were few / the most postcards that Arne sold has
'It was few/the most postcards that Arne sold.'

- b. Es war eine Postkarte / waren ein paar Postkarten, die Arne verkauft hat.
it was a postcard / were a few postcards that Arne sold has
'It was a/a few postcard(s) that Arne sold.'

Focus in the canonical sentence as an answer to an explicit question as well as the cleft are both infelicitous with quantified phrases involving a partitive quantifier, such as in (17-a) and (18-a) involving *few* and *most*. The non-partitive quantifiers *a* and *a few*, in contrast, are felicitous. This shows that the cleft is related to the canonical. Moreover, it suggests that the cleft is related to a question.

1.3 The Discourse Function of Clefts

Besides the approaches that focus on the structure of clefts, there are several approaches to clefts that analyze them on the level of discourse. There is a line of research originating in Prince (1978), which is based on corpus data and which analyzes, on a descriptive level, which functions clefts may realize in the discourse. Delin (1990), Hedberg (1990), Delin and Oberlander (1995), among others, follow this line of research. They claim various functions for the cleft, such as marking contrastivity or givenness, changing the temporal order of events, commenting, backgrounding, marking a source or a cause etc. However, these approaches are not formalized.

In contrast to this, there are approaches that analyze clefts in a question-based discourse model, e.g. Seuren (1985), van Kuppevelt (1991), or Velleman et al. (2012). I will follow these approaches and will provide a formal analysis based on Roberts' (2012) Question Under Discussion (QUD) model. Let us first consider the motivation for treating clefts as a phenomenon that needs to be discussed with respect to its discourse function.

My analysis is inspired by clefts which I discovered in crime novels. Here, I provide two such examples.

- (19) Context: *A couple is on their way to their new house at the sea. He suffers from cancer. The two are passing by a clinic for him to get special treatment before they reach their new house.*

“Fühlst du dich gut?”, fragte sie nach einer Weile. „Du bist so still.”
“Alles bestens.”

Sollte er ihr beichten, dass er das Gefühl hatte, in seiner Brust atme die Lunge eines Spatzes? “Die Landschaft ist einfach unschlagbar”, sagte er.

“Ja, das ist sie. Du wirst sehen, das Meer und die Ruhe werden Wunder bei dir bewirken.”

“Ein wenig hoffe ich auch auf die Spezialklinik in Rostock”, schränkte er ein.

Nun war es Julia, die still wurde.

(Berg, E. (2018). So bitter die Rache. München: Limes. p. 29)

‘ “Are you feeling good?” she asked after a while. “You’re so quiet.”
“Everything’s fine.”

Should he confess to her that he felt as if the lungs of a sparrow were breathing in his chest? “The landscape is just amazingly beautiful”, he said.

“Yes, it is. You will see, the sea and some peace and quiet will do wonders for you.”

“I am also kind of hoping for the special hospital in Rostock”, he responded.
Now, it was Julia who became quiet.’

- (20) Context: *The protagonist is writing about a murder that she found out about.*

Jeder Mord, von dem wir erfahren, ist für sich beängstigend genug. Aber wir meinen, ein Bild von den jeweiligen Verbrechern zu haben – von den Räubern, die Rentner in ihrem Gartenhäuschen erschlagen, den Sexualverbrechern, die Frauen und Kindern auflauern, den reizbaren jungen Männern, die uns auf Bahnhöfen überfallen. **Es ist der massive Einbruch der Gewalt in unseren Alltag und unsere Ordnung, den wir so sehr fürchten.**

(Berg, E. (2016). Die Schattenbucht. München: Limes. p. 10)

‘*Every murder that we learn of is per se scary enough. But we assume we know what such criminals look like – of the burglars who slay pensioners in their garden shed, of the sexual offenders, that wait for women and children, of the easily irritated young men who mug us at train stations. It is the massive invasion of violence into our everyday life and our order that we fear so much.*’

These clefts smoothly fit into the text passages. We see here that even though clefts are not frequent in German, there are still clefts, such as the ones above, that sound very natural. Crucially, the discourse effect changes significantly when the cleft is replaced with a canonical sentence, such as in (21-a) and (21-b), respectively.

- (21) a. Nun wurde Julia still.
 Now became Julia quiet
 ‘Now, Julia became quiet.’
- b. Wir fürchten den massiven Einbruch der Gewalt in unsern
 we fear the massive invasion of the violence into our
 Alltag und unsere Ordnung so sehr.
 everyday.life and our order so much
 ‘We fear the massive invasion of violence into our everyday life and our order
 so much.’

It seemed to me that the clefts in (19) and (20) successfully address issues which are not expected to be addressed in that context. In (19), the cleft seems to refer back to Julia’s boyfriend being quiet and establishes a contrast. This reference is not expected to be established at that point since the utterance *Du bist so still* (‘You are so quiet’) is no longer fresh in the reader’s memory. In line with this, I observe that the replacement with the canonical does not establish this contrast (at least not as obviously). The canonical sentence is likely to be interpreted as just expressing what happened next, an issue that is much more expected.

Similarly for (20), if the cleft was replaced with (21-b), it would be interpreted as yet another aspect of crime that we assume to have an idea of. However, the cleft unambiguously gives the impression that we are addressing something else now, namely the less expected issue of what is really scaring us and the kind of criminality we know little about.

These intuitions are in line with question-based approaches that investigate the relation of clefts to certain questions in the discourse. The big question for the question-based discourse approaches is how the acceptability of a cleft in a context differs from the acceptability of its canonical counterpart. Contrary to the examples in (19) and (20), it is frequently the case for German that replacing a real occurrence of an *es*-cleft with its canonical counterpart does not affect the acceptability of the sentence in the same context (see Tönnis, 2018). Even in examples (19) and (20), replacing the cleft would not lead to unacceptability but just to the different discourse relations between the context and the canonical sentence. This makes the proposed effect hard to test empirically.

We need an example in which the cleft is acceptable while the canonical sentence is not. Such an example could provide interesting insights into how an *es*-cleft structures the discourse differently than an unclefted sentence would do. Inspired by the clefts in the crime novels, I constructed an example, which clearly favors the *es*-cleft over the canonical sentence. Example (22) presents a context in which the *es*-cleft is more acceptable than its canonical equivalent. The context in (22) and its respective continuation in (22-a) or (22-b) are intended to be uttered or written by the same speaker/author.

- (22) Lena hat gestern auf der Party mit einem Typen₁ gesprochen. Die beiden haben viel gelacht und sich direkt für den nächsten Abend verabredet. Dann ist Lena glücklich nach Hause gefahren.

'Yesterday at the party, Lena talked to some guy₁. The two of them laughed a lot and they agreed to meet again the next evening. Then, Lena went home happily.'

- a. Es war Peter₁, mit dem sie gesprochen hat.
it was Peter₁ with whom she talked has
'It was Peter₁ she talked to.'
- b. ?Sie hat mit Peter₁ gesprochen.
She has with Peter₁ talked
'She talked to Peter₁.'

Native speakers of German judge (22-a) as acceptable, while they are confused about the function of (22-b). The problem with (22-b) is that it is very difficult for the addressee to figure out how it could relate to the preceding discourse. In case of the *es*-cleft in (22-a), in contrast, the coherence is established without problems because the cleft can easily be interpreted as referring back to the discourse referent introduced by *einem Typen* ('some guy'). Hence, the reader interprets *Peter* as referring to that referent.

The canonical sentence in (22-b) seems to be incapable of referring back to the discourse referent introduced by *einem Typen* ('some guy'), and leaves the addressee a bit puzzled. For the addressee, the canonical sentence in (22-b) does not clearly identify Peter as the guy Lena talked to. If the speaker indeed did not want to identify the guy Lena talked to as Peter, the canonical sentence does not seem relevant in this context. Hence,

discourse coherence is not established at all or only with a lot of extra effort, such as repair strategies. The canonical sentence would only constitute a coherent continuation if the speaker and the addressee knew Peter before.

In this thesis, I will investigate which questions a cleft is capable of addressing compared with the questions a canonical sentence may felicitously address. Example (22) suggests that the cleft addresses the question *Which guy did Lena talk to?*, which cannot be addressed by the canonical sentence in this context. However, there are other contexts in which the canonical sentence perfectly addresses this question. An explicit question would be such a context, as illustrated in (23).

(23) Mit welchem Typen hat Lena gesprochen?

'Which guy did Lena talk to?'

- a. ?Es war Peter₁, mit dem sie gesprochen hat.
it was Peter₁ with whom she talked has
'It was Peter₁ she talked to.'
- b. Sie hat mit Peter₁ gesprochen.
She has with Peter₁ talked
'She talked to Peter₁'

Interestingly, the acceptability is reversed now. The canonical sentence provides a good answer while the cleft sounds odd. Based on these and many more examples to come, this thesis develops the following hypotheses.

- (24)
- a. An *es*-cleft addresses a question that came up in the preceding context, but that the addressee does not expect to be answered at that point in the discourse compared to other questions.
 - b. Those questions that are more expected are preferably addressed with a canonical sentence instead of a cleft.
 - c. Those questions that are neither particularly expected nor particularly unexpected can equally well be addressed by a cleft or a canonical sentence.

I will extend existing question-based discourse models by integrating the concept of EXPECTEDNESS. I will formalize expectedness as a function that assigns an expectedness value to a question given a previous discourse. Expectedness not only models the acceptability of clefts but the acceptability of discourse moves in general. I will show that it accounts for relevance constraints and accommodation conditions as well. Eventually, my model can subsume at least some of the various discourse functions claimed previously under the function of unexpectedness or consequences of unexpectedness.

1.4 Structure of the Thesis

This thesis is structured as follows. It contains two parts. Part I (chapter 2 – chapter 4) is concerned with the *theoretical background*, which introduces the important notions that I use for my analysis of clefts. Part II (chapter 5 – chapter 11) is concerned with the *analysis of cleft sentences* in German, but also other languages. If you have a solid knowledge of formal analyses of questions, focus, and discourse models, and are just curious about my analysis of clefts, I suggest skipping to Part I.

In chapter 2, the *semantics and pragmatics of questions* are presented. I model the discourse function of clefts in a question-based discourse model, which requires a basic understanding of what the linguistic representation of a question should be. Furthermore, I analyze clefts as addressing some questions rather than other questions. Hence, chapter 2 distinguishes different kinds of questions in order to provide a basis for approaching which question the cleft could address. Furthermore, it distinguishes different levels of questions, distinguishing the form of a question from its semantic representation and its realization of an interrogative speech act.

Chapter 3 provides an overview of the topic of *focus*. Focus is another essential part of the discourse model that I am proposing. Furthermore, clefts are often compared to plain focus and to expressions involving focus-sensitive expressions. Therefore, section 3.2 introduces focus and focus sensitivity in the framework of alternative semantics. The chapter develops a definition of focus that is related to questions.

Building on the previous two chapters, chapter 4 develops a *question-based discourse model*, which defines the most important concepts that my analysis of clefts uses. These include the notion of a context update, of a question under discussion, and of a potential question. Furthermore, this chapter explains what it means to accept a discourse move made by a speaker, and what makes an assertion or a question relevant to the preceding discourse. The chapter makes some adjustments to previous versions of question-based discourse models, and provides a new perspective on the role of the current question.

Chapter 5 is concerned with the *syntax of clefts* and, in parts, the semantics that is associated with it. The syntactic analysis of clefts has turned out to be a huge challenge that has not yet been met. This chapter summarizes various approaches to the syntax of clefts and points out the discrepancy between its syntactic structure and its semantic/pragmatic function.

In chapter 6, the *existence presupposition* of clefts is addressed. Since the existence presupposition is not debated much, this chapter just provides a brief overview that includes the derivation of the presuppositions and some complications. Furthermore, it compares the existence presupposition of clefts to the (potential) existence inference of focus.

Chapter 7 presents several theoretical approaches to *exhaustivity* in clefts. It discusses whether exhaustivity of clefts is a semantic or a pragmatic inference and whether it is at-issue or not-at-issue. Furthermore, it relates the theoretical approaches to empirical results and shows that they diverge quite a lot.

In chapter 8, the previous approaches that analyzed *clefts in discourse* are discussed. I am not the first to analyze the effect of a cleft in discourse. This chapter shows that there is a wide range of functions that the cleft has been claimed to have in discourse, ranging from marking givenness or contrast to establishing discourse coherence or addressing a certain question. Those approaches were the first to ever look at actual occurrences of clefts in novels or newspapers, which gave valuable insights. The approaches discussed in this chapter are quite diverse and formalize their claims to different degrees.

In chapter 9, the *empirical studies* on clefts and their results are summarized and discussed. The chapter includes corpus studies, experiments on exhaustivity, and experiments on contrastivity, which are the main fields of empirical research on clefts. I point out in this chapter that the discourse function of clefts has not yet been investigated much empirically. Nevertheless, the discourse function interfered with the results of many of the experiments.

Chapter 10 is the centerpiece of this thesis and develops a new *analysis of German *es-clefts* in discourse*. It introduces the concept of expectedness, which is argued to explain the acceptability of clefts in some contexts and the unacceptability in other context. The analysis presented in this chapter takes a broader perspective on clefts than many of the previous approaches did, for example by including large parts of the preceding context into the analysis of clefts. Furthermore, the analysis manages to subsume many of the proposed discourse functions under one discourse function.

Finally, chapter 11 concludes the thesis.

Part I

Theoretical Background

Chapter 2

Semantics and Pragmatics of Questions

In this chapter, I provide an overview over the semantics and pragmatics of questions. This and the next two chapters constitute the theoretical background that I consider necessary for the analysis of *es*-clefts in German. For the analysis of clefts in German, I will build a question-based discourse model (see chapter 4 and section 10.3), which requires an understanding of the semantic representation of the questions that arise in discourse. Furthermore, the definition of focus congruence in chapter 3 also relies on the semantic representation of questions. A discourse update in the discourse model, however, is not only based on well-defined semantics questions, it also requires an understanding of what it means to update a context by an interrogative speech act. Moreover, the different linguistic forms of questions affect the acceptability of a cleft sentence in German in different ways. Therefore, it is useful to have an overview of linguistic realizations of questions in German.

In section 2.1 of this chapter, I disambiguate different meanings of the term *question*. I first present different linguistic means realizing questions in different languages. This part also includes an overview over different kinds of questions, such as *wh*-questions, polar questions, embedded questions, and alternative questions. Secondly, I distinguish questions as interrogative speech acts from the semantic object that represents a ques-

tion. Section 2.2, section 2.3, section 2.4, and section 2.5 present different theoretical approaches towards questions as semantic objects. Based on these, section 2.7 discusses pragmatic features of questions, in particular their existence inference and exhaustivity. In section 2.6, I present different subtypes of questions. Those subtypes will turn out quite useful in the model that can predict the difference between clefts and canonical sentences in German (see chapter 10).

2.1 Introduction

In this section, I present an overview on the semantics and pragmatics of questions, distinguishing them from assertions. In the course of that, I will refer to four levels that play a role for differentiating the meaning of questions and assertions: the LINGUISTIC FORM, the DISCOURSE EFFECT, the SEMANTIC OBJECT, and the SPEECH ACT.

Consider first the LINGUISTIC FORM of a question. Questions and assertions are marked by different linguistic means, which are visible on the surface level in most languages. Example (1-a) shows a question, and example (1-b) an assertion in German.

- (1) a. Wen mag Laura?
who.ACC likes Laura
'Who does Laura like?'
- b. Laura mag Natalio.
Laura likes Natalio
'Laura likes Natalio.'

In many languages assertions are marked with a falling final intonation while questions show a rising final intonation. Furthermore, questions and assertions differ in their default word order in many languages. Hence, if we call something *question* on this level we are referring to a string of words that has certain syntactic, morphological, prosodic etc. features that distinguish it from an assertion. Below, I will elaborate on how languages mark questions focusing on German. However, the linguistic form is not always sufficient for disambiguating between questions and assertions, as (2) illustrates.

- (2) Du willst nicht wirklich dieses Auto kaufen.
you want not really this car buy
'You don't really want to buy this car.'

This example can be used for both an assertion and a question with the same intonation and with the exact same form as in (2). It can only be disambiguated by including the discourse context. In a context in which the speaker is an expert on cars while the addressee has no clue about cars, the speaker could assert (2) as a good advice to not buy the car, as in (3).

- (3) Glaub mir! Du willst nicht wirklich dieses Auto kaufen. Es ist eine Schrottkarre!
'Trust me! You don't really want to buy this car. It's junk!'

In a context in which the speaker is surprised about the addressee buying this car, the same string of words represents the question whether the addressee actually wants to buy the car, as in (4).

- (4) Ich kann es gar nicht glauben. Sag mal ehrlich. Du willst nicht wirklich dieses Auto kaufen.
'I cannot believe it. Honestly, you do not really want to buy this car.'

As in this example, such questions frequently involve the speaker's disbelief towards one of the possible answers, here *Yes, I want to buy that car*. Still, example (4) seems to rather be a question than an assertion. Hence, the categorization of (3) as an assertion and (4) as a question must be located on a different level than the linguistic form.

Another approach to distinguish questions from assertions is analyzing their respective DISCOURSE EFFECT. Questions and assertions contribute in different ways to achieving the goal of a conversation. This goal is roughly finding out what the world is like (c.f. Roberts, 2012), or, in other words, singling out our actual world out of a set of possible worlds. Given a set of worlds in which all the mutually shared propositions are true (context set), questions and assertions change this set of worlds in different ways. This topic is discussed in much more detail in section 4.2.

Assertions update the context set, i.e. they exclude certain worlds from the context set, namely those in which the assertion is true. This brings the discourse participants a bit closer to the discourse goal since there are fewer worlds than before that could be the actual world. Questions do not exclude any world from the context set. Instead, they restrict possible updates in future discourse. The only future updates that are allowed are those that are compatible with a possible answer to the question. Example (5) illustrates a future assertive update that is not allowed.

- (5) Did Mary dance? – #If Mary danced, John danced.

The reply in this example is unacceptable because it is compatible with both possible answers to the question, *Mary danced* and *Mary did not dance*. Questions do not themselves exclude worlds, but they predict which sets of worlds may be excluded by new assertions in the future discourse and which may not. Intuitively, questions have an immediate effect on the addressee. Once a question is uttered, it is very strongly expected of the addressee to at least try to answer this question. If the addressee does not want to answer this question, she/he needs to make some effort to mark this deviation from the expected direction of the discourse, as in (6).

- (6) Speaker A: Was hast du gerade in meinem Zimmer gesehen?

'What have you just seen in my room?'

Speaker B:

- a. ?Was versteckst du in deinem Zimmer?
what hide you in your room
'What are you hiding in your room?'
- b. Was versteckst du denn in deinem Zimmer?
what hide you Q.PRT in your room
'Why, what are you hiding in your room?'
lit. *'What are you Q.PRT hiding in your room?'*

Asking a different question instead of answering speaker A's question is only acceptable, when it is marked with the question particle *denn* in German, as in (6-b). Strictly

speaking, speaker B is not supposed to ask such a question. These and other observations about the behavior of questions and assertions in discourse motivate different semantic analyses of questions because the semantics of questions and assertions, respectively, is derived from their discourse effects. The different semantic approaches will be presented in this chapter. Questions in discourse will be discussed in more detail in chapter 4.

The term *question* can also refer to a SEMANTIC OBJECT. The semantic meaning of an assertion is usually represented as a proposition. Following Stalnaker (1976), a proposition is a set of worlds, namely those worlds in which the declarative sentence is true. In inquisitive semantics, propositions are type-raised to sets of sets of worlds (see section 2.5). The meaning of a question, in contrast, can be quite different objects depending on which approach one wants to follow.

In the framework of structured meanings, questions are represented by an incomplete proposition. The idea is that questions point out that there is some information missing that needs to be filled in the discourse. In alternative semantics, questions are assumed to be raising alternative utterances for the addressee to choose from (Rooth, 1985; Hamblin, 1973; Karttunen, 1977). The partition approach also involves the idea of incomplete propositions but adds an equivalence relation between worlds or world-time-pairs (Groenendijk and Stokhof, 1984). This results in a representation of questions involving sets of mutually exclusive alternatives that necessarily cover all possibilities.

Finally, there is inquisitive semantics, in which both questions and assertions are represented as sets of sets of worlds (Ciardelli et al., 2013; Groenendijk and Roelofsen, 2009). However, the representations of questions and assertions differ with respect to formal features in the following way: Questions include more than one maximal set while assertions contain just one maximal set. This amounts to something very similar to alternatives again, i.e. having one set (assertion) or more than one set (question) to choose from. However, inquisitive semantics adds informativity to the definition of an assertion and inquisitiveness to the definition of questions. These notions describe their function in discourse. Informativity, roughly, says that an assertion should exclude some worlds from the context set. Inquisitiveness disallows an alternative that covers all possible worlds for

a question since this would mean that the addressee does not necessarily need to exclude worlds with her/his answer.

In the inquisitive approach, the semantic object representing questions already covers a lot of its discourse effects and its use-conditions by requiring informative or inquisitive propositions. In the other approaches, discourse effects are modeled to a greater degree by posing discourse rules that apply to the semantic object. The different semantic approaches to questions will be discussed in section 2.2 – 2.5.

Ultimately, we can refer to questions as INTERROGATIVE SPEECH ACTS. The representation of speech acts is used to model what is actually happening in the world when a question is asked or an assertion is made. Krifka (2011) identifies the speech act of a question as expressing a lack of information to the addressee. It can be seen as an invitation to the addressee to provide that information. Whether or not the addressee actually follows the invitation is not part of the speech act or the semantics of questions, at least not entirely. This is illustrated by example (7).

- (7) Did Mary dance? – #I want to buy a cake this afternoon.

This answer, just like (5), is compatible with worlds in which Mary danced and worlds in which she did not dance. Hence, the addressee did not follow the invitation to provide the missing information. A further reason why the answer is odd is that the provided information is irrelevant with respect to the preceding context. The sentence would be almost as odd in the context of the assertion that could answer the question, as in (8).

- (8) Mary danced. – #I want to buy a cake this afternoon.

Thus, the oddness of the reply is only partly due to requirements of the question. There are other discourse rules that regulate when discourse participants accept a discourse move and when they do not. This is dependent on the relevance of the discourse move (c.f. Roberts, 2012). Relevance effects are discussed in chapter 4, which will deal with this aspect as part of the discourse model.

Above, we saw that the different semantic approaches incorporated the proposed speech act to different degrees. I will follow Stenius' (1967) distinction of a speech act operator and a sentence radical. The latter is the semantic object, and the former models the speech act and how the speech act is derived from the semantic object. This operator can apply to incomplete proposition or to alternatives, which has the effect of expressing the request to fill in the missing information (structured meanings), to choose an alternative (alternative semantics), or to locate the actual world in one of the alternatives in the partition (partition approach). In inquisitive semantics, it is more complex to apply the operator to the semantic object, but the result should finally be similar as in the partition approach, i.e. locating the actual world in a set of worlds, or maybe in more than one, but not all, sets of worlds.

In the remainder of this introduction, I will elaborate on the linguistic form of questions, particularly in German, and on the distinction between the semantic object and the speech act of a question.

Linguistic form Questions can be realized by quite different linguistic expressions. In German, questions are usually realized by a final rising intonation, and sometimes a change of word order, or a *wh*-pronoun, also called a question pro-form⁴. The linguistic form of the question differs with respect to what kind of question we are dealing with. Therefore, I will now go through the different types of questions, and will show how those questions are realized in German. I will also provide some examples of how other languages express questions differently than German does.

The first type is the POLAR QUESTION (or *yes/no*-question), exemplified in (9). For comparison, the corresponding assertion is provided in (10).

- (9) Mag Laura Natalio?
likes Laura Natalio
'Does Laura like Natalio?'

⁴I will use the term *wh*-pronoun in this thesis.

- (10) Laura mag Natalio.
 Laura likes Natalio
 ‘Laura likes Natalio.’

A polar question, like (9), just allows two answers, and the person asking is only considering two options. In (9), it is either *Yes*(, *Laura likes Natalio*) or *No*(, *Laura does not like Natalio*) (hence, *yes/no*-question). In German, the polar question is marked by a rising intonation at the end of the sentence and movement of the finite verb into a sentence initial position. The assertion in (10), in contrast, is marked by a falling final accent and has a verb-second word order. Turkish, for instance, forms polar questions with the clitic vowel-harmonic -*mi*, as the example in (11) shows.

- (11) Ali dün iskambil oyna-di mi?
 Ali yesterday card play.PAST Q.PRT
 ‘Did Ali play cards yesterday?’ (Kamali and Krifka, 2020:6)

This clitic is obligatory for polar questions in modern Turkish (Kamali and Krifka, 2020). As the English translations of (9) and (11) indicate, English requires *do*-support in polar questions that do not include an auxiliary verb. Hence, a finite form of *do* fills the sentence initial position in English polar questions. In case there is an auxiliary verb, it is moved to the sentence initial position instead, as in (12).

- (12) Have you met Ted?

Some languages have been claimed to not distinguish assertions from polar questions, at least not on the level of the linguistic form (e.g. Mitterer and Stivers, 2007). We saw in example (2) above that also in German there are cases that are ambiguous between a question and an assertion on the surface level.

The second kind of question is the *wh*-QUESTION or CONSTITUENT QUESTION, as illustrated in (13) and (14) for German.

- (13) Wer mag Natalio?
 who_{NOM} likes Natalio
 ‘Who likes Natalio?’

- (14) Wen mag Laura?
 who_{ACC} likes Laura
 ‘Who does Laura like?’

Constituent questions or *wh*-questions are used to ask for specification of a constituent by using a *wh*-pronoun. Example (13) illustrates a case of a subject *wh*-question, since this question is used to ask about the subject. For such a question, German simply fills the subject position with the nominative *wh*-pronoun.⁵ Moreover, the *wh*-question also receives a rising intonation at the end of the sentence. Example (14) represents an object *wh*-question. Parallel to (13), the object position is filled with the accusative *wh*-pronoun *wen*. Usually, this *wh*-pronoun is moved to the sentence initial position, as it is the case in (14). In spoken German, however, cases like (15) can be found as well, again with a rising final accent.

- (15) Laura mag wen?
 Laura likes who.ACC
 ‘Laura likes who?’

In this question, the *wh*-pronoun is left *in-situ*. This variant of a *wh*-question, however, is only used in specific contexts, such as clarification questions, echo-questions, or for the purpose of expressing surprise or disbelief. Mandarin Chinese, in contrast, is a language, that always leaves question pro-forms *in-situ*, as indicated in (16) taken from Cheng (1991:113).

⁵A simple replacement of the subject with a *wh*-pronoun is not always sufficient in order to form a question, as illustrated by the following examples.

- (i) a. Gestern hat Laura geschlafen.
 yesterday has Laura slept
 ‘Yesterday, Laura slept.’
- b. Gestern hat wer geschlafen?
 yesterday has who slept
 ‘Yesterday, who slept?’
- c. Wer hat gestern geschlafen.
 who has yesterday slept
 ‘Who slept yesterday?’

When there is an adverbial in the prefield, such as *gestern* (‘yesterday’) in (i-a), replacing the subject with the *wh*-pronoun only leads to an echo question, as in (i-b), instead of a ‘normal’ question. The default question to ask for the constituent *Laura* in (i-a) is given in (i-c).

- (16) Hufei chi-le sheme (ne)
 Hufei eat.ASP what (WH.PRT)
 ‘What did Hufei eat?’

As the example illustrates, Mandarin Chinese also uses a *wh*-pronoun *sheme*, which is, unlike in English or German, not fronted. Some languages require the *wh*-pronoun to move to a specific focus position. In Hungarian, it is moved to a position immediately preceding the verb (Szabolcsi, 1981), while Chadic requires movement to a postverbal position (see Tuller, 1992).

The list of constituents that can be asked for by a *wh*-pronoun is quite elaborate in German. Next to time and place adverbials, as in (17) and (18) respectively, Krifka (2011:1744) points out that it is even possible to ask for ordinals in German, which is not possible in English, as illustrated by (19).

- (17) Wann hat Natalio geschlafen?

‘When did Natalio sleep?’

- (18) Wo hat Natalio geschlafen?

‘Where did Natalio sleep?’

- (19) Den wievielten Geburtstag feiert Maria?
 the how.many.ORD birthday celebrates Maria
 lit. ‘The how manyth birthday is Maria celebrating?’

Some sub-constituents in German can only be *wh*-moved by pied-piping (see Horvath, 2017), i.e. by fronting a larger constituent than just the *wh*-pronoun. Such examples are presented in (20) and (21).

- (20) a. [In welchem Haus] wohnt Laura?
 in which house lives Laura
 ‘Which house does Laura live in?’
- b. *Welchem₁ wohnt Laura in t₁ Haus?
 which lives Laura in t₁ house
 ‘Which house does Laura live in?’

- (21) a. [Wessen Katze] mag der Peter?
 whose cat likes the.NOM Peter
 ‘Whose cat does Peter like?’
- b. *Wessen₁ mag der Peter *t*₁ Katze?
 whose likes the.NOM Peter *t*₁ cat
 ‘Whose cat does Peter like?’

Example (20) represents a case of prepositional pied-piping. As (20-b) indicates, it is ungrammatical to front the argument of the preposition *in* alone. The question is only well-formed when the whole prepositional phrase is moved, as in (20-a). The English translation, again, shows that English differs from German by not requiring pied-piping in this case. The second example constitutes an example of pied-piping with a possessive phrase. Again, moving just *wessen* (‘whose’) is ungrammatical, as in (21-b), while moving the whole possessive phrase yields a well-formed question, as in (21-a).

Some constituents, such as verb phrases cannot be asked for by just a *wh*-constituent. They need a *wh*-pronoun and a higher order verb, such as *machen* (‘make’) in (22).

- (22) Was hat Laura gemacht?
 what has Laura made
 ‘What did Laura do?’

Moreover, it is possible to ask for multiple constituents in one question by using multiple *wh*-pronouns, as in (23).

- (23) a. Wer mag wen?
 who.NOM likes who.ACC
 ‘Who likes whom?’
- b. Wem hat Laura was geschenkt?
 who.DAT has Laura what.ACC given.as.a.present
 ‘Whom did Laura give what as a present?’

In German, only one *wh*-pronoun is fronted, while the second *wh*-pronoun stays in-situ. This is particularly obvious in example (23-b), where the dative *wh*-pronoun is fronted and the accusative *wh*-pronouns stays in-situ.

The last category of questions that I want to consider here is alternative questions, such as the one in (24).

- (24) Mag Laura Natalio oder Felipa?
 likes Laura Natalio or Felipa
 ‘Does Laura like Natalio or Felipa?’

Alternative questions in German, but also in English, show a different intonation than other questions. In (24), there is a rising accent on *Natalio* and a falling sentence final accent on *Felipa*. In case of a rising final intonation, we would be dealing with a polar question. Alternative questions are like constituent questions that reduce the acceptable answers to a few alternatives. Example (24) corresponds to the question *Who does Laura like?* just allowing the answers *Laura likes Natalio*, *Laura likes Felipa*, and *Laura likes Natalio and Felipa*.⁶

The kinds of questions discussed so far are called ROOT QUESTIONS (Krifka, 2011). Next to root questions, there are embedded questions, which involve a question embedding predicate, such as in (25) and (26).

- (25) Lena weiß, ob Laura Natalio mag.
 Lena knows whether Laura Natalio likes
 ‘Lena knows whether Laura likes Natalio.’
- (26) Lena fragt sich, wer Natalio mag.
 Lena asks herself who Natalio likes
 ‘Lena is wondering who likes Natalio.’

Example (25) embeds a polar question (*Does Laura like Natalio?*) under the predicate *know*. Example (26) embeds a *wh*-question (*Who likes Natalio?*) under *wonder*. Embedded questions show the word order that subordinate clauses in German usually do with the question word as the complementizer, *ob* (‘whether’) for a polar embedded question and the respective *wh*-pronoun for embedded *wh*-questions. Example (26) is of particular interest for this thesis since it provides a good context for clefts. The root question, in

⁶Possibly, it also allows for the answer *Laura likes neither Natalio nor Felipa*. This will be discussed in section 2.3.

contrast, constitutes a rather unacceptable context for the cleft. Compare example (27) and (28).

- (27) Lena fragt sich, wer Natalio mag. – Es ist Simon, der ihn mag.
'Lena is wondering who likes Natalio. It is Simon who likes him.'

- (28) Wer mag Natalio? – ?Es ist Simon, der ihn mag.
'Who likes Natalio? – ?It is Simon who likes him.'

That clefts do not provide good answers to explicit questions was discussed and empirically attested by Velleman et al. (2012) and Destruel and Velleman (2014), among others, and will be picked up in section 8.6.

This overview covered the main types of questions and their linguistic form in German, and gave some impressions of how other languages realize questions. In the following, I will come back to the distinction of the semantic object, representing a question, from the interrogative speech act, performing the question.

Semantic object vs. speech act As mentioned above, Stenius (1967) divides each utterance into a sentence radical, which denotes a semantic object, and a speech act operator. Applying the speech act operator to the sentence radical yields the communicative act of asking a question (Krifka, 2011). I will refer to the semantic object representing a question as $?φ$ or q , while I will use $A_{?φ}$ or A_q for the speech act of asking $?φ$. Further, I will be using $φ$ for propositions, and $A_φ$ for assertive speech acts. For assertions, the sentence radical is a proposition $φ$, to which the assertion speech act operator ASSERT applies, as in (29-a). In (29-b), an example of an assertion is analyzed as a speech act.

- (29) a. ASSERT($φ$)
b. Laura likes Natalio.
ASSERT(**like(Laura,Natalio)**)

According to Stenius (1967), interrogative speech acts are composed of the sentence radical in form of an incomplete proposition, which is represented by $?φ$, and the interrogative speech act operator QUEST, as in (30-a).

- (30) a. QUEST($?φ$)
 - b. Who likes Natalio?
- QUEST(**like(x,Natalio)**)

Following Krifka (2011), I argue that this operator can be applied to different representations of questions, apart from incomplete propositions, as we will see for the different semantic analyses in section 2.2–2.5. The only difference is that QUEST is doing slightly different things depending on which semantic object it is applied to. Krifka (2011:1743) describes the interrogative speech act as “expressing lack of information of a specified type”. How this type is specified, depends on the semantic analysis of questions.

Embedded questions, in contrast, are analyzed as involving no QUEST speech act operator, as (31) indicates.

- (31) Lena knows who likes Natalio.
- ASSERT(**know(Lena,like(x))**)

The predicate *know* directly combines with the interrogative sentence radical by taking it as its argument. Accordingly, embedded questions do not realize the speech act of expressing a lack of information, which is an effect that can be observed for examples like (31). However, both unembedded and embedded questions involve the interrogative sentence radical. In other words, they involve the same semantic object (interrogative sentence radical) but express different speech acts.

Contrary to this, there are also cases where an embedded question may realize an interrogative speech act, as in (32).

- (32) Ich frage mich, wer Natalio mag.
 I ask myself who.NOM Natalio likes
 ‘*I wonder who likes Natalio.*’

This example can be used to express the request to provide an answer to the embedded question *Who likes Natalio?*. This possibility seems much more prominent with the embedding predicate *wonder* than with *know*. Therefore, Krifka (2001b) puts these two predicates in different categories of embedding verbs. He proposes that certain question embedding predicates, such as *wonder*, do not embed interrogative sentence radicals but interrogative speech acts. Other examples of such INQUISITIVE PREDICATES are *want to know* and *ask* (Krifka, 2011). The difference between inquisitive predicates and other question embedding predicates also shows syntactic differences in some languages. For German, the following differences can be observed.

- (33) a. Wer mag Natalio, fragte Lena sich.
who.NOM likes Natalio asked Lena herself
'Who likes Natalio, Lena was wondering.'
- b. *Wer mag Natalio, wusste Lena.
who.NOM likes Natalio knew Lena
'Lena knew who liked Natalio.'
- c. Wer Natalio mag, wusste Lena.
who.NOM Natalio likes knew Lena
'Lena knew who liked Natalio.'

Inquisitive predicates can embed a question with a verb-second word order, as in (33-a). This word order is not well-formed for *wissen* ('know') in German, as (33-b) indicates. For non-inquisitive embedding verbs, only verb-last word order of subordinate clauses is allowed, as in (33-c).

Another perspective on interrogative speech acts is concerned with how asking a question affects the discourse, and what requirements are imposed on the common ground. In section 4.2, this is captured in the definition of an interrogative update of the common ground or the context set. There, I define context updates for speech acts since it seems that the discourse effect of a question only arises when the speech act is actually performed, as the examples (34) and (35) indicate.

(34) A: When will you finally go on a date with me?

B:

a. Now!

b. #Then ask me!

(35) A: I wonder when you will finally go on a date with me.

B:

a. Now!

b. Then ask me!

In (34), the discourse effect of the question is that only a direct answer is appropriate. Therefore, (34-b) is unacceptable given that it does not provide such an answer. Obviously, it is also unacceptable because it demands a speech act that has already been performed. However, for the inquisitive predicate *wonder* in (35), both answers are acceptable. Hence, on some level the speech act must be involved, but it can also be interpreted as not performed, as indicated by the acceptability of (35-b). Hence, the speech act of a question has immediate effects on the discourse.

After having broadly distinguished the linguistic form of questions, their discourse effect, the semantic object representing a question, and the speech act of asking a question, I will now discuss some cases of questions which do not display question features on all of these levels but which are still considered questions. There are, for instance, ways to express a question which use none of the typical question markings in German, such as (36).⁷

(36) Du möchtest gerne die ganze Wohnung putzen.
you would like gladly the whole flat clean
'You would like to clean the whole flat.'

This example can express a question even without a rising intonation, e.g. in a context in which the speaker is surprised that her/his flatmate does not have a problem with cleaning the whole flat. By uttering (36), the speaker expresses that she/he cannot believe that,

⁷This example behaves like (2), mentioned above.

and she/he is asking for a confirmation. To illustrate the underlying question, (36) could be rephrased as in (37).

- (37) Möchtest du wirklich gerne die ganze Wohnung putzen?
would.like you really gladly the whole flat clean
'Would you really like to clean the whole flat?'

On the semantic level, example (36) can be analyzed just like (37), as involving an interrogative sentence radical. Moreover, (36) seems to perform a kind of interrogative speech act, although possibly not exactly the same as (37) does. The speech act performed in (36) rather asks for confirmation of an already given information and might also express disbelief. What is crucial here is that (36) does not show any question markings and is still perceived as asking a question. Hence, it is important to separate the linguistic expression of a question from its semantics and from interrogative speech acts.

Example (38) represents the opposite case of a sentence that has formal features of a question, but does not represent the speech act of asking a question.

- (38) Seid ihr verrückt geworden?
are you insane become
'Have you gone insane?'

This is an example of a rhetorical question, which is not used to express the lack of information because the speaker knows the answer already (see Biezma and Rawlins (2017) for an analysis of rhetorical questions). Hence, what superficially looks like a question does not express an interrogative speech act. It might involve the semantics of a question, though (c.f. Caponigro and Sprouse, 2007). Another example that has the form of a question but realizes a different speech act is the following.

- (39) Could you pass me the salt?

This is the polite way to express a command in English, and also in German. It has the formal features of a question, such as movement of the auxiliary verb and a rising intonation.

At the end of this introduction, I want to point out the importance of these distinctions for building a theory of discourse. Many discourse models, including the one I am proposing, refer to questions as a central component. Whether one wants to follow Roberts (2012), who introduces the question under discussion (QUD), or Büring (2003), who uses discourse trees, or Onea (2016) and van Kuppevelt (1995), who use implicit questions, one should always know which level of *question* is appropriate. What are the questions in Roberts' QUD stack? Is it semantic objects or interrogative speech acts? The same could be asked about an update of a context with a question. Section 10.2 and 10.3 will elaborate on this issue.

In the following, I will present a summary of the main approaches towards the semantics of questions. An approach that suggests itself is one which defines questions via answerhood. This seems to be the natural way to approach questions since the semantics of assertions – and answers are assertions in most cases – is well-understood, and can be exploited to derive a semantics for questions. The role of answers for the semantics of questions has been treated and formalized in different ways. I will distinguish four main approaches: the FUNCTIONAL APPROACH, the ALTERNATIVE SEMANTICS APPROACH, the PARTITION APPROACH, and the INQUISITIVE APPROACH.

2.2 Functional Approach

In the FUNCTIONAL approach, a question is treated as an incomplete proposition which is missing the “part” that is questioned.⁸ In other words, the question is represented as an assertion that misses one argument. Following (Krifka, 2011:1754), my example is formally represented as in (40).

$$(40) \quad \llbracket \text{Who likes Natalio?} \rrbracket = \lambda i \lambda x \in \text{PERSON}_i [\text{like}_i(x, \text{Natalio})]$$

The variable i is an index that models the dependence of the meaning on the world and time, represented as a tuple $\langle w, t \rangle$. This approach allows a semantic definition of the

⁸This holds just for *wh*-questions, not polar questions.

relation of the question and a term answer/elliptical answer. If we assume an elliptical answer to (40), such as *Laura*, the relation between the question and the answer is derived by functional application (FA), as in (41).

(41) Who likes Natalio? – Laura.

$$\begin{aligned}
 & [[\text{Who likes Natalio?}]] ([[Laura]]) \\
 & = \lambda i \lambda x \in \text{PERSON}_i [\text{like}_i(x, \text{Natalio})] (\text{Laura}) \\
 & = \lambda i [\text{like}_i(\text{Laura}, \text{Natalio})]
 \end{aligned} \tag{FA}$$

The function in (40) representing the question applies to the denotation of *Laura* yielding TRUE depending on the index i . The relation between a question and a full answer, such as *LAURA likes Natalio*, is more indirect. It requires the function to distinguish the focused part (*Laura*) of the answer from the backgrounded part (*likes Natalio*). Then, it can apply to the focused part like it was described for elliptical answers. Focus will be discussed in detail in chapter 3.

For polar questions, Krifka (2011) assumes a function which takes truth-values as its argument, and which has two operators in its domain, as illustrated in (42).

$$(42) [[\text{Does Laura like Natalio?}]] = \lambda i [\lambda f \in \{\lambda t.t, \lambda t.\neg t\} [f(\text{like}_i(\text{Laura}, \text{Natalio}))]]$$

The operator $\lambda t.t$, the identity function, corresponds to the answer *yes*, and the operator $\lambda t.\neg t$ corresponds to the answer *no*. Following Krifka (2011:1756), the polar question meaning applies to the answer *yes*, as illustrated in (43).

(43) Does Laura like Natalio? – Yes.

$$\begin{aligned}
 & [[\text{Does Laura like Natalio?}]] ([[yes]]) \\
 & = \lambda i [\lambda f \in \{\lambda t.t, \lambda t.\neg t\} [f(\text{like}_i(\text{Laura}, \text{Natalio}))]](\lambda t.t) \\
 & = \lambda i [\text{like}_i(\text{Laura}, \text{Natalio})}
 \end{aligned} \tag{FA}$$

The answer *no* applies in the same way but picking the function $\lambda t.\neg t$. These functional representations of questions describe the interrogative sentence radicals (following Stenius,

1967), which can then serve as arguments for a speech act operator. Under the operator QUEST, a *wh*-question, such as in (41), amounts to the speaker trying to get the addressee to specify the values for x under which the function yields TRUE, as well as those values for which it yields FALSE. Applying QUEST to (43), yields a request of the speaker to specify which function out of $\lambda t.t$ and $\lambda t.\neg t$ makes the proposition true and which makes it false.

As mentioned above, question embedding predicates, such as *know* in (44), are assumed to take the interrogative sentence radical as their argument.

- (44) Lena knows who likes Natalio.

Such examples are easy to model in this approach. Krifka (2011) assumes a sentence like (44) to mean that Lena either knows the whole range of x that make the incomplete proposition true or false, or she knows of at least one x that makes the proposition true. These two versions describe the exhaustive and the non-exhaustive reading, respectively, which I will get to in section 2.7.

2.3 Alternative Semantics Approach

The alternative semantics approach, introduced by Hamblin (1973), Karttunen (1977), and Rooth (1985), presents an approach to the semantics of questions that establishes a direct relation between questions and non-elliptical answers. Basically, the meaning of a question is represented as a set of propositions that correspond to the (possible) answers. The meaning of *Who likes Natalio?* is illustrated informally in (45).

- (45) $\llbracket \text{Who likes Natalio?} \rrbracket = \{ \text{Laura likes Natalio}, \text{Simon likes Natalio}, \text{Dani likes Natalio}, \dots \}$

This set is also called the set of Q-ALTERNATIVES, based on Rooth's (1985) alternative semantics. The question in (45) raises alternatives of the form x *likes Natalio* for all x

from a certain domain, e.g. the domain of persons. If we disregard focus in the answer for the moment, the elements in the set of q-alternatives have exactly the same form as the semantic representation of each possible answer to the question, namely the form of a proposition. Hence, each answer is an element of the q-alternatives. In contrast to the functional approach, the relation between questions and their answers is not modeled via functional application but via the an “element-of” relation between q-alternatives and the semantic representation of each answer (to be revised below).

This approach also covers elliptical answers assuming that those have an underlying full proposition and that the linguistic material is only deleted on the surface level, as (46) indicates.

- (46) Laura likes Natalie.

In chapter 3, I will spell out how focus interacts with the q-alternatives, and it will turn out that only the ordinary semantic value (c.f. Rooth, 1992) of the answer describes a proposition that is an element of the q-alternatives. The focus semantic value of the answer will itself be a set of alternatives. We will see that the subset relation, instead of the “element-of” relation, will be crucial in the end. For now, I point out the parallelism between the elements in the q-alternatives and the semantic representation of (possible) answers.

There is some disagreement in the literature about what kind of answers, i.e. what kind of alternatives, should be included in the q-alternatives in order to adequately represent the meaning of a question. Hamblin (1973) assumes that the q-alternatives include all possible answers, even those that are false. Following Hamblin, the meaning of (45) can be formally represented as in (47).

- (47) $\llbracket \text{Who likes Natalie?} \rrbracket = \{\llbracket x \text{ likes Natalie} \rrbracket \mid \llbracket x \rrbracket \in \text{PERSON}\}$

This question meaning covers all alternatives of the form x *likes Natalie* without distinguishing whether it is actually true. The requirement of x being a person is argued for

by Rooth (1992). He states that the question pronoun *who* is only used for asking for a person and, thus, the domain for x in (47) must be restricted. This does not hold for the question pronoun *what*, as in (48-a). Its formal representation in (48-b) must, thus, contain the restriction of x not being a person.

- (48) a. What does Laura like?
b. $\llbracket \text{What does Laura like?} \rrbracket = \{\llbracket \text{Laura likes } x \rrbracket \mid \llbracket x \rrbracket \notin \text{PERSON}\}$

The representations in (47) and (48-b) also include answers to the respective questions that are false. However, we might just be concerned with the true answers of a question when it comes to its meaning. Hence, Karttunen (1977) argues that the q-alternatives should contain only those propositions that represent true answers to the question.

- (49) $\llbracket \text{Who likes Natalio?} \rrbracket = \{\llbracket x \text{ likes Natalio} \rrbracket \mid \llbracket x \rrbracket \in \text{PERSON} \wedge \llbracket x \text{ likes Natalio} \rrbracket = 1\}$

It is conspicuous that the q-alternatives resemble the function that the functional approach derived for the semantic representation of questions. Example (50) shows the q-alternatives of (47) in Krifka's (2011:1757) notation, to make the resemblance even more apparent.⁹

- (50) $\llbracket \text{Who likes Natalio?} \rrbracket = \{\lambda i \text{ [likes}_i(x, \text{Natalio})] \mid x \in \text{PERSON}\}$

Indeed, Krifka (2011) argues that the question function can always be transformed into a set of propositions, just by taking the function to define the propositions in the q-alternatives. In other words, the question function is the characteristic function of the q-alternatives set. In contrast, not every set of q-alternatives can be transformed into a function that represents the question in a sensible way. Beaver and Clark (2008:91–92) present a formal mapping from structured meanings to alternatives semantics, and they show that the reverse is not possible. I will not go into the formal details, but what is crucial is that the functional approach is generally capable of expressing more than alternative semantics. A mapping from functions to alternatives will necessarily map two

⁹I will stick to the notation used in (47) in the remainder of this thesis.

different functions onto the same set of alternatives. Hence, the reverse mapping, from alternatives to functions, will not include this difference anymore.

The alternative semantics approach to questions, like the functional approach, can also account for polar questions and embedded questions. Hamblin (1973) proposes the following semantics for polar questions.

- (51) $\llbracket \text{Does Laura like Natalio?} \rrbracket = \{\llbracket \text{Laura likes Natalio} \rrbracket, \llbracket \text{Laura does not like Natalio} \rrbracket\}$

This analysis assumes that the meaning of polar questions is represented by a set that contains just two alternatives, a proposition and its negation. This proposition corresponds to the positive full-sentence answer, while its negation corresponds to the negative full answer. The analysis of embedded questions is very similar to the one in the functional approach. For an embedded question, such as (44), we can assume that question embedding predicates, such as *know*, apply equally well to sets of propositions. They simply functionally apply to every proposition in that set.

Krifka (2011) points out some short-comings of the alternative semantics approach, for instance that it cannot account for *yes* and *no* as term answers to polar questions in a direct way (see Krifka (2011:1760) and Krifka (2001a) for more details on the comparison of the functional approach and the alternative semantics approach). This thesis will not be concerned with polar questions in much detail, but rather with *wh*-questions. The reason is that clefts always address *wh*-questions due to their narrow focus (see section 8.5).

2.4 Partition Approach

This approach is an extension of the previous approach. It makes yet another prediction about the kind of alternatives that are relevant, namely only those alternatives that correspond to complete answers. This approach goes back to work by Groenendijk and Stokhof (1984). This approach involves a lot of formal details which make valuable predictions, but which I will not discuss here. I interpret their representation of questions as

a partition with mutually exclusive alternatives. Recall the example *Who likes Natalio?* in a context in which the only persons are Laura and Simon (Natalio is a cat). In this context, the alternatives proposed by Groenendijk and Stokhof are the ones in (52).

- (52) {*Laura and Simon like Natalio, Only Laura likes Natalio, Only Simon likes Natalio, Nobody likes Natalio* }

These alternatives partition the context set (the set of worlds compatible with the propositions in the common ground), as illustrated in Figure 2.1.

L & S like N	only L likes N
only S likes N	nobody likes N

Figure 2.1: Example partition for *Who likes Natalio?* with PERSON = {L, S}

Each cell contains only those worlds in which the respective proposition is true. Given that the alternatives are mutually exclusive, there is no world that can be contained in two or more cells of the partition. Each cell only contains worlds in which one and the same answer is true. This is formally modeled by an equivalence relation on worlds or world-time-pairs (c.f. Groenendijk and Stokhof, 1984). The underlying idea is that questions are asked in order to find out which cell of the partition contains the actual world.

The partition approach is the first approach that accounts for the fact that *nobody* is a well-formed reply to a constituent question. Moreover, this approach takes the standard case of an answer to be an exhaustive answer, which I will get to below. Formally, this means that Groenendijk and Stokhof (1984) need to introduce an exhaustification operator with the semantics of *only*, that is applied to each proposition in the partition and that depends on the focus (see section 2.7 for exhaustivity in questions, and section 3.2 for more details on the focus-sensitive operator *only*).

The partition approach can also account for the inference pattern in (53).

(53) Lea knows who likes Natalio.

Maria does not like Natalio.

Lea knows that Maria does not like Natalio.

The existence of this inference is still debated and empirically investigated. There is, for instance, some empirical evidence of the availability of an intermediate exhaustive reading of question-embedding *know*, for which the inference in (53) would not hold (see Cremers and Chemla (2016) for English, and Fricke et al. (to appear) for different German embedding verbs). Under this reading, the sentence *Lea knows who likes Natalio* only requires Lea to know about all the individuals who actually like Natalio that they do. However, she might have uncertainties about those individuals that in fact do not like Natalio. Hence, it is possible that she does not know whether Maria liked Natalio. In that case, the conclusion in (53) would not follow.

I will still take (53) to be a valid inference for now. Given that assumption, the partition approach is so far the only approach that can account for this inference without having to introduce additional operators or rules. The reason is that the inference only follows when the embedded question *who likes Natalio* is interpreted exhaustively, which is only required in the partition approach. Hence, the inference immediately follows from the semantics of questions.

Comparing the partition approach to the alternative semantics approach, notice that the q-alternatives can be transformed into a partition. The Hamblin-alternatives would produce a partition in which the cells overlap, as in Figure 2.2.

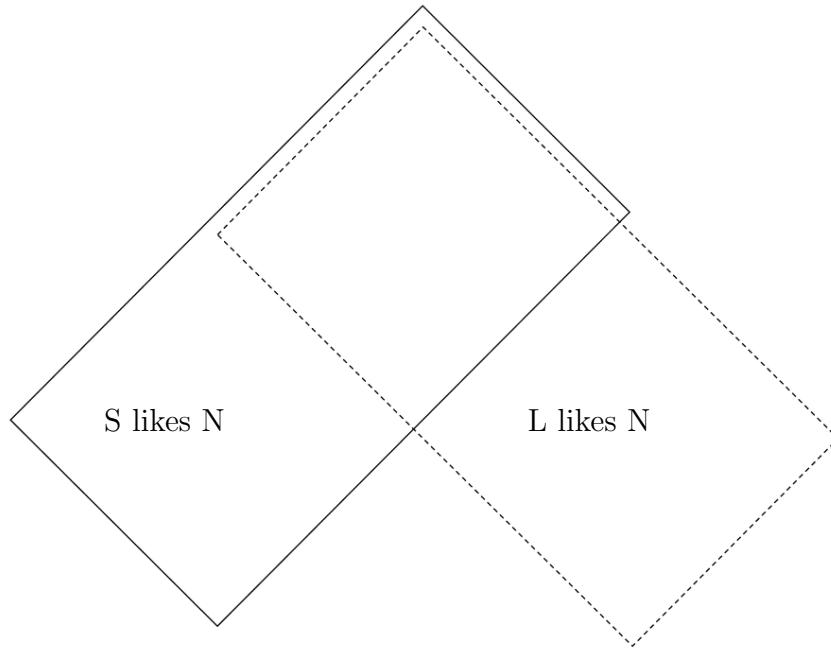


Figure 2.2: Partition derived from Hamblin-alternatives for *Who likes Natalio?* with $\text{PERSON} = \{L, S\}$

A world in which both Laura and Simon like Natalio is contained in both cells, the one determined by *Laura likes Natalio* and the one determined by *Simon likes Natalio*.

The analysis of polar questions is straightforward in the inquisitive approach. A polar question partitions the context set into two parts, as depicted in Figure 2.3.

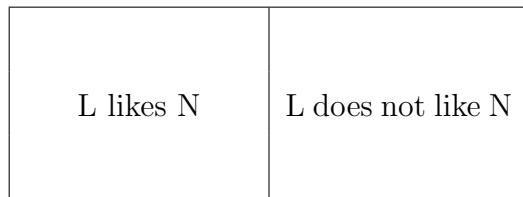


Figure 2.3: Partition for the polar question *Does Laura like Natalio?*

One part of the partition contains all the worlds in which the proposition *Laura likes Natalio* is true and the other part contains those worlds in which it is false, which is equivalent to what alternative semantics predicted. The exhaustification has no effect here, since there are just two mutually exclusive options anyway.

Applying a speech act operator to this representation would be similar to the alternative semantics approach, having the effect of a request to locate the actual world in one of the cells of the partition.

2.5 Inquisitive Approach

INQUISITIVE SEMANTICS is a framework that describes the semantics of assertions and questions establishing a very close relation to their function in discourse (c.f. Groenendijk and Roelofsen, 2009; Ciardelli et al., 2013). Based on dynamic semantics and the idea of a context change potential (c.f. Heim, 1982; Groenendijk and Stokhof, 1991), an assertion is conceptualized as an INFORMATIVE update of the context, and a question as an INQUISITIVE update. Formally, the semantic representations of questions and assertions are of the same type, namely sets of sets of worlds. All the other approaches, mentioned above, assumed distinct types for propositions and questions, which is problematic for the following example.

- (54) Lea knows who likes Natalio and that Laura likes Felipa.

This examples coordinates an embedded question and an embedded assertion, which is problematic for approaches that assume different types for questions and assertions, given that only semantic objects of the same type can be coordinated. Only inquisitive semantics can account for the well-formedness of (54), without having to type-shift the assertion just for this specific case. In inquisitive semantics, the type of assertions is raised to the same type as questions anyways, from just a set of worlds to a set of sets of worlds. Nevertheless, it is still possible to distinguish questions from assertions, which is illustrated in (55).

- (55) a. $\llbracket \text{Laura likes Natalio} \rrbracket = \{\{w \mid \text{Laura likes Natalio in } w\}\}$
b. $\llbracket \text{Who likes Natalio?} \rrbracket = \{\{w \mid x \text{ likes Natalio in } w\} \mid x \in \text{PERSON}\}$

For the assertion in (55-a), we get the set of what was previously defined as a proposition. In other words, an assertion is defined as the set that contains just one element, namely

the set of all worlds in which the assertion is true. This is a simplified version, which I will nevertheless use for illustration in the following. Below in Definition 2, downward closure is added. For the question in (55-b), we get a set that has more than one element. If we assume that PERSON consists of Laura, Dani, and Simon, for instance, the question is represented by (56), also simplified by omitting downward closure.

$$(56) \quad [\text{Who likes Natalio}] = \{\{w \mid \text{Laura likes Natalio in } w\}, \{w \mid \text{Dani likes Natalio in } w\}, \{w \mid \text{Simon likes Natalio in } w\}\}$$

This representation expresses the alternatives corresponding to the answers to the question, just as in alternative semantics and the functional approach. As long as there is more than one element in PERSON, which is reasonable to assume, we will get a set with more than one element for the question, in contrast to the assertion, which always contains just one element. Intuitively, this question representation describes the addressee's choice of an update in the following discourse.

Formally, inquisitive semantics distinguishes questions from assertions via the notion of INFORMATIVENESS. Ciardelli et al. (2013) represent both assertions and questions as type-raised propositions in their version of inquisitive semantics, as we saw above. Their understanding of propositions is based on INFORMATION STATES, given in Definition 1.

Definition 1 (Information State). *An information state s is a set of possible worlds, i.e., $s \subseteq \omega$, where ω is the set of all possible worlds.*

(Ciardelli et al., 2013:461)

An information state can, for instance, be the set of worlds in which all the commonly shared propositions are true. This information state is called the *context set* and will be discussed further in chapter 4. More generally, an information state can be interpreted as representing the worlds that are still possible given a certain context. Of course, this does not follow from Definition 1, but Ciardelli et al. (2013) provide some more formal machinery, that models discourse with such information states. In chapter 4, I will introduce a discourse model, based on Krifka (2015) and Kamali and Krifka (2020),

which basically does the same thing. Onea (2020) shows that Kamali and Krifka's (2020) system can be naturally transformed into inquisitive semantics.

Based on the concept of information states, Ciardelli et al. (2013) define propositions, as in Definition 2.

Definition 2 (Proposition in Inquisitive Semantics). *A proposition is a non-empty, downward closed set of states.* (Ciardelli et al., 2013:465)

This definition does not state anything as to whether a proposition represents a question or an assertion. However, it does say that a proposition should not contain no information state at all. A contradictive assertion, for instance, is not considered a proposition since it would exclude all worlds. Moreover, it requires downward closure under subset relation. This means that every subset of each element in the proposition must itself be an element of the proposition. Downward closure is necessary in order to model entailment patterns correctly, and it makes the system easier to handle on a technical level (see Onea and Zimmermann, 2019:45–46).

Ciardelli et al. (2013) formulate the differences between INFORMATIVE propositions, that correspond to assertions, and INQUISITIVE propositions, that correspond to questions, as defined in Definition 3.

Definition 3 (Informative and inquisitive propositions).

1. *A proposition φ is **informative** iff $\bigcup \varphi \neq \omega$, where ω is the set of all possible worlds.*
2. *A proposition φ is **inquisitive** iff $\bigcup \varphi \not\subseteq \varphi$.* (Ciardelli et al., 2013:466)

According to Definition 3.1, a proposition is informative if the information states contained in it do not cover all possible worlds. Intuitively, this means that when we get new information we should be able to exclude some worlds based on this information. This is what we observe when an assertion is accepted in discourse, as described in section 4.2.

Definition 3.2 says that an inquisitive proposition must contain more than one state (or alternative), in which the actual world could be located in.

2.6 Subtypes of Questions

At the beginning of this section, I explicate the notion of a POSSIBLE QUESTION, before getting to more specific subtypes. Depending on the analysis, everything is a possible question provided that it corresponds to a semantic object which fulfills the requirements posed for questions by the respective theory. The semantics proposed for questions, by any of the approaches above, even allows semantic representations for questions that might not be syntactically well-formed in a language. Recall example (19), which was not well-formed in English, as (57) indicates.

- (57) *The how manyth birthday of his son is John celebrating?

However, this question can be represented as a well-formed semantic object, for instance using the Rooth-Hamblin system as in (58).

- (58) $\{\llbracket \text{John is celebrating the } x^{\text{th}} \text{ birthday of his son} \rrbracket \mid \llbracket x \rrbracket \in \text{NUMBER}\}$

It will turn out that we do want such a question to be expressed semantically, even if it is not syntactically well-formed. The explanation is related to focus congruence with respect to a question. In English, it is also possible to have a narrow focus on an ordinal, as in (59).

- (59) John is celebrating the SECOND_F birthday of his son.

Anticipating the discussion in section 3.2, focus is analyzed as an answer to an (implicit) question, which would be the one represented in (58). If syntactically not well-formed questions were not allowed in general, we would incorrectly predict that English should not allow a focus as in (59). Hence, the semantics of questions allows for much more questions than those that are possible to find a linguistic form for. However, this should not be problematic. In principle, it is not impossible for an English speaker to request exactly the information that a German speaker could request with just one question. The only difference is that the English speaker needs more steps to form a question asking for

the same information that a German speaker can address with one question. One way to ask this would be the question in (60), which assumes that John's son is between 1 and 5 years old.

- (60) Is John celebrating the first, the second, the third, the forth, or the fifth birthday of his son?

Another way to formulate the question would be the following.

- (61) Given that John is celebrating his son's birthday, how old did his son turn?

The most likely thing an English speaker would do is use an underspecified question, which is disambiguated through the context, such as (62).

- (62) Which birthday of his son's is John celebrating?

It is not very likely that the speaker wants to know whether it was, for instance, the best, the worst or the most beautiful birthday although this would be a possible answer to (62). There are more examples of possible answers that are congruent to syntactically ill-formed questions. The example in (63) violates *wh*-movement restrictions.

- (63) a. That Laura liked NATALIO surprised Peter.
b. *Who₁ did that Laura liked *t*₁ surprised Peter?

In this example, the *wh*-pronoun moves out of a syntactic island (as introduced by Ross, 1967), which makes the resulting question ungrammatical. Nevertheless, we are certainly able to process the answer and relate it to an implicit question containing alternatives of the form *That Laura liked x surprised Peter*. Concluding, we explicitly want to allow semantic representations of both questions that do have a linguistic realization, as well as questions that do not.

The category of possible questions is a rather broad category. Therefore, it is useful to distinguish different subtypes of questions. In the preceding sections, we already became

acquainted with so-called root questions and embedded questions. All examples of root questions presented above were EXPLICIT QUESTIONS. In most of the cases, explicit questions express an interrogative speech act and are addressed towards the interlocutor(s). However, there are some exceptions, such as rhetorical questions or examples such as (64) or (65).

- (64) Wer mag wohl Natalio?
who.NOM likes PRT Natalio
'Who could like Natalio, I wonder.'
- (65) Maria war verzweifelt. Was sollte sie nur tun?
Maria was desperate. What should she only do
'Maria was desperate. What could she possibly do?'

Zimmermann (2011a) argues that by using the particle *wohl* in an explicit question, "the speaker indicates her awareness that the addressee may not be fully committed to her answer" (Zimmermann, 2011a:2020). Hence, such questions result in a slightly different speech act than questions without *wohl*. They express an invitation to the addressee to provide her/his best guess of an answer. Given the uncertainty that is allowed for the answers to such questions, (64) is also a very suitable way to address a question to oneself or a question for which the speaker does not expect an answer at all, which is the reading of (64) I am interested in here.

The question in (65) is an example of free indirect discourse (see Eckardt, 2014), where the speaker of an utterance is switched from the narrator to a person in the narration (e.g. the protagonist) without using quotation. We are dealing with an explicitly asked question that is, however, not addressed towards the reader, but towards the person whose perspective the reader was led towards. In (65), the question is addressed towards the protagonist Maria herself.

We also saw that embedded questions do not invite the addressee to provide an answer, at least for non-inquisitive question embedding predicates. Furthermore, there is a different type of question that also does not request anything from the addressee. It is the so-called IMPLICIT or POTENTIAL QUESTIONS developed by van Kuppevelt (1995), Onea

(2016) and others. Those are questions that arise to the addressee in the course of the conversation without being explicitly uttered. Example (66) shows that a sentence involving an indefinite expression evokes an implicit/potential question.

- (66) Maria met a mysterious man yesterday.

Implicit/potential question: Who was that man?

This question is not uttered, and does not need to be addressed by either the speaker or the addressee. An implicit/potential question is rather something that is on the addressee's mind when processing, or it is on the speaker's/writer's mind when constructing a story. In section 4.6, potential questions will be discussed in detail. Furthermore, chapter 4 will be concerned with how (implicit) questions structure the discourse, in the form of the current question and the discourse question.

Finally, there are two more notions that I want to mention: SUPER-QUESTIONS and SUB-QUESTIONS. Roberts (2012) defines super- and sub-questions via entailment. Consider the following two questions.

- (67) a. Who likes Natalio?
b. Does Laura like Natalio?

The question in (67-a) is the super-question of the question in (67-b), and (67-b) is the sub-question of (67-a). According to Roberts (2012), a super-questions q-entails all its sub-questions. Entailment between questions (q-entailment), as opposed to entailment between propositions, is defined by Groenendijk and Stokhof (1984), as repeated in Definition 4.

Definition 4 (Question Entailment). *A question q_1 q-entails another question q_2 iff every proposition that gives a complete answer to q_1 also gives a complete answer to q_2 .*

(Groenendijk and Stokhof, 1984:16)

Accordingly, (67-a) entails (67-b). To illustrate this, assume the context, mentioned above, in which Laura and Simon are the only persons in the world (Natalio is a cat).

There are four propositions that constitute complete answers to the question in (67-a), listed in (68).¹⁰

- (68) a. Laura and Simon like Natalio.
 b. Only Laura likes Natalio.
 c. Only Simon likes Natalio.
 d. Nobody likes Natalio.

All of these answers also represent complete answers to the question in (67-b). Thus, (67-b) q-entails (67-a), and is identified as a sub-question of (67-a).

Moreover, answering several sub-questions of a question may jointly answer that question. If our context only contained the persons Laura, Dani, and Simon, then answering the sub-questions of (67-a), in (69), would automatically answer (67-a).

- (69) a. Does Laura like Natalio?
 b. Does Dani like Natalio?
 c. Does Simon like Natalio?

These relations between questions will play a role for modeling discourse, as we will see in chapter 4. Note that these relations are semantic relations and, thus, hold between the semantic objects that represent questions, not between speech acts. Furthermore, we can identify sub-questions of implicit questions, or questions that do not have a linguistic form, as discussed above.

2.7 Pragmatic Features of Questions

After having discussed the semantics of questions, I now point out some pragmatic features of questions that should be accounted for: The EXISTENCE INFERENCE of questions and EXHAUSTIVITY.

¹⁰The answers in (68) correspond to the alternatives proposed by Groenendijk and Stokhof (1984), but q-entailment can be implemented for the other approaches to questions as well.

Existence Inference Some *wh*-questions seem to have an EXISTENCE PRESUPPOSITION, as example (70-a) illustrates, which seems to presuppose (70-b).

- (70) a. What did you just steal from my pocket?
b. You just stole something from my pocket. (Krifka, 2011:1755)

Krifka (2011) states that if the addressee of (70-a) did not steal anything from the speaker's pocket, she/he could reply with the utterance "*nothing*", but she/he would still feel accused by the speaker. This impression can be argued to arise because the question presupposes that the addressee stole something. In general, it has frequently been claimed that *wh*-questions have existence presuppositions, but this assumption has also been questioned. Another argument in favor of questions having an existence presupposition is the unacceptability of the answer in (71).

- (71) A: Who likes Natalio?
B: #Somebody likes Natalio.

B's answer does not seem informative in this context even though it excludes the alternative of nobody liking Natalio. The reason could be that it is already presupposed that somebody liked Natalio. Hence, nobody liking Natalio is excluded anyway and the answer in (71) would not provide any new information. Onea and Zimmermann (2019) suggest a solution to this problem by analyzing the alleged existential presupposition as an existence impudicature. They argue that A could alternatively have asked the more direct question in (72) in order to get the information provided by B.

- (72) Does anyone like Natalio?

Since A chose to ask the *wh*-question instead, she might take the answer to (72) for granted. It is reasonable to assume that she takes this answer to be *yes*, given that she is asking (71). If the answer to (72) was known to be *no*, asking the *wh*-question in (71) would not promise any new insight, since the answer was already known, namely *nobody*. Thus, (71) gives rise to the impudicature that somebody likes Natalio.

There are also questions that do not seem to presuppose existence, such as (73-a).

- (73) a. Who can name all the capitals of Europe?
b. Somebody can name all capitals of Europe.

In this case, it is not presupposed that there is somebody who can name all the capitals of Europe. Hence, (73-a) does not presuppose (73-b). Another argument against the existential presupposition is the possibility of negative answers, as in (74).

- (74) Who likes Natalio? – Nobody.
↗ Somebody likes Natalio.

Since presuppositions are usually not cancelable in such a simple way, this example shows that we are not dealing with an existence presupposition. In order to question the presupposition, the addressee should use a marking such as *wait a minute*, as attested for cancellation of other presuppositions. However, this does not produce an acceptable reply for questions, as indicated in (75-c).

- (75) Who likes Natalio?
a. Nobody likes Natalio.
b. Actually, nobody likes Natalio.
c. ?Wait a minute, nobody likes Natalio.

This is another counterexample against questions having an existence presupposition. In contrast, the typical marking for canceling implicatures, such as *actually*, is acceptable, though not necessary either, as indicated in (75-a) and (75-b).

In order to capture both cases, those with and without an existence inference, Onea and Zimmermann (2019) suggest to treat this inference as a supposition, instead of a presupposition. A supposition is a weaker inference than a presupposition, which is not part of the truth-conditions but part of the speaker's expectations about future developments of the discourse. Hence, the speaker expects a different answer than *nobody* or *nothing*.

Onea and Zimmermann state that “when uttering a wh-interrogative, a speaker normally indicates that s/he assumes or expects there to be some individual(s) with the relevant property in question, and that s/he expects an affirmative response specifying these individual(s).” (Onea and Zimmermann, 2019:27) This way, the existence inference is not part of the semantics of questions anymore, and can be canceled in case of a negative answer or specific contexts, such as (73-a).

A formal analysis that captures the difference between the expected answer and the unexpected answer is presented in Roelofsen and van Gool (2010). They introduce an extension of inquisitive semantics that involves HIGHLIGHTING of alternatives. Highlighting is exemplified for *wh*-questions in (76). The highlighted alternatives are underlined.

- (76) [Who likes Natalio?] = {Laura likes Natalio, Nina likes Natalio, Simon likes Natalio, ... Nobody likes Natalio}

Roelofsen and van Gool (2010) argue that a *wh*-question, such as (76), highlights certain possibilities, while others are not highlighted. In particular, only the “negative” alternative in the q-alternatives of the *wh*-question is not highlighted. Their notion of highlighting can be interpreted to mean that the speaker expects one of the highlighted alternatives to be the answer. Formally, this is also reflected given that the union of the highlighted alternatives of the *wh*-question are equivalent to the existence presupposition. Hence, the expectation of an answer naturally leads to assuming existence. What is important here is that the q-alternatives do not exclude the “negative” alternative, it is just not highlighted and, thus, less expected.

Exhaustivity Another frequently discussed property of questions is EXHAUSTIVITY. As mentioned in section 2.4, questions are sometimes assumed to require exhaustive answers. An exhaustive question meaning looks like (77), following Onea and Zimmermann (2019:32).

$$(77) \quad [\![\text{Who likes Natalio?}(w)]\!] \\ = \lambda p. \exists x [p(w) \wedge p = (\lambda v. (x = \max(\lambda y. \text{like}(y, \text{Natalio}, v))))]$$

What is the maximal sum individual that contains all persons who like Natalio?

This representation contains a covert maximality operator that is part of the question meaning. This operator is derived from the maximality operator originally introduced by Link (1983) for plural definite descriptions like *the girls*. However, it is not that obvious whether questions actually do require exhaustive answers. It does not seem entirely unacceptable to provide a partial answer to a question like (77), naming just some of the people that like Natalio.

To approach exhaustivity in questions, it is useful to look at embedded questions, such as (78), because exhaustivity can be evaluated more clearly in those questions.

(78) Lea knows who likes Natalio.

\rightsquigarrow Lea knows about all the individuals that like Natalio that they like Natalio and she knows of all the individuals that do not like Natalio that they do not.

As the paraphrase indicates, this example can be assumed to mean that Lea knows the exhaustive list of individuals that like Natalio. Heim (1994), among others, supports such a strong exhaustive analysis of question embedding *know*. However, she treats exhaustivity here as a lexical feature of *know*, instead of as feature of the embedded question. Moreover, as mentioned earlier, there are also other readings of (78) that involve weaker exhaustivity, that are discussed in the literature. Example (78) can be uttered in a context where Lea knows of all the individuals that actually like Natalio that they like him, but she can be uncertain about some of the individuals that do not like Natalio (intermediate exhaustive reading), or even have false beliefs about some of the individuals that do not like Natalio (weak exhaustive reading). Those weaker readings not confirmed by all speakers, though (see Fricke et al., to appear).

Moreover, question embedding predicates can also embed mention-some questions, which do not even require weak exhaustivity, as in (79).

- (79) Lea knows where one can get good cake in Hanover.

This sentence is true if Lea knows one place where one can get good cake in Hanover. She does not need to know all the places. Thus, a mention-some question explicitly does not require an exhaustive answer, while other questions, such as the *wer alles* questions (lit. ‘*who all*’) in German (80), explicitly require an exhaustive answer.

- (80) Wer mag alles Natalio?
who likes all Natalio
lit. ‘Who all likes Natalio?’

This question does not allow a non-exhaustive answer, but explicitly requests to name all individuals that like Natalio (see Zimmermann (2007) for more details on *wer alles*).

In general, these observations support theories that allow for more than just exhaustive answers to a question, or just mutually exclusive q-alternatives. This, provides a counter-argument against Groenendijk and Stokhof (1984), whose question partition approach requires mutually exclusive alternatives.

Furthermore, there are many approaches showing for embedded questions that the exhaustivity observed in some readings can be implemented by optionally including an exhaustification operator, (see, e.g., Klinedinst and Rothschild, 2011; Uegaki, 2015). This makes it possible to assume a non-exhaustive representation of questions as the underlying representation, and at the same time to account for exhaustive readings of embedded questions. (Onea and Zimmermann, 2019:34) state for embedded questions that “exhaustification is conceived of as a way of integrating them into their linguistic environment” in such approaches. In chapter 3, exhaustivity will be discussed as a pragmatic effect of focus, instead of the question.

Chapter 3

Focus

In this chapter, I provide an overview over the topic of FOCUS. In section 3.1, I present some intuitive definitions of focus that have been proposed in the literature, and point out problems for these definitions. Section 3.2 deals with focus sensitivity which provides a good formal diagnostic for the identification of focus. In section 3.3, I introduce a formal definition of focus, based on alternative semantics following Rooth (1985) and Rooth (1992). This leads to an understanding of focus that is not defined by its means of prosodic realization but by its information-structural function. Eventually, section 3.4 defines focus as marking implicit questions, as put forward by van Kuppevelt (1991), Roberts (2012), and Krifka and Musan (2012), among others. In section 3.5, I discuss pragmatic effects of focus, and how they can be traced back to the function of focus, following Krifka and Musan (2012).

3.1 Introduction

The motivation for including an introduction to focus into this thesis is two-fold: 1. The discourse model that I will be using in my analysis of clefts is based on a certain interpretation of focus, which is based on questions. Therefore, I will motivate this understanding of focus here independently of clefts. 2. Clefts are often compared to canonical sentences with focus on the element that constituted the cleft pivot, as in (1).

- (1) a. It is Nina who is petting Mavi.
b. NINA_F is petting Mavi.

Capital letters mark the nuclear pitch accent in the sentence, while the subscript *F* marks the focus. This section provides the basics about focus that are necessary to approach this comparison, which will be discussed in detail in section 5.5 and section 8.5.

Focus, as in (1-b), has often been described as highlighting the most important part of a sentence, or as the part of a sentence that is new. Krifka and Musan (2012), however, argue that those descriptions are not useful definitions, given the vagueness of the concepts of newness and importance. A counter-example against focus representing new information is given in (2), taken from Schwarzschild (1999:145).

- (2) Who did John's mother praise? – She praised JOHN_F.

In example (2), the focus is commonly assumed to be the constituent *John*. However, the individual that *John* refers to is not new in the answer, since *John* has already been mentioned in the question. What is new is that *John's mother* praised him. Even though newness coincides with focus in many examples, examples like (2) show that it can still not be used to define focus.

The problem with the concept of importance is that it is very hard to define. Often, it is not really clear what is important in a sentence or what a sentence highlights to be important. Krifka and Musan (2012) exemplify this with a clefted example as in (3), which is commonly assumed to focus the clefted constituent, here again *John*.

- (3) It wasn't JOHN_F who stole the cookie. (Krifka and Musan, 2012:17)

Krifka and Musan suggest that what is important in this example is not *John* or the person that *John* refers to but the fact that somebody else stole the cookie, which is definitely not what the concept of focus is supposed to express.

Another way to illustrate focus is by a question-answer-pair, as in (4).

- (4) a. Who is Nina petting? – Nina is petting MAVI_F.
 b. Who is petting Mavi? – NINA_F is petting Mavi.

Example (4) indicates that focus depends on the question that is asked. Here, the focus represents the answer to the respective question. As we will see later, focus can also relate to implicit questions (Krifka and Musan, 2012). Before digging further into this approach, I want to discuss yet another approach that suggests itself in the light of (4), namely the one that focus is the most linguistically marked constituent of a sentence, which is realized by prosody in example (4). In this example, the answer to the question coincides with the prosodically most prominent element in the sentence, *Mavi* or *Nina*, respectively.

In the following, I take an excursion into how languages mark focus, and why there is not always a one-to-one mapping of the actual focus and observable focus realization. In German (but also English and many other languages), focus is realized by intonation. In particular, (some part of) the focused constituent receives the nuclear pitch accent, more specifically an A-accent (Bolinger, 1958). This accent is described as follows.

Accent A: A relative leveling off of the accentable syllable followed by a relatively abrupt drop, either within the accentable syllable (which is prolonged for the purpose) or in the immediately following syllable. In very rapid speech the drop may be postponed to the second following syllable, but rarely beyond this.

(Bolinger, 1958:142)

Example (5) provides an instance of narrow object focus and narrow subject focus, where the respective word that carries the main accent is marked with capital letters.¹¹

- (5) a. Nina streichelt MAVI_F. (narrow object focus)
 Nina pets MAVI.
‘Nina is petting MAVI.’
- b. NINA_F streichelt Mavi. (narrow subject focus)
 NINA pets Mavi.
‘NINA is petting Mavi.’

¹¹The capital letters are usually used to mark just the nuclear pitch accent of the sentence, not a specific accent such as the A-accent. Whenever I use this marking in this thesis, the A-accent marking is actually supposed to be indicated.

However, languages employ a lot of different means to mark focus. Hungarian, for instance, uses syntactic movement for marking the so called *identificational* focus (c.f. É. Kiss, 1998). Hungarian is said to have a designated syntactic focus position, into which the focused constituent is moved, exemplified in (6) and (19).

- (6) Mari fel hívta PÉTER.
 Mari.NOM up called PÉTER.ACC
 ‘Mari called PÉTER.’
- (7) Mari PÉTER hívta fel.
 Mari.NOM PÉTER.ACC call up
 ‘It is PÉTER Mari called.’ (É. Kiss, 1998:256)

In (6), we find the canonical word order for Hungarian, which is how information focus is expressed, following the analysis of É. Kiss (1998). In this case, the focused object *Pétert* is sentence-final. In (19), in contrast, *Pétert* has moved to the preverbal position. É. Kiss (1998) argues that it represents a case of identificational focus (see section 8.5 for details).

Some languages mark focus morphologically, such as Japanese or Cuzco Quechua. The latter marks evidential focus by the focus enclitic *-mi* when the topic marker *-qa* is present, as indicated in (8).

- (8) Law-ta-**mi** pay-**qa** mikhu-rqa-n
 soup.ACC.**mi** (s)he.**qa** eat.PAST.3
 ‘(S)he_{Top} ate soup_F.’ (Faller, 2002:150)

Syntactic constructions have also been proposed to realize focus in some languages. French, for instance, sometimes uses clefting for marking subject focus (Féry, 2001). Consider the French translation of the German (5-b) in (9).

- (9) C'est NINA qui caresse Mavi.
 it.is Nina who.NOM pets Mavi
 ‘NINA is petting Mavi.’

In French it is not possible to stress the subject, as it is done in German. This example indicates that clefts seem to have something to do with focus (see more in section 5.5

and section 8.5). In many examples, it is not just one focus marking that is used in a language. As already indicated by the accent marking in (19), Hungarian marks the focused constituent prosodically additionally to the syntactic movement. Likewise in the French example in (9), the clefted constituent also receives the nuclear pitch accent. I will argue later that even in German clefting is used for subject focus marking in written texts (Tönnis et al., 2018), which will be discussed in 8.5.

However, defining focus by how it is marked is not a useful approach, either. The cases that cannot be explained by that approach are those where focus PROJECTS. Example (10) illustrates several cases of focus projection. German behaves like English with respect to focus projection.

- (10) a. *What did Mary buy a book about?*
 Mary bought a book about [BATS]_F
- b. *What kind of book did Mary buy?*
 Mary bought a book [about BATS]_F
- c. *What did Mary buy?*
 Mary bought [a book about BATS]_F
- d. *What did Mary do?*
 Mary [bought a book about BATS]_F
- e. *What happened?*
 [Mary bought a book about BATS]_F

(Selkirk, 1996:554)

In order to distinguish the representation of intonation and focus in languages such as German and English, linguists mark the focused constituent with the subscript of *F*, while intonation, more specifically the nuclear pitch accent, is marked with capital letters. All the answers in (10) show exactly the same focus realization by assigning *bats* the nuclear pitch accent. However, the sentence with this intonation can answer quite different questions, which leads to different foci. The accent is said to percolate from *bats* to higher phrases containing *bats* since it is not allowed by (German) prosody to assign an accent to every word inside the focus, as illustrated in (11).

- (11) *What did Mary do?*

*Mary [BOUGHT A BOOK ABOUT BATS]_F

Hence, I conclude that the feature *intonation* cannot be used to define focus in German. This holds for all languages which either do not mark focus by intonation at all, or which do mark focus by intonation but also have focus projection. This most likely include all languages.

Returning to the question-answer approach, we can observe that (4-a) has the inference that Nina is neither petting Theo, nor Lilly, nor anybody else except for Mavi. Admittedly, it is not a very strong inference, but it is there. This observation motivates an approach promoted by Rooth (1985), that considers alternatives, such as *Nina is petting Theo* for (4-a), to define focus. Recall that also questions were defined in terms of alternatives (see chapter 2). Hence, the fact that focus depends on questions is yet another motivation to involve alternatives in the definition of focus. Based on this, Krifka and Musan (2012) propose a preliminary definition of focus, as in (12).

- (12) Focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions. (Krifka and Musan, 2012:7)

So far, all the focus related categories mentioned above were pragmatic ones. However, it has been observed that focus can have semantic effects as well. Those effects occur with focus-sensitive particles, such as *only*. Example (13) illustrates the truth-conditionally different readings of one sentence depending on whether the focus is on *Bill* (13-a) or on *Sue* (13-b). The meanings are added by the author.

- (13) a. Mary only introduced BILL_F to Sue.

Meaning: *Mary introduced nobody else to Sue except for Bill. She might have introduced Bill to other people, though.*

- b. Mary only introduced Bill to SUE_F.

Meaning: *Mary introduced Bill to nobody else than Sue. She might have introduced other people to Sue, though.* (Rooth, 1992:77)

It is not straightforward how the pragmatic approach to focus, based on questions and alternatives, relates to those semantics differences, based on focus-sensitive particles. On the one hand, the pragmatic approaches have problems explaining why semantic differences can arise due to focus. On the other hand, focus-sensitive particles seem to provide the best formal diagnostic so far for identifying focus. Precisely, those semantic effects make it possible to formally distinguish one focus from the other. Hence, the next subsection will elaborate on focus sensitivity.

3.2 Focus Sensitivity

This section will be concerned with FOCUS-SENSITIVE expressions, such as *only*, *even*, and *fortunately*, following an analysis of Beaver and Clark (2008). In case a sentence contains a focus-sensitive expression, truth-conditional differences arise depending on focus. The standard example is the one mentioned above in (13), containing the focus-sensitive particle *only*. Such examples are discussed by Beaver and Clark (2008), Büring and Hartmann (2001), Krifka and Musan (2012), and many others.

In a situation where Mary introduced Bill to Sue, Jane to Sue, and made no other introductions, (13-a) is false, while (13-b) is true. Hence, example (13) shows a semantic effect of intonation. Since example (13) does not show the same effect in the German translation, I provide a different example for German in (14), which is inspired by (Krifka and Musan, 2012:14), and contains the focus-sensitive adverbial *glücklicherweise* ('fortunately').

(14) Context: *Nina is supposed to bring all the documents.*

- a. Glücklicherweise hat Nina die WENIGER wichtigen Unterlagen
fortunately has Nina the LESS important documents
vergessen.
forgot
'Fortunately, Nina forgot the LESS important documents.'
- b. #Glücklicherweise hat Nina die weniger wichtigen Unterlagen
fortunately has Nina the less important documents
VERGESSEN.
FORGOT
'Fortunately, Nina FORGOT the less important documents.'

In the context in (14), where the speaker wants Nina to bring all the documents, the speaker is certainly not happy about Nina forgetting something, which makes (14-b) false in this context. However, (14-a) is true since the speaker just expresses that she/he is happy that Nina forgot the less important documents as opposed to more important documents. One might argue that (14-b) is perceived as “less false” than (13-b) in a context where Mary introduced Bill to Jane and Sue. The difference between focus-sensitive adverbials, such as *fortunately*, and focus-sensitive particles, such as *only*, lies in their different at-issueness status. The *yes, but...* test (see Onea and Beaver, 2009) explicates this difference, as (15) and (16) indicate.

(15) Context: *Nina is supposed to bring all the documents.*

Speaker A:

- a. Glücklicherweise hat Nina die weniger wichtigen Unterlagen
fortunately has Nina the less important documents
VERGESSEN.
FORGOT
‘Fortunately, Nina FORGOT the less important documents.’
- b. Speaker B: Yes, but that is not fortunate.
- c. Speaker B: #No, that is not fortunate.

(16) Context: *Mary introduced Bill and Jane to Sue.*

- a. Speaker A: Mary only introduced BILL to Sue.
- b. Speaker B: #Yes, but she also introduced Jane to Sue.
- c. Speaker B: No, she also introduced Jane to Sue.

This test distinguishes at-issue from not-at-issue content. When not-at-issue content is violated, it is more acceptable to address this using *yes, but....* When at-issue content is violated, a reply using *no,...* is most acceptable. For the focus-sensitive adverbial in (15), the test shows that the inference that it is fortunate that Nina forgot the less important documents is not-at-issue. Accordingly, the answer in (15-b) is acceptable, while the answer in (15-c) is not. For the focus-sensitive particle in (16), the test shows that the inference that nobody else was introduced to Sue by Mary is at-issue, as the answers in

(16-b) and (16-c) indicate. Hence, a semantic effect of intonation can be attested with focus-sensitive expressions for English and German, either on the at-issue or not-at-issue level.

However, it is actually focus that is relevant here, not intonation. This follows from the possibility of focus projection in combination with focus-sensitive expressions, as exemplified in (17).

- (17) a. Tonight, Mary only introduced Bill to [SUE]_F.

Meaning: *Mary introduced Bill to nobody else than Sue. She might have done other things tonight, though.*

- b. Tonight, Mary only [introduced Bill to SUE]_F.

Meaning: *The only thing Mary did tonight was introducing Bill to Sue.*

This example presents the same sentence with the same intonation but expressing two different foci. In (17-a), the accent and the focus coincide and we get a narrow focus reading. In (17-b), in contrast, the *F*-marking concerns a larger constituent than the one that receives the accent. The focus projects to the verb phrase. In a context in which Mary introduced Bill to Sue, did no other introductions, but also danced with Dave, (17-a) is true, while (17-b) is false.¹² Hence, we observe two truth-conditionally different readings for one and the same sentence, at least as it appears on the surface level. However, these truth-conditional differences are a good indication for a different underlying form. In other words, the semantic differences are considered to arise due to different foci.

We can use focus sensitivity as a test for identifying the focus. This provides a diagnostics that makes identification of focus unambiguous because one can test the semantic effect by evaluating the truth-conditions. Beaver and Clark (2008) assume that focus-sensitive expressions introduce an operator that associates with focus, producing semantic differ-

¹²This does not work the other way around since (17-b) entails (17-a). Whenever Mary did nothing else tonight than introducing Bill to Sue, she also introduced Bill to nobody else than Sue.

ences. Such an operator can only associate with the focused element. This leads to a preliminary Definition 5 of focus.

Definition 5 (Focus – preliminary I). *The focus is what focus-sensitive expressions associate with.*

Beaver and Clark (2008) further distinguish three categories of association with focus with different expressions falling into each category: QUASI ASSOCIATION, FREE ASSOCIATION, and CONVENTIONAL ASSOCIATION. Focus particles, such as *only*, *even*, and *also*, are represented as operators that associate conventionally with focus. Those expressions involve a lexically-encoded, conventionalized dependency on focus. It is hard-wired in the semantics and cannot be cancelled since its semantics depends on the focus, as we saw above. Next to those cases, Beaver and Clark (2008) describe cases of quasi association, which they refer to as a special kind of implicature. Expressions that fall into this category are negation, *either ... or*, possibility modals etc. I provide an example with negation in (18).

- (18) a. Kim doesn't study [LINGUISTICS]_F at Northwestern.
b. Kim doesn't study linguistics at [NORTHWESTERN]_F.

(Beaver and Clark, 2008:45)

In this example, there are no truth-conditional differences. However, the two versions lead to different inferences. Example (18-a) implies that Kim studies something else than linguistics at Northwestern, while this certainly does not hold for (18-b). That example, in contrast, implies that Kim studies linguistics somewhere else than at Northwestern. These inferences are, however, not part of the semantics, in contrast to inferences of conventionally associated expressions. Inferences of quasi-associated expressions are assumed to be implicatures, since they can be canceled, as illustrated in (19)

- (19) a. Kim doesn't study [LINGUISTICS]_F at Northwestern. Actually, she does not study anything at Northwestern.

- b. Kim doesn't study linguistics at [NORTHWESTERN]_F. Actually, she does not study linguistics anywhere.

Hence, for quasi associated expressions, focus just affects what kind of implicatures are evoked. The third category, free association, is concerned with expressions that quantify over or set up a comparison with an implicit domain. I will discuss this category using *always* as an example, as in (20).

- (20) a. Kim always serves Sandy [Johnnie WALKER]_F.
 b. Kim always serves [SANDY]_F Johnnie Walker.

(Beaver and Clark, 2008:53)

Beaver and Clark (2008) interpret *always* as a quantificational operator that quantifies over a domain of events, which is not made explicit in (20), but which is easily retrieved via pragmatic reasoning. Depending on the focus in (20), this domain is resolved to a different set of events. This is illustrated by the paraphrases of (20-a) and (20-b), respectively, in (21).

- (21) a. **In all the events of Kim serving something to Sandy**, Kim served Sandy [Johnnie WALKER]_F.
 b. **In all the events of Kim serving somebody Johnnie Walker**, Kim served [SANDY]_F Johnnie Walker.

The phrases in bold describe the implicit domain for the quantification of *always* that is retrieved via a pragmatic process. Again, the two paraphrases depend on the focus under *always*.

Note that all of these occurrences of association with focus needed the existence of an operator, the exclusive operator for *only*, the negation operator, and the quantifying operator for *always*. In cases when there is such an operator in a sentence, analyzing the truth-conditional, or pragmatic effects described above may be a good diagnostic for identifying focus. However, the test runs into problems when there is a focus without an

operator. Hence, focus cannot be defined based on an explicit focus-sensitive operator, as observed in the examples of this chapter. Therefore, the next section explores yet another definition of focus, based on alternatives.

3.3 Alternative Semantics

This section starts with how to formally model the association of focus-sensitive operators with focus. The question of what exactly the operator applies to brings us closer to what focus actually is. Beaver and Clark (2008) assume that there is always a current question at each stage of a conversation (the term *current question* will be discussed in more detail in section 4.4). For now, we take the current question to be the question that the speaker intends to address with her/his utterance. Focus-sensitive expressions are concerned with alternatives similar to the question alternatives of this current question. Based on this thought, Krifka and Musan (2012) formulate the following, still preliminary, definition for focus.

Definition 6 (Focus – preliminary II). *Focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions.*

(Krifka and Musan, 2012:7)

According to Krifka and Musan (2012), this does not mean that alternatives are not present without focus. Focus simply highlights the presence of particular alternatives, which will turn out to depend on the current question below. The operators that associate in one way or the other with focus are represented as applying to those alternatives in different ways.

I will now present the framework of *alternative semantics*, developed by Rooth (1985), in which focus is a semantic category which is concerned with alternatives to what is actually uttered. The element that is focused in the actual utterance is varied in the alternatives, while all the other elements remain the same. Rooth (1992) assumes that each sentence

has an ordinary semantic value, marked $\llbracket \] \circ$, and a focus semantic value, marked $\llbracket \]^f$.

Example (22) presents the two values for example (5-a) above.¹³

The subject focus in (23-a) leads to the same ordinary meaning as in (22-b), but it has a different focus semantic value, as (23-b) indicates.

The focus semantic values in (22-c) and (23-c) can be described as taking the ordinary meaning and substituting the focused phrase with all possible elements from the same domain. Here, it is D_e , the domain of entities, since *Nina* and *Mavi* refer to entities. This procedure results in a set of propositions that vary with respect to the focused constituent. In example (22), the focus semantic value could be the set $\{Nina \text{ is petting } Theo, Nina \text{ is petting } Laura, \dots\}$. The focus semantic value is also often referred to as the (set of) FOCAL ALTERNATIVES. Each substitution yields an alternative to what was actually uttered. According to Rooth's (1992) definition, the ordinary meaning is also itself an alternative. For (22), possible alternatives would be *Nina is petting Lilly* or *Nina is petting Nala* etc., while (23) would evoke the alternatives *Lena is petting Mavi* or *Laura is petting Mavi* etc., for instance. The alternatives do not always only vary with respect to entities, as in example (24).

¹³A similar approach is developed by Klein and von Stechow (1982), Krifka (1992), and others, who introduce STRUCTURED MEANINGS. “A structured meaning is a pair consisting of a background part and a focus part. The background is of a type that can be applied to the focus. If this application is carried out, we arrive at the ordinary semantic representation.” (Krifka, 1992:18) I will use the framework of alternative semantics in this thesis.

- (24) Nina STREICHELT_F Mavi.
 Nina IS.PETTING Mavi
 ‘Nina is PETTING Mavi.’

Here, the alternatives vary with respect to all the things Nina could do with respect to Mavi, which is formally explicated in (25).

- (25) $\llbracket(24)\rrbracket^f = \{P(\mathbf{Nina}, \mathbf{Mavi}) \mid P \in D_{e, \langle e, t \rangle}\}$, where $D_{e, \langle e, t \rangle}$ is the domain of predicates with two arguments.

The focal alternatives in (25) contain *Nina is following Mavi*, *Nina is tickling Mavi*, etc. Furthermore, the cases of projected focus are also covered in alternative semantics. Consider (22-a) in a more specific context, such as in (26).

- (26) Context: *We are wondering what Nina is doing at the moment.*
 Nina is [petting MAVI]_F.

In this example, the alternatives are not of the kind *Nina is petting x* but of the kind *Nina is X-ing*, where *X* is a predicate. The alternatives the speaker and addressee consider are concerned with what else Nina could be doing. Rooth (1992) would derive the focus semantic value of (26) as in (27).

- (27) $\llbracket(26)\rrbracket^f = \{P(\mathbf{Nina}) \mid P \in D_{e, t}\}$, where $D_{e, t}$ is the domain of predicates with one argument.

But how are those alternatives related to the focus-sensitive expressions? Beaver and Clark (2008) propose that focus-sensitive expressions comment on these alternatives. Recall the examples discussed above:

- (28) a. Mary only introduced Bill to SUE_F.
 b. Kim doesn’t study LINGUISTICS_F at Northwestern.
 c. Kim always serves SANDY_F Johnnie Walker.

The focus in example (28-a) evokes the alternatives of the form *Mary introduced Bill to x* and the focus-sensitive operator requires *Mary introduced Bill to Sue* to be the only true alternative. Example (28-b) evokes alternatives of the form *Kim doesn't study x at Northwestern* but also *Kim studies x at Northwestern* (Beaver and Clark, 2008). Negation pragmatically implies that one alternative of the latter form is true. Hence, it can be inferred that Kim studies something else than linguistics at Northwestern. For example (28-c), alternatives of the form *Kim serves x Johnnie Walker* are derived. The operator representing *always* quantifies over events of the form *Kim serves x Johnnie Walker* requiring that in all those events *x* includes Sandy. Hence, those events are excluded in which Kim serves Johnnie Walker to people but Sandy is not among them.

This seems to work well when focus is embedded under an operator. However, what happens in cases of focus that is not associated with a focus-sensitive expression? How can the alternatives be integrated for those cases? These questions are addressed in the next section.

3.4 Focus and Questions

If focus indicates the presence of alternatives, what is the reason for doing so if no operator applies to those alternatives? It is not obvious how the alternatives are integrated into the semantics of simple focus sentences, such as in (22-a) and (23-a). Following Paul (1880), von Stechow (1991), Beaver and Clark (2008), Rooth (1992), and Krifka and Musan (2012), among others, I assume that the focus alternatives are integrated via the question they relate to, more precisely via FOCUS CONGRUENCE. Recall that also questions were defined as sets of alternatives (see chapter 2). Given that questions are understood as sets of answers, it is reasonable to assume that focus in the answer corresponds to the other possible answers in some way. Depending on its focus, an utterance is FOCUS CONGRUENT to a specific question, as defined in Definition 7, based on Roberts (2012:31) and Rooth (1992).

Definition 7 (Focus congruence). *Move β is congruent to a question q iff its focal alternatives are a superset of the question alternatives determined by q .*

We will see in chapter 4 that the question q in Definition 7 will turn out to be the current question. For now, however, we leave aside how this question arises and will just assume that it is there. The impression that we get in (26) about what kind of alternatives are relevant is related to what kind of question the speaker wants to answer or address with her/his utterance. Example (10), repeated in (29), illustrates that different foci are required depending on which question we have in mind.

- (29) a. *What did Mary buy a book about?*
Mary bought a book about [BATS]_F
- b. *What kind of book did Mary buy?*
Mary bought a book [about BATS]_F
- c. *What did Mary buy?*
Mary bought [a book about BATS]_F
- d. *What did Mary do?*
Mary [bought a book about BATS]_F
- e. *What happened?*
[Mary bought a book about BATS]_F

(Selkirk, 1996:554)

Example (29-a) represents an instance of NARROW FOCUS, where the main accent coincides with the F -marking. Moreover, there are several other questions that can be addressed with the same sentence, such as the ones (29-b) to (29-e). Those vary with respect to F -marking, though, and represent WIDE FOCUS. Still, the answer in (29) cannot address just any question, as indicated by (30).

- (30) *Who bought a book about bats?*
#Mary bought a book about BATS.

The answer in (30) is odd because, intuitively, its intonation feels wrong. Theoretically speaking, the problem is that its focus is not focus congruent with the question. Whatever the F-marking is in (30), it must be a constituent that contains the accented word *bats*. This either leads to alternatives that do not vary with respect to the subject, as in (29-a), or it leads to alternatives that vary for the entire sentence including the subject, as in (29-e). Both those *F*-markings introduce different kinds of alternatives than the question in (30) does.

Let us now see how focus congruence applies. In (31) and (32), the question alternatives and focal alternatives, respectively, of a congruent question-answer-pair are presented.

- (31) a. Who bought a book about bats?
- b. **Set of q-alternatives:** $\{\llbracket x \text{ bought a book about } \text{bats} \rrbracket \mid \llbracket x \rrbracket \in \text{PERSON}\}$

- (32) a. MARY_F bought a book about bats.
- b. **Set of focal alternatives:** $\{\llbracket x \text{ bought a book about } \text{bats} \rrbracket \mid \llbracket x \rrbracket \in D_e\}$

Rooth (1992:85) argues that the set of question alternatives is not identical but smaller than the set of focal alternatives since the question alternatives are restricted by features of the *wh*-pronoun. In the focal alternatives in (32-b), *x* is required to refer to an entity. In (31-b), in contrast, *x* must be an entity that is also a person. This is a requirement of the interrogative pronoun *who*, which can only ask for persons. Given that the focal alternatives in (32-b) are a superset of the question alternatives in (31-b), (31-a) and (32-a) represent a congruent question-answer-pair according to Definition 7.

The the question-answer-pair in (30), however, is not focus congruent. The focal alternatives of the answer, as illustrated in (33), are not a superset of the question alternatives, which are the same as in (31-b).

- (33) a. Mary bought a book about BATS_F.
- b. **Set of focal alternatives:** $\{\llbracket \text{Mary bought a book about } x \rrbracket \mid \llbracket x \rrbracket \in D_e\}$

Here, I assumed a narrow F -marking. Almost all the other possible focus projections from *bats* to a higher constituent do not produce focus alternatives that are a superset of the question alternatives, either. The one exception is broad focus, as in (34).

- (34) a. Who bought a book about bats?
b. [Mary bought a book about BATS] $_F$.

Definition 7 (focus congruence) makes wrong predictions for this kind of examples, a problem that has been noticed before (for instance by Krifka, 2011; Onea, 2013). The definition would predict the question-answer-pair in (34) to be focus congruent, which it is obviously not. However, there is nothing that prevents focus from projecting from *bats* to the entire sentence. Hence, the focus alternatives for (34-b) would be the set of all propositions, which would certainly be a superset of the question alternatives of (34-a), as represented in (31-b). A less complex example, giving rise to the same problem, is presented in (35).

- (35) a. What did Mary buy a book about?
b. Mary bought [a book about BATS] $_F$.

This question-answer-pair is also incorrectly predicted to be focus congruent. The set of focus alternatives of (35-b) of the form *Mary bought x* is certainly a superset of the set of q-alternatives of (35-a) of the form *Mary bought a book about x*. Hence, the requirements for focus congruence are satisfied. In order to avoid such incorrect predictions, one could assume some kind of pragmatic optimization requirement, as formulated informally in (36).

- (36) Focus congruence should be established by the smallest possible F-marking.

Schwarzschild (1999) proposed something similar. This would lead to narrow focus on *bats* in (35-b) and on *Mary* in (34-b). I will not go into detail of solving this problem in this thesis. However, since focus congruence will be an important component of my

discourse model in section 10.3, it is necessary that this problem is solvable (see Onea (2013:107–115) for a formal solution to this problem).

In order to reach the final definition of focus, one more aspect needs to be added: the integration of implicit questions. Especially in written texts, most of the occurrences of focus do not include an explicit question that focus can associate with. I follow Krifka and Musan (2012), Kamali and Krifka (2020), and Simons et al. (2017), among others, who analyze focus as a device to mark (covert) questions on the basis of the context. This leads to the final definition of focus, which uses focus congruence.

Definition 8 (Focus - final version). *Focus in a discourse move α indicates the presence of a (covert) question q such that α is focus congruent to this q .*

In order to understand this definition, consider the examples from Krifka and Musan (2012), repeated in (37), which illustrate how implicit questions may arise.

- (37) 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.

(Krifka and Musan, 2012:11)

The first sentence in example (37-a) induces the question *What happened?*, which is picked up by the second sentence by a broad focus. Hence, focus congruence is given. The first sentence in (37-b) evokes the question *What did she do there?*, and again the second sentence is focus congruent to that question.

Obviously, the first sentences in (37) evoke other questions, too. The kind of questions that can be evoked by an utterance are discussed in more detail in chapter 4. What is important here is the conceptualization of focus marking in the second sentences in (37) as helping the addressee to reconstruct the question that is addressed. The addressee, in a way, reconstructs the question-answer-pair since there is no explicit question uttered. In other words, the pragmatic effect of focus is to mark the question without having to make it explicit. Of course, this does not mean that the question is necessarily implicit. The approach holds equally for question-answer-pairs with explicit questions.

Recalling example (26), repeated in (38), it is now apparent that the context induces the question *What is Nina doing at the moment?*, here even as an embedded question.

- (38) Context: *We are wondering what Nina is doing at the moment.*

Nina is [petting MAVI]_F.

This question leads to a wider focus than narrow focus on *Mavi*. To explicate the proposed function of focus even more, consider the following example in (39).

- (39) *Mavi mag sehr gerne kuscheln! Und gerade wirkt sie ganz fröhlich. Ich kann es nicht genau erkennen, aber ich glaube jemand streichelt sie. Übrigens, was macht denn Nina gerade?*

'Mavi likes cuddling a lot! And she seems quite happy at the moment. I cannot make out exactly what's happening, but I think someone is petting her. By the way, what is Nina doing right now?'

- a. DIE_F streichelt Mavi.
DEF.PRO.FEM is.petting Mavi.
'SHE is petting Mavi.'
- b. Die [streichelt MAVI]_F.
DEF.PRO.FEM is.petting MAVI.
'She is petting MAVI.'

This context evokes two questions, one explicitly (q_1 : *What is Nina doing?*), and one by an indefinite description *someone* (q_2 : *Who is petting Mavi?*). In (39), we observe two possible replies that only differ with respect to focus. Moreover, they address different questions. The reply in (39-a) addresses q_2 , while the reply in (39-b) addresses q_1 . This example illustrates how focus helps the addressee to reconstruct the question-answer-pair that the speaker had in mind in case there are two possible questions to address.

Concluding, this definition of focus integrates the alternatives evoked by focus via focus congruence to a question, that may be implicit. In section 4.2, a formal definition of focus based on marking (covert) questions is provided, which is set in the framework of

Commitment Space Semantics. This understanding of focus naturally leads to its function in discourse, which will be the topic of the next section.

Finally, we need to check whether this is a definition that just applies to non-associated focus (without an operator), or whether it covers focus in the scope of a focus-sensitive operator as well. Krifka and Musan (2012) argue for the latter. Also Beaver and Clark (2008) propose that focus in the scope of focus-sensitive operators still indicates the presence of the same question in the context as it would without the focus-sensitive operator. Hence, focus remains a pragmatic phenomenon that requires focus congruence between a question and its answer. The only thing that focus-sensitive operators do is that they impose additional requirements on how the focal alternatives relate to the *q*-alternatives, as described above. Beaver and Clark (2008) state that they comment on the way in which the answer relates to the question marked by focus.

However, indicating the presence of a question might not be the only pragmatic function of focus. Krifka (2008), among others, mentions that it is also used to mark contrast, parallelism, or exhaustivity. Those phenomena are discussed in the next section.

3.5 Pragmatic Effects

In this last section, I present some specific uses and features of focus, which partly cause problems for the presented definition of it. Krifka and Musan (2012) mention different kinds of foci that have different pragmatic effects. It is still a point of debate how these should be analyzed. Here, I am concerned with the question whether the final definition of focus (Definition 8 above) can account for those uses of focus as well, and which problems occur.

Focus can, for instance, be used to HIGHLIGHT PARALLELS, as in (40).

- (40) An AMERICAN_F farmer was talking to a CANADIAN_F farmer.

(Rooth, 1992:80)

In this example, the two focused elements contrast with each other, and can be said to both receive contrastive focus. However, it does not seem very plausible that (40) is addressing the question *Which farmer was talking to which farmer?*. It is much more likely to address a broad question, such as *What happened?*. One way to deal with such examples would be to assume that we are not dealing with focus accents here. One of them or both could be contrastive topics, for instance (c.f. Büring, 2003). Given that the two foci symmetrically contrast, one could also assume that there are two focus congruent questions, raised sentence-internally: The focus on *American* is focus congruent to the question *Which farmer talked to a Canadian farmer?*, while the focus on *Canadian* is focus congruent to the question *Which farmer did an American farmer talk to?*. I will not solve this issue in this thesis (see, for instance, Rooth (2009) for a solution).

This example can also be seen as a special case of CONTRAST. A more general example is given in (41).

- (41) Speaker A: What did Peter buy?
Speaker B: Peter bought a KANGAROO_{CF}!

Speaker B's reply is not only answering the question in the context, it also expresses that speaker B does not believe that speaker A will consider it likely that Peter bought a kangaroo and something else (c.f. Zimmermann, 2011b:1167). Taking a closer look at this description of the observed effect, we see that the proposition that is considered unlikely depends on the question present in the context. This becomes more apparent when the question is not explicit, as in (42).

- (42) Peter left the party. Then, he WALKED_{CF} home.

The speaker implies that it is not likely that Peter walked home, because he usually takes a taxi. In order to derive this inference, the addressee needs to consider the question *How did Peter get home?*, and determine whether walking is an expected alternative compared to the other alternative in the q-alternatives of this question. Hence, contrast in general seems to be in line with the definition of focus above.

Furthermore, focus marking can be used to CORRECT or CONFIRM information, as in (43) (following Krifka and Musan, 2012:11).

- (43) Speaker A: Nina is petting Mavi.
- Speaker B: (No), LENAF is petting Mavi.
 - Speaker B: Yes, NINA_F is petting Mavi.

Krifka and Musan (2012) argue that correction and confirmation require that the focal alternatives contain a proposition that was uttered in the immediate context, such as provided by speaker A. I assume that correction and confirmation (or reassertion) are performed before this proposition is actually accepted. Following Farkas and Bruce (2010), I assume that assertions are ‘parked’ on a table of negotiation for some time. As long as they are on the table, the questions they address are not finally settled and can be considered open. Hence, correction and confirmation can be treated in exactly the same way as focus before, only that the questions they indicate must be on the table. Confirmation differs from correction, though, as far as the explicitness of the question is concerned, as indicated in (44) and (45).

- (44) a. Speaker A: NINA is petting Mavi.
b. Speaker B: Yes, NINA is petting Mavi. / (No), LENAF is petting Mavi.
- (45) a. Speaker A: Nina is petting MAVI.
b. Speaker B: ?Yes, NINA is petting Mavi. / (No), LENAF is petting Mavi.

It seems that both confirmation and correction are appropriate when the proposition they address indicates the same question as the confirming or correcting statement, as in (44). If the foci diverge comparing the proposition and the confirmation/correction statement, as in (45), only correction is appropriate. The confirmation in (45-b) is marginally appropriate. It only seems to be interpretable if it does not only confirm but also expresses a contrast between Nina and other individuals in the context. This contrast inference does not arise in (44), and it does not arise for correction in (45-b). Hence, it seems that

confirmation needs an explicit question while the question addressed by correction may be inferred in the context (but on the table).

Moreover, Krifka and Musan (2012), and many others, argue that focus has an EXHAUS-TIVITY inference, as in (46).

- (46) It is JOHN_F who stole the cookie!

The cleft sentence is argued to have the exhaustivity inference that nobody other than John stole the cookie. However, the existence of this inference is also questioned. Furthermore, it could be a feature of the cleft structure, instead of being evoked by exhaustive focus. Velleman et al. (2012), for instance, argue that the cleft structure introduces a focus-sensitive exhaustivity operator as part of its semantics just like *only* does, with the difference that *only* leads to at-issue exhaustivity and clefts to not-at-issue exhaustivity. This issue will be discussed in more detail in chapter 7. Plain focus, in contrast, is not assumed to introduce such an operator.

In general, a focus marked by intonation, instead of clefting, is said to just have an exhaustivity implicature. For the example *Nina is petting MAVI_F*, (47) illustrates how this implicature can be canceled.

- (47) Nina is petting MAVI_F, and actually she is also petting Theo.

The implicature is derived by Gricean reasoning as follows: The speaker is assumed to obey the maxim of quantity. She could have said that *Nina is petting Mavi, Lena, and Theo etc.* Since she did not do that, she must have had no evidence for Nina petting Lena, or anybody else. Accordingly, Nina is only petting Mavi. This reasoning accesses the alternative of the focus congruent question of *Nina is petting MAVI_F*, and negates all stronger alternatives. Benz and Jasinskaja (2017) show that this kind of implicature and also other implicatures can be analyzed as depending on the question under discussion and, accordingly, on the focus (see also van Kuppevelt, 1996). Hence, this pragmatic effect is in line with the approach to focus presented here.

Chapter 4

Discourse Model

For the analysis of German *es*-clefts in discourse, one needs a solid basis for how to model discourse in general. This chapter provides this foundation and presents a new perspective on the current question, resulting in revised principles of *relevance* and *accepting a discourse move*.

Section 4.1 presents an overview of what discourse is and what should be covered by a discourse model. Furthermore, I will explain the motivation for choosing a discourse model based on questions instead of discourse relations. One important concept that is necessary for modeling discourse is the *context update*, which is discussed in section 4.2 in the framework of *Commitment Space Semantics* by Krifka (2015) and Kamali and Krifka (2020). The basics of the framework I use for modeling discourse is introduced in section 4.3, namely the *Question Under Discussion* framework based on Roberts (1996) and Roberts (2012). Section 4.4 and 4.5 extend this framework by distinguishing *current questions* from *discourse questions*, and by including both the speaker's and the addressee's perspective. Finally, section 4.6 will introduce *potential questions*, following Onea (2016), as kinds of questions that arise in discourse and that are of particular relevance to the analysis of *es*-clefts in German.

4.1 Introduction

A discourse model is concerned with predicting which information exchange is well-formed and which is not. With respect to texts, this amounts to predicting whether a sequence of sentences represents a coherent text or not. The model should, for instance, be able to explain the difference in well-formedness between (1-a) and (1-b).

- (1) a. My friend Laura lives in Japan. She grew up in Colombia, though.
b. My friend Laura lives in Japan. #It is raining in Colombia(, though).

The second sentence in (1-a) seems to be a natural continuation of the first sentence. The second sentence in (1-b), in contrast, seems odd, independently of the presence of the discourse connective *though*. The second sentence simply has no relation to the first sentence. This is something we usually do not find in natural discourses. If this occurs, extra marking is necessary (Onea and Zimmermann, 2019; Krifka and Musan, 2012), such as in (2).

- (2) a. My friend Laura lives in Japan. In Colombia, it is raining, though.
b. My friend Laura lives in Japan. And, irrespective of that, it is raining in Colombia.

Example (2-a) shows a non-default word order, which is referred to as *topicalization*. The movement of the constituent *in Colombia* directs the addressee's attention the discourse referent that *Colombia* refers to (c.f. Roberts, 2011). Furthermore, coherence is also gained in (2-a) because *Colombia* is a contrastive topic, which contrasts with *Japan* in the previous sentence. Example (2-b) explicitly states the non-relatedness of the two sentences by the phrase *irrespective of that*. This extra marking allows continuations that would result in a rather incoherent text without the marking.

The aim of this chapter is to explicate what kind of relation two or more sentences should have in order to result in a coherent text or a well-formed discourse. I follow Stalnaker (1978), Roberts (2012), and many others who analyze discourse as an inquiry of our

world. This inquiry is jointly conducted by the interlocutors who share the goal of finding out “the way things are” (Roberts, 2012:4). For this purpose the discourse participants share information with each other. Following Wittgenstein, Carlson (1982) describes this process as DIALOGUE GAMES, that consist of “cooperative activities of information exchange” (Carlson, 1982:18). The goal of this game is to publically agree on information about our world, and it includes discourse rules that the discourse participants adhere to, for instance *Do not utter something completely irrelevant*. The discourse participants make moves according to these rules, e.g. they make an assertion or ask a question. These moves might involve more or less complex strategies that lead to sub-inquiries. Sub-inquiries evolve because it is impossible to complete the inquiry of “the way things are”. It would be an endless conversation. Hence, one might want to start with a sub-inquiry, for instance the sub-inquiry of “the way the weather is today”.

Carlson (1982) describes a DISCOURSE MOVES as follows, excluding imperatives:

[Discourse] Moves in a dialogue game will consist of complete sentences put forward by dialogue participants and addressed to other participants of the dialogue. [...] the only admissible moves consist of declarative and interrogative sentences. (Carlson, 1982:17)

Carlson (1982) states that the game can be extended to also include imperatives. However, imperatives do not play a role for cleft sentences, and will, thus, be ignored in this thesis. Carlson further assumes that the participants of the dialogue game pursue strategies to win the cooperative game, i.e. to share as much information as possible. Following Grice (1975), he assumes that these strategies evolve from rational reasoning of the participants based on the Gricean maxims, in particular the maxim of relevance. For discourse, Carlson describes RELEVANCE as follows:

The crucial consideration in the well-formedness of a dialogue will be whether its individual moves have bearing on the topic or subject matter of the dialogue, or whether they are irrelevant, idle, or beside the point.

(Carlson, 1982:91)

What Carlson calls “topic or subject matter” is conceptualized by Roberts (2012) as a question that the discourse participants want to answer. She presents a discourse model that is based on questions, though not necessarily explicitly uttered questions. She calls such questions the QUESTIONS UNDER DISCUSSION (QUDs). She assumes that at every point of a discourse it should be clear to the discourse participants what issue they are discussing, e.g. the weather. She conceptualizes and formalizes this issue as an underlying question that the discourse participants are trying to answer together. The issue about the weather translates to the question *What is the weather like?*, for instance. The QUD framework will be presented in section 4.3.

There are other approaches to modeling discourse next to the framework I am focusing on in this thesis. Asher and Lascarides (2003), for instance, present *Segmented Discourse Relation Theory* (SDRT), which analyzes rhetorical relations between discourse segments. Discourse segments consist of one sentence or a set of adjacent sentences in a discourse that represent a rhetorical unit, such as *narration*, *elaboration*, *explanation* etc. Consider the famous discourse example by Asher and Lascarides (2003), repeated in (3).

- (3) a. John had a great evening.
b. He had a great meal.
c. He ate salmon.
d. He devoured cheese.
e. He then won a dance competition. (Asher and Lascarides, 2003:80)

Consider just the sentences (3-b)–(3-d), which I will use to exemplarily show how discourse relations and discourse segments are determined. The whole discourse is illustrated in Figure 4.1 below. The sentences in (3-c) and (3-d) form a discourse segment expressing the rhetorical relation of *Elaboration* with respect to the sentence in (3-b). In other words these two sentences provide more details on the great meal that John had. The discourse segments of a coherent text or monologue are hierarchically organized, as illustrated in Figure 4.1 for example (3).

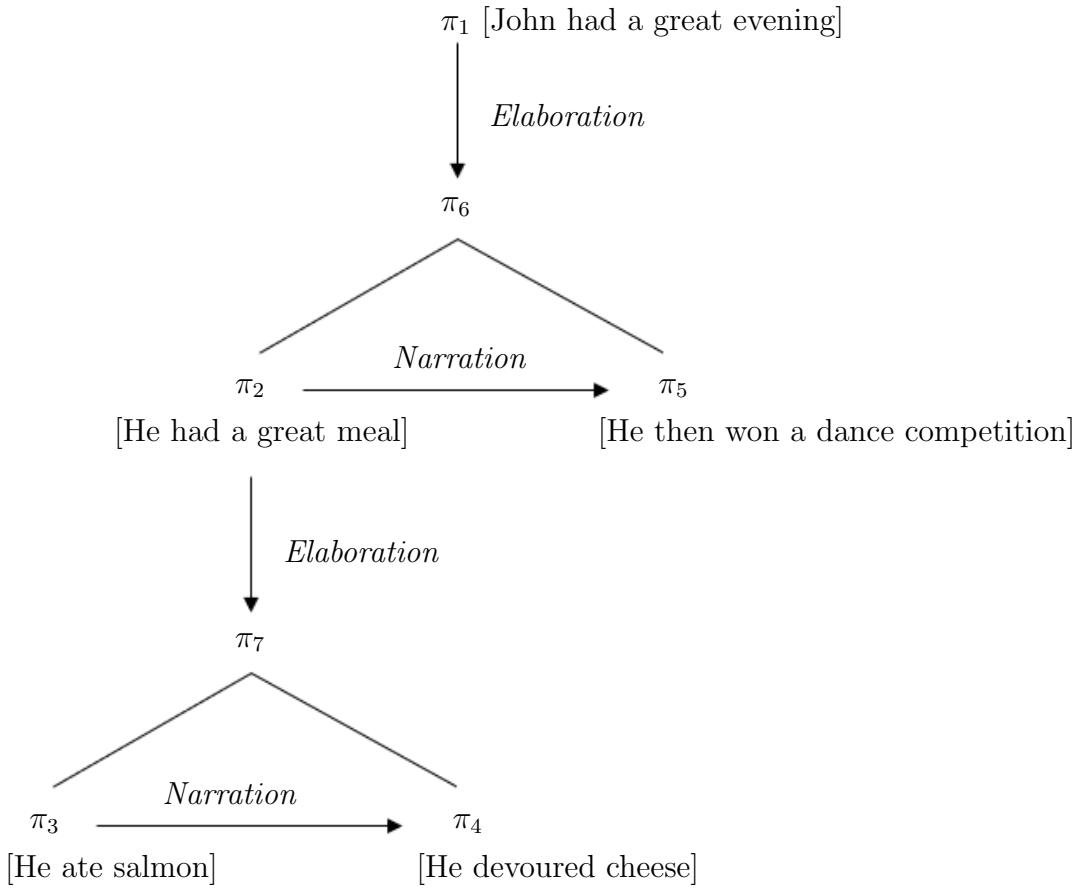


Figure 4.1: SDRT graph for the discourse in (3) taken from Asher and Lascarides (2003:147)

With this structure and the RIGHT FRONTIER CONSTRAINT, first mentioned by (Polanyi, 1985), this model is capable of making predictions about the well-formedness of a discourse. Based on the right frontier constraint, Asher and Lascarides (2003) define available attachment sites for additional discourse segments as follows: Discourse segments may only be attached to the segment at the right edge of a SDRT graph, which would be π_5 in Figure 4.1, or those elements that dominate it, hence π_6 and π_1 in Figure 4.1. Accordingly, the continuation in (4-f) is predicted to be well-formed, given that it attaches to π_5 , while the continuation in (5-f) is not well-formed since it violates the right frontier constraint by attaching to π_4 .

- (4) a. John had a great evening.
 b. He had a great meal.
 c. He ate salmon.
 d. He devoured cheese.
 e. He then won a dance competition.
 f. **Then, he celebrated with his competitors.**
- (5) a. John had a great evening.
 b. He had a great meal.
 c. He ate salmon.
 d. He devoured cheese.
 e. He then won a dance competition.
 f. **#He had ice cream for desert.**

However, there are violations of the right frontier constraint that nevertheless constitute well-formed discourses. Especially clefts seem to frequently break the constraint, as the example from chapter 1 repeated in (6) indicates.

- (6) Lena hat gestern auf der Party mit einem Typen gesprochen. Die beiden haben viel gelacht und sich direkt für den nächsten Abend verabredet. Dann ist Lena glücklich nach Hause gefahren.
'Yesterday at the party, Lena talked to some guy. The two of them laughed a lot and they agreed to meet again the next evening. Then, Lena went home happily.'
- a. Es war Peter, mit dem sie gesprochen hat.
 it was Peter with whom she talked has
'It was Peter she talked to.'
- b. ?Sie hat mit Peter gesprochen.
 She has with Peter talked
'She talked to Peter.'

The cleft sentence in (6-a) seems to attach to the preceding discourse without problems by elaborating on the first sentence. However, it does then not attach at the right edge, as the SDRT graph of the context of (6) illustrates in Figure 4.2.

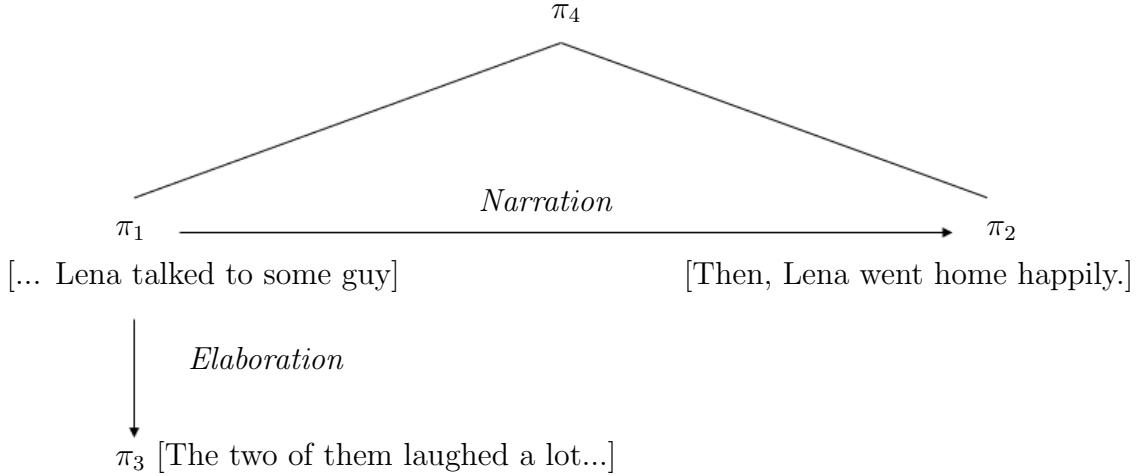


Figure 4.2: SDRT graph for the context of the discourse in (6)

It seems that the cleft in (6-a) must attach to π_1 or maybe π_3 , in Figure 4.2, in order to express the rhetorical relation of providing more details about the guy mentioned in the first sentence. However, neither π_1 nor π_3 are at the right edge or a segment dominating the right edge. Hence, the cleft sentence breaks the right frontier constraint without leading to an incoherent text. The framework of SDRT cannot account for the discourse behavior of the cleft in this example. The unclefted equivalent in (6-b), in contrast, is indeed infelicitous, and behaves as predicted by the right frontier constraint.

Furthermore, there are examples, such as (7), that show not entirely incoherent texts but very unexpected continuations of a text. Such examples cannot be accounted for in the SDRT framework.

- (7) Lena talked to Peter. ?She uttered 2.79 words per second.

The continuation in (7) attaches at the right edge. Thus, the example cannot be ruled out because of a right frontier violation. Such problematic examples are much better captured by approaches that are question-based and relevance-based. Addressing the question *How*

many words per second did Lena utter? does not seem relevant in the context in (7). In section 10.2, this example and also (6) will be accounted for based on the concept of expectedness of addressing a question.

Hence, a question-based discourse model serves the purpose better to explain the behavior of *es*-clefts in German. Anyways, I assume that it is possible to translate discourse relations and the predictions that can be made on their basis into a question-based framework. I will not spell this out in this thesis but refer to Onea (2019b).

Another aspect that needs to be modeled is what it means when discourse participants are approaching their shared goal of finding out how the world is. Stalnaker (1978) proposed that the discourse participants are incrementally adding information to the shared COMMON GROUND, the set of mutually shared information. In the ideal case, the common ground would eventually contain all the information necessary to know how things are in our world. The more information is agreed on by all discourse participants, the more the common ground grows. And for each new accepted information, the common ground is updated. Section 4.2 will elaborate on this issue.

Finally, Onea (2013) and van Kuppevelt (1995), among others, argue that not only the QUD is relevant for making predictions about discourse, but also a certain kind of implicit questions. Onea (2013) calls them *potential questions*, as already mentioned in chapter 2. Those are questions that are evoked by an utterance without being part of a strategy. Section 4.6 will be concerned with such questions, and section 10.3 will incorporate them into an adapted discourse model.

4.2 Context Update

In this section, I introduce the notions of CONTEXT UPDATE, COMMON GROUND and COMMITMENT SPACES. As mentioned above discourse involves a common goal. This goal can be formulated as increasing the common ground or decreasing the context set. For

every new information that is gathered by the discourse participants the common ground or the context set is updated.

Generally, one can assume that, at every point of a conversation, there is some information which (i) all interlocutors share and which (ii) each interlocutor believes that all interlocutors share. The second part means that it is not information that the interlocutors share coincidentally. This shared information is referred to as COMMON GROUND (CG).

All speech acts to be performed in the conversation are based on the common ground. For instance, one cannot perform the speech act of an assertion that presupposes something that is not in the common ground, not even if it is coincidentally known to everybody. In the simplest case, a piece of information enters the common ground when it is asserted in a conversation and all interlocutors agreed that it is true. It is, however, possible that a piece of information is in the common ground that has not explicitly been uttered but can still be assumed by all interlocutors to be true, such as the fact that a cat is an animal.

According to Stalnaker (2002), Grice (1989) was the first to use the term *common ground* even though he did not define it. Stalnaker (2002:704) describes the common ground as “*the mutually recognized shared information in a situation in which an act of trying to communicate takes place*”. He defines common ground in terms of acceptance, in the sense that it contains all propositions that are accepted to be true by all participants of the conversation. The pieces of shared information, mentioned above, are modeled as accepted propositions. According to Stalnaker, the simplest and most immediate way to accept a proposition is to believe it to be true. I will make the simplified assumption that propositions are always accepted in this way. Hence, I follow Stalnaker’s preliminary proposal that the common ground is equivalent to the common beliefs shared by all the participants of the conversation, except for the coincidentally identical beliefs. Stalnaker discusses some cases where the two notions diverge¹⁴, which are, however, orthogonal to

¹⁴Stalnaker (2002:716) describes these cases as follows: “One may simplify or idealize in an inquiry, one may presume innocence to ensure fairness, one may make assumptions for the purpose of contingency planning, one may grant something for the purpose of an argument.” We see that those are very specific situations that do not have to concern us here.

the interpretation of *es*-clefts in German. Definition 9 defines common ground as I will use it, adapted from Stalnaker (2002:716).

Definition 9 (Common Ground). *The common ground \mathcal{CG} is a set of propositions, such that for each proposition $\varphi \in \mathcal{CG}$, all participants of a conversation believe that φ , and all believe that all believe that φ , and all believe that all believe that all believe that φ , etc.*

Depending on which representation of propositions is assumed, the common ground looks slightly different. If one assumed that propositions are sets of worlds, namely those in which the proposition is true, the common ground would be a set of sets of worlds. Each set of worlds in CG represents one proposition.¹⁵

A related notion is the CONTEXT SET, which describes the set of possible worlds compatible with the propositions in the common ground (Beaver and Clark, 2008; Roberts, 2012). When a proposition φ is defined as the set of worlds where φ is true, then the context set can be defined as in Definition 10 (c.f. Roberts, 2012).

Definition 10 (Context Set). *For a common ground \mathcal{CG} , the context set S_{con} is such that $S_{con} = \bigcap \mathcal{CG}$, i.e. the set of possible worlds where all the propositions in the common ground are true.*

The context set includes only those worlds that are in each set of the CG, i.e. in each proposition in CG, yielding the worlds in which all so far accepted propositions are true. The two notions, COMMON GROUND and CONTEXT SET, illustrate two sides of a conversation. On the one hand, the aim of a conversation is to gather and share information. This means that the common ground grows during the conversation, when new propositions are added. On the other hand, the aim of a conversation is to find out which world out of all possible worlds is our actual world. Since there are always questions that we do not yet have answers to, there are always different possible worlds that could be ours depending on the answers to these questions. For each answer in a conversation, i.e. whenever a proposition is added to the common ground¹⁶, the context set shrinks. Those worlds that

¹⁵The definition of CG does, however, not hinge on a particular representation of propositions.

¹⁶I assume that only informative propositions are added to the common ground

are incompatible with the new proposition are eliminated from the context set. In other words, interlocutors want to gather information in the common ground and they want to narrow down possible worlds in the context set to approach the actual world. Actually, these are not two separate goals, as it might have seemed above. The common ground and the context set are rather two perspectives on the same goal of finding out how things are, as Roberts (2012) phrases it.

I will now introduce the notion of a CONTEXT UPDATE, which has already played a role in the description above. Every speech act that is performed in a conversation changes the common ground and the context set. For an assertion, a proposition φ is added to the common ground, and the context set is reduced by intersection with φ . This is called an UPDATE of the common ground or the context and will be formally defined in the following.

I will define context updates in the framework of *Commitment Space Semantics*, introduced by Cohen and Krifka (2014), and refined by Krifka (2015) and Kamali and Krifka (2020). This framework is useful for my purposes since it defines updates with respect to speech acts not just with respect to semantic objects, such as questions or propositions. Cohen and Krifka describe the starting point for their approach, as follows.

We understand speech acts as changing commitments of the interlocutors. For example, in asserting a proposition Φ , the speaker takes on a commitment to be responsible for the truth of Φ , and in promising to behave in a way described by a proposition Φ , the speaker takes on the commitment to behave in that way. (Cohen and Krifka, 2014:48)

Before getting to the definition of context updates, I present some relevant concepts developed in *Commitment Space Semantics*. First of all, Cohen and Krifka (2014) introduce the basic notion of COMMITMENT STATES, which they describe as “the set of public commitments that the interlocutors have accumulated up to the current point in conversation.” (Cohen and Krifka, 2014:48) Furthermore, they require that commitment states should not contain contradicting commitments. This understanding of commitment states cor-

responds to the notion of common ground in other models. In this thesis, I will use a less restricted definition of commitment states, as in Definition 11.

Definition 11 (Commitment State). *A commitment state is a set of non-contradicting propositions.*

Cohen and Krifka (2014) describe commitment with respect to a proposition. Following this approach, Definition 11 reduces commitment states to the set of those proposition that the interlocutors commit to. According to this definition, however, not every commitment state has to correspond to the common ground. A commitment state can be any set of propositions as long as it is consistent. It could, for example, represent the propositions that only one of the discourse participants commit to. For now, the common ground is described as the *current commitment state*. Below, we will get back to what actually corresponds to the common ground in this system.

Krifka (2015), further, suggests that every new speech act performed in the discourse updates the current commitment state. An update of a commitment state c with the speech act of an assertion is given in Definition 12.

Definition 12 (Assertive Update of a Commitment State). *For a commitment state c and a proposition φ , $+A_\varphi$ is the assertive update with the speech act of asserting φ , such that $c + A_\varphi$ is a commitment state and $c + A_\varphi = c \cup \{\varphi\}$.*

(Krifka, 2015:329)

This definition describes that when a commitment state c is updated via the speech act of an assertion, a proposition φ that corresponds to the assertion is added to the commitment state c . Krifka (2015) notes that φ should ideally neither be entailed by c nor contradict propositions in c . The former is necessary in order to prevent updates with redundant assertions. The latter is already covered by the definition of commitment states and $c + A_\varphi$ is required to be a commitment state.

In order to not only capture assertions but also questions¹⁷, Krifka (2015) introduces COMMITMENT SPACES, which include not only the current commitment state but also appropriate future developments of the discourse. The motivation for including those future developments arises from the assumption that the speech act of a question does not add a proposition to the common ground, but instead it introduces restrictions concerning the kind of future updates that are allowed. I will adopt the notion of commitment spaces as well since it is crucial for my model of clefts to include both assertive as well as interrogative updates.

Informally speaking, a commitment space models the current information state in a discourse (the shared information), but it also models possible and acceptable continuations that can be performed by a speech act (Cohen and Krifka, 2014). In general, acceptable continuations in a commitment state c can be described as those updates of c that do not lead to inconsistencies between the propositions of the updated c . Cohen and Krifka formally define commitment spaces as in Definition 13.

Definition 13 (Commitment Space). *C is a commitment space iff:*

- *C is a set of commitment states;*
- $\exists c \in C \forall c' \in C [c \subseteq c']$

(Cohen and Krifka, 2014:50)

Hence, a commitment space C subsumes a number of commitment states, that include the current commitment state and possible future commitment states resulting from further updates. Formally, a commitment space is a set of sets of propositions. The second condition in Definition 13 ensures that there is a smallest non-empty commitment state in the commitment space C that is contained in all the other commitment states in the commitment space. This commitment state is then identical to $\bigcap C$, the set that contains exactly those propositions that are contained in every commitment state in C . This aims

¹⁷Krifka (2015) also includes the speech act of denegation, as in (i), which I will not discuss here since it is not of particular relevance for the analysis of German clefts.

(i) I don't promise to come to your party.

to formalize the idea that no future commitment state should be allowed in a commitment space that removes any of the formerly accepted propositions. Hence, the propositions contained in the current commitment state should also be contained in all the updated commitment states to come.

Linking this to the common ground, Krifka (2015) calls $\bigcap C$ the ROOT of the commitment space and identifies it as the common ground. For illustration, (8) presents an example of a well-formed commitment space according to Definition 13.¹⁸

- (8) a. Assertion: Lena overslept.
- b. Common ground: {oversleep(Lena), female(Lena)}
- c. Commitment space C : $\{\{oversleep(Lena), female(Lena)\}, \{oversleep(Lena), female(Lena), \neg drink(Lena, coffee)\}, \{oversleep(Lena), female(Lena), meet(Lena, Peter)\}\}$

For this example, assume a context with the assertion in (8-a), resulting in the common ground in (8-b). This CG is of course extremely simplified, and would include much more implicitly shared information. For illustration, however, I assume only these two propositions to constitute the CG. Example (8-c) shows one possible commitment space C , which contains two more commitment states next to the one corresponding to the CQ. The first additional commitment state is updated with the proposition asserted as *Lena did not drink coffee*. The second additional commitment state is updated with the proposition asserted as *Lena met Peter*. As required by Definition 13, there is a commitment state in C , namely the first one, that is contained in all the other commitment states in C . Moreover, $\bigcap C$ is identical to the common ground in (8-b). This is, however, not the only possible commitment space with a root identical to the CG in (8-b). In principle, infinitely many commitment states could be added to the commitment space in (8-c) as long as they are supersets of the CG in (8-b), and do not contain contradicting propositions.

¹⁸I use here a simplified representation of propositions using predicate logic. Hence, the proposition *Lena is female* is represented as *female(Lena)*.

When an assertion is accepted by the discourse participants, the commitment space is updated in the following way (adapted from Kamali and Krifka, 2020:18).

Definition 14 (Assertive Update of a Commitment Space). *For a commitment space C and a proposition φ , $+A_\varphi$ is the assertive update of C with the speech act of asserting φ iff $C+A_\varphi = \{c \in C \mid \varphi \in c\}$.*

Definition 14 states that each commitment state in C that did not contain φ as a possible continuation is deleted in $C+A_\varphi$. Those commitment states in C that did contain φ remain in $C+A_\varphi$. Particularly, this means the former $\bigcap C$ is not in $C+A_\varphi$, assuming that φ represents new information. The new $\bigcap(C+A_\varphi)$, i.e. the new common ground, has been updated with φ .

Just as for the update of a commitment state, the assertive update of a commitment space prevents an update with a contradictory proposition. It is, however, derived slightly differently. By the definition of commitment states, it is not possible to have a proposition and its negation (i.e. a contradiction) in one commitment state. Therefore, if a commitment space is already updated with the speech act of φ , all commitment states will include φ and none will include $\neg\varphi$. Hence, if one then wanted to update with $\neg\varphi$, the result would be the empty set, i.e. an empty common ground, which is obviously not something we want to end up with in a conversation.

According to Definition 13, a commitment space does not necessarily contain all the possible commitment states resulting from consistent updates, but at least one commitment state. If we follow Krifka (2015) in assuming that $\bigcap C$ is the CG, this one commitment state would be identical to CG. For the analysis of natural language discourse, I argue that only those commitment spaces should be considered which contain a commitment state for each combination of consistent propositions. This restriction ensures that there will never be a consistent update that cannot be executed because it is not in any commitment state of the commitment space. According to Definition 14, if a proposition φ , which is consistent with CG in a commitment space C , is not included in any commitment state of C , the update of C with asserting φ would fail and yield the empty set. In general, this

is not a desirable result for an update with a consistent proposition. The update should only be prohibited by relevance restrictions, which will be defined as a discourse principle in section 4.4.

There might be some exceptions of updates that are consistent with CG, but that could nevertheless be excluded. These could be cultural, social, or other contextual restrictions arising, for instance, from taboo topics or a lack of expertise. For a conversation of two four-year-olds, for instance, the commitment space will probably not include commitment states that contain propositions about higher mathematics. In such a situation, one could reasonably exclude commitment states containing those propositions even though they do not conflict with the CG. Apart from such kind of restrictions, I propose to include all commitment states that contain any number of propositions that are consistent with each other and with CG.

Even though updates are prohibited to involve propositions that are inconsistent with the CG, it should still be possible for a discourse participant to disagree or reject a previous statement, e.g. by making a contradicting statement. However, such a rejective assertion would lead to a contradiction with a proposition in the common ground, i.e. in every commitment state of the input commitment space. If the update was nevertheless executed, the result would be the undesirable empty common ground.

In order to allow rejective speech acts and at the same time avoid every commitment state to be deleted, Krifka (2015) introduces sequences of commitment spaces (inspired by the “table” introduced by Farkas and Bruce, 2010). Those sequences of commitment spaces keep track of the development of the commitment space. A discourse participant can go back to a previous commitment space in the sequence in order to retract an update. Farkas and Bruce (2010) model the same effect by assuming that each update is first proposed and stays on a *table* until it is decided to be properly executed. As long as the proposed update is on that table, the speaker commits to it, but not necessarily all the addressees. Thus, one or more of the addressees can still reject it.

Kamali and Krifka (2020:20), furthermore, define different interrogative updates of a commitment space. The basic idea for questions is that they differ from assertions by not changing the root of a commitment space. They just restrict possible continuations. In the following, I will provide a conceptual understanding of interrogative updates by way of two special cases of questions, monopolar and bipolar questions. It is not necessary to extensively discuss the formal details since my discourse model is compatible with any formal model that includes an assertive and an interrogative update for speech acts (see Krifka (2015) for a more elaborate formal implementation of interrogative updates).

The first special case is what Krifka (2015) calls a *monopolar question*. It is a kind of biased polar question, such as the one in (9).

(9) I passed the exam?

The speaker asking (9) rather expects the answer to be *yes* than *no*. The question is biased towards the positive answer. Hence, Krifka (2015) assumes that replying with *no* to such a question would require some kind of retraction operation of the positive answer, before updating with the negative answer. This leads to Definition 15 of the update with a monopolar question.

Definition 15 (Interrogative Update with a Monopolar Question). *An update of a commitment space C with the speech act of the monopolar question $?φ$ is defined as $C+A_{?φ} = \bigcap C \cup \{c \in C \mid φ \in c\}$*

(Kamali and Krifka, 2020:20)

In this definition, $?φ$ corresponds to the interrogative sentence radical corresponding to a polar question, such as (9) for instance, while $φ$ refers to the proposition expressing the positive answer, i.e. *I passed the exam* for (9). Definition 15 defines the update with a monopolar question just like the assertive update with the proposition expressing the positive answer, with the only difference that the former common ground, $\bigcap C$, is not removed from the updated commitment space, i.e. $\bigcap C = \bigcap(C+A_{?φ})$. Hence, only commitment states that contain the positive answer $φ$ to the monopolar question are allowed in the commitment space that is updated by a monopolar question, just like in

the assertive update. Unlike in the assertive update, however, the update is not actually executed since $\bigcap C$ does not contain φ . By Definition 15, this also means that $\bigcap(C+A_{?\varphi})$ does not contain φ . This models the understanding of a question as just the proposal of an update but not as the actual update. As mentioned above, answering question (9) with *no* would require retraction of the proposed update.

A non-biased question, such as (10-a), could be analyzed as a bipolar question, as in (10-b).

- (10) a. Did I pass the exam?
- b. Did I pass the exam, or not?

Kamali and Krifka (2020) define such a bipolar question as a disjunction of updates with monopolar questions, as follows.

Definition 16 (Interrogative Update with a Bipolar Question). *An update of a commitment space C with the speech act of the bipolar question $[?\varphi \vee ?\neg\varphi]$ is defined as:*

$$C+A_{[?\varphi \vee ?\neg\varphi]} = \bigcap C \cup \{c \in C \mid \varphi \in c \vee \neg\varphi \in c\}$$

(Kamali and Krifka, 2020:21)

The speech act of a bipolar question can be understood as a proposal to update with φ or with $\neg\varphi$. The common ground, $\bigcap C$, is again left intact. This time, none of the answers would need a retraction before the assertive update. Moreover, the effect of the update with a bipolar question is not as strong as for the monopolar question. The only commitment states that are excluded in the updated commitment space are those that contain neither φ nor $\neg\varphi$. Intuitively, this update is not a proposal of a certain update, as for the monopolar question. Instead, it is rather a proposal of considering to make a choice between two possible updates. It is left open what the update should be.

Definition 16 concerns just polar alternative questions, such as (10-b), but it is possible to extend it to other alternative questions such as (11).

- (11) Does Laura like Natalio or Felipa?

Krifka (2015) extends interrogative updates to also cover *wh*-questions. Basically, he treats *wh*-questions like alternative questions, and translates (12-a) to (12-b).

- (12) a. Who does Laura like?
b. Who does Laura like, Natalio or Felipa?

This alternative question corresponding to the *wh*-question would include as many alternatives as there are in the context. For (12), there are only two relevant individuals in the context, Natalio and Felipa. Krifka (2015) introduces a lot of extra machinery in order to model all kinds of questions, which I will not include in this thesis. My approach to discourse and clefts in discourse does not rely on this specific definition of common ground and context update. Any other system that incorporates the concept of an update with an assertive or an interrogative speech act would serve the same purpose. It is important for my approach that the system involves the relation to the speech act since my analysis will be concerned with judging the expectedness of certain speech acts to be performed.

Based on the definition of focus developed in chapter 3, repeated in (13), focus can now be modeled in commitment space semantics as well.

- (13) Focus in a discourse move α indicates the presence of a (covert) question q such that α is focus congruent to this q . (Definition 8 - repeated)

Kamali and Krifka (2020:23) state that “focus is modeled as a restriction on the input commitment space C .” This means that the input C for an assertion that is focus congruent to a question q is required to be partitioned with respect to q . For the example $NINA_F$ is petting Mavi, the input C has to be partitioned with respect to the question *Who is petting Mavi?*. In other words, only those continuations should be included in the input commitment space C that result in an update with a proposition of the type *x is petting Mavi*. This input commitment space is identical to a commitment space that has been updated with the interrogative speech act of asking *Who is petting Mavi?*. Kamali and Krifka (2020) point out that it is also possible to accommodate such a restriction on

the input commitment space. In section 4.5, I will argue that it is actually the current question that is accommodated.

4.3 Question under Discussion Framework

In this section, I introduce the Question under Discussion framework as the foundation for my own discourse model, which will be discussed in chapter 10. The term *Question under Discussion* (QUD) was introduced by Roberts (1996). I will cite Roberts (2012), an updated version of that article. Roberts (2012) argues that at every point of a discourse there is a set of unanswered, but answerable, QUDs that the interlocutors aim to answer together. These may be explicitly uttered in the discourse or implied by the context. QUDs correspond to what Carlson (1982) calls *topics or subject matter*, and Roberts argues that answering the QUDs constitutes the cooperative goal in the dialogue game of agreeing on the way the world is.

Roberts (2012) models the set of QUDs as a push-down store, and calls it the QUD STACK (Roberts, 2012:15). Every new question that is accepted by the discourse participants is added to the stack as the top-most element. Once an answer to a question Q is accepted, Q is removed from the QUD stack. Roberts' model also makes the concept of ACCEPTING A DISCOURSE MOVE more explicit.

According to Roberts (2012), the discourse participants have the choice to ACCEPT a discourse move or REJECT it. In case of an assertion, accepting a discourse move means that the common ground is updated, or worlds are excluded from the context set. Furthermore, the question that this assertion answers is removed from the QUD stack. As we have seen in the previous section, however, an update with a question does not really result in a change in the common ground. Instead, Roberts (2012) argues that with accepting a question the discourse participants form the intention to answer it, and this intention is added to the common ground. The addressee commits to answering the question. In Roberts' model, accepting a question further means that the question is added to the QUD stack.

Accepting a discourse move (assertion or question) also relies on RELEVANCE. An assertion only results in a context update in case it is relevant for the discourse. A question is only added to the QUD stack if it is relevant to the former top-most question. Roberts (2012) defines relevance for a discourse move as follows.

Definition 17 (Relevance of a Discourse Move). *A move m is RELEVANT to the top-most question under discussion q , iff m either introduces a (partial) answer to q (m is an assertion) or is part of a strategy to answer q (m is a question).*

(Roberts, 2012:21)

Providing a partial or a complete answer to the QUD means making an assertion that excludes at least one but not all alternatives raised by the QUD (partial answer), or that removes all except for one of the alternatives raised by the QUD (complete answer).¹⁹ The relevance of a question, in contrast, involves the concept of a STRATEGY introduced by Roberts (2012), which I will address now. The following example of a discourse illustrates a strategy in a context in which the only individuals are Hilary and Robin and the only food items are bagels and tofu.

(14) Q_{main} : Who ate what?

- a. What did Hilary eat?
 - i. Did Hilary eat bagels?
 $\text{Ans}(a_i) = \text{yes}$
 - ii. Did Hilary eat tofu?
 $\text{Ans}(a_{ii}) = \text{yes}$
- b. What did Robin eat?
 - i. Did Robin eat bagels?
 $\text{Ans}(b_i) = \text{yes}$
 - ii. Did Robin eat tofu?
 $\text{Ans}(b_{ii}) = \text{yes}$

(Roberts, 2012:16)

¹⁹This holds at least for those approaches that assume mutually exclusive alternatives for the semantics of questions.

It is apparent in (14) that the strategy is based on sub-questions, which were defined in section 2.6 via entailment of the answers. The questions in (14-a) and (14-b) are sub-questions of the hierarchically higher question Q_{main} . Moreover, the questions (a-i) and (a-ii), and (b-i) and (b-ii) are sub-questions of (14-a) and (14-b), respectively. When exposed to the highest question Q_{main} , the addressee could immediately provide a complete answer, such as *Hilary and Robin both ate bagels and tofu*. However, it is also possible to follow a strategy by asking a number of sub-questions, the answers of which jointly provide the answer to the super-question. Example (14) demonstrates a successful strategy, given that the information provided by the answers to the sub-questions entail the answer to Q_{main} .

This example, furthermore, illustrates that assertions may answer more than one question at a time. At the point of the discourse in (14) when the question *Did Robin eat tofu?* is asked, there are two QUDs on the stack: (b-ii) (*Did Robin eat Tofu?*) and Q_{main} (*Who ate what?*). The questions in (14-a) and (b-i) have been removed from the stack because they were answered. The answer $Ans(b_i) = yes$ settles both questions that remained on the QUD stack at that point of the discourse. Hence, when the answer is accepted, question (b-ii) and Q_{main} are both removed from the stack. More generally, when a question is removed from the stack, all the questions that are entailed by this question will also be removed.

These considerations about acceptance conditions of discourse moves and relevance lead to the following definition of ACCEPTING A DISCOURSE MOVE, based on Roberts (2012).

Definition 18 (Accepting a Discourse Move). *A discourse move m is accepted iff it is relevant to the top-most question under discussion. Then,*

- i. *the common ground is updated, if m is an assertion, or*
- ii. *m is added to the QUD stack as the top-most element, if m is a question.*

This definition follows Roberts (2012:5), who notes that she defines discourse moves as semantic objects, not as speech acts. Recall that Krifka (2015), in contrast, considers

context updates, but not common ground updates, to involve speech acts (section 4.2). My analysis in chapter 10 will also need to include speech acts in order to account for the acceptability of discourse moves realized by clefts in German which involve considerations about potential speech acts.

Formally, the QUD stack is an ordered set of questions. The questions are ranked with respect to the time they entered the stack. Again, what Roberts (2012) calls *questions* here, is the semantics objects not the speech act of asking a question. The top-most question is the highest ranked question in the ordered set and it turns out to play a special role. Roberts calls it the *immediate* question, and it is often referred to as the *current question*. In the following, I present an extension of Roberts' model, based on Simons et al. (2017), who distinguish the current question and the discourse question.

4.4 Current Question and Discourse Question

Roberts (2012) defines the immediate question under discussion as the last question added to the QUD stack at the time of a discourse move. This immediate QUD has later been referred to as the CURRENT QUESTION (CQ). On a more conceptual level, Beaver and Clark (2008:35) describe the current question as a “question that is proffered and mutually accepted by the interlocutors as the most immediate goal of the discourse”. The term *proffered content* is introduced by Roberts (2012:5) as the content that is asserted in an assertion and as the content that is non-presupposed in a question. The latter basically amounts to whatever is contributed by the semantics of the question, apart from presuppositions.

Put more formally, a CQ is a semantic object of a question, just like the semantic representation of the actual utterance of a question. Hence, it is a set of q-alternatives, when following the alternative semantics approach from section 2.3. Beaver and Clark (2008) propose some further restrictions for the CQ in form of the CURRENT QUESTION RULE, repeated in (15).

(15) **Current Question Rule**

The Current Question must contain at least one true alternative, and contain multiple alternatives which are not resolved as true or false in the common ground.

(Beaver and Clark, 2008:36)

This rule basically describes the requirement that the CQ should be an open, hence unresolved, question in the given context. Given the understanding of CQ as a question that is supposed to be accepted as a discourse goal, it is not reasonable to accept a question that does not raise at least two alternatives to choose from. This is reminiscent of inquisitiveness, as defined in Definition 3 in section 2.5. Under such an inquisitive definition of questions, the second clause in (15) would not be necessary, since it describes a general feature of questions in inquisitive semantics.

The first part of the CQ rule requires CQ to have an existence presupposition. This presupposition has been discussed for questions in general in section 2.7. Although it is still debated, whether questions have an existence presuppositions, most approaches assume that they do. Observations like the following speak in favor of an existence presupposition of questions.

- (16) Who likes Natalio? – #Somebody likes Natalio.

If there was no existence presupposition for questions, the answer in (16) would be incorrectly predicted to be appropriate addressing the preceding question. It would be a partial answer, which excludes one alternative, namely the one where nobody liked Natalio. However, if it is presupposed that somebody liked Natalio (as required by the CQ rule), then the answer in (16) is correctly predicted to be uninformative and, thus, unacceptable.

Other observations are problematic for the existence presupposition of questions, such as the “negative answers” in (17).

- (17) a. Who likes Natalio? – Nobody.
b. What does Laura like? – Nothing.

Such an answer would cancel the existence presupposition, which is usually only possible for presuppositions using extra linguistic marking, such as *wait a minute* (see section 2.7 for a more detailed discussion).

In any case, the question whether questions have an existence presupposition or not is something that should be decided for questions in general. Hence, the CQ rule does not need to contain the requirement of at least one true alternative. If one would want to still include it, it would be problematic for those approaches to questions that do not include “negative” answers as alternatives. Such approaches would exclude CQs for which the only true answer would be a negative answer. However, those approaches usually come up with a way how to deal with negative answers. Again, I want to point out that those properties are not specific for CQs but for questions in general, and should, thus, be defined independently of CQs.

Definitely, we want to avoid defining different semantic objects for questions that correspond to an actual speech act and CQs, i.e. questions that are elements of discourse. This distinction is not desirable since one very obvious way of introducing a CQ into the discourse is by making a speech act of asking exactly that question. Accordingly, precisely that question should be added to the QUD stack, not a slightly different object. Concluding, the CQ rule needs to either be adjusted depending on the approach taken towards the semantics of questions in general, or it can even be discarded entirely given that it does not contain anything that does not hold for questions anyways.

I will now present a different understanding of the CQ, which is inspired by Velleman et al. (2012). Their basic assumption is that the CQ is a parameter that is provided by the context, and its semantic representation does not differ from interrogative sentence radicals used in speech acts. This understanding is embedded in the assumption that there is something like Beaver and Clark’s (2008) focus principle, repeated in Definition 19.

Definition 19 (Focus Principle). *Some part of a declarative utterance should evoke a set of alternatives containing all the Rooth-Hamblin alternatives of the CQ.*

(Beaver and Clark, 2008:37)

This principle is very similar to what was defined as *focus congruence* in Definition 7 (section 3.4). There, it was already pointed out that there is a strong relationship between focus and questions in general. The final definition of focus in chapter 3, repeated in (18), refers to some (covert) question, as well.

- (18) Focus in a discourse move α indicates the presence of a (covert) question q such that α is focus congruent to this q . (Definition 8 - repeated)

Beaver and Clark (2008) identify this question q as the current question. Hence, focus is dependent on the CQ. Here are some examples that illustrate the relation between focus and the CQ. Example (19-b) shows a narrow focus resulting from the CQ in (19-a).

- (19) a. **CQ:** Who talked to Peter?
 b. LENA_F talked to Peter.

This is basically the same as was demonstrated for focus congruence in section 3.4, only specified with respect to the current question. For a cleft sentence, I assume that its pivot associates with the CQ, as (20) indicates.

- (20) a. **CQ:** Who talked to Peter?
 b. It was Lena who talked to Peter.

This assumption is based on the observation that narrow focus canonical sentences and clefts behave similarly, which has been pointed out frequently. É. Kiss (1998), De Veaughe-Geiss et al. (2015), and many others argued that clefts mark focus on their pivot. For clefts with a narrow focus inside of the pivot, as in (21), I assume that still the entire pivot is associated with the CQ, not just the focused part of the pivot. This leads to (20-b) and (21-b) addressing the same question.

- (21) a. **CQ:** Who talked to Peter?
 b. It is LENA'S_F boyfriend who talked to Peter.

This assumption is still debated and is probably not correct. Example (21-b) seems to rather address *Whose boyfriend did Peter talk to?*, instead of (21). I will not be able to solve this issue in this thesis and will make this simplifying assumption in order to not over-complicate things (see Velleman et al. (2012) and É. Kiss (1998), among others, for discussion).

Based on Beaver and Clark's (2008) CQ rule, Simons et al. (2017) propose the following definition of current question.

Definition 20 (Current Question for an Utterance). *The CQ for an utterance is a privileged subset of the focal alternative set of the uttered sentence (given a structural analysis of that sentence, including focus marking) which meets the following conditions:*

- i. *The proposition expressed is a member of the CQ and*
- ii. *the CQ has at least one additional member.*

(Simons et al., 2017:194)

This definition covers the Current Question Rule, and adds the requirement of focus congruence to it. This requires that the focus alternatives must be known independently of the CQ, given that Definition 20 defines CQ based on the focal alternatives. I consider this problematic, at least when focus is defined as in (18), namely as indicating the CQ. It seems circular, as focus is defined based on knowing the identity of the CQ, and the CQ is defined based on knowing what the focal alternatives are. The only case where the focus alternatives can be derived from the linguistic form of an utterance only is when focus is unambiguously marked at the surface, such as it is the case with narrow subject focus in German in (22).

(22) NINA was petting Mavi.

This utterance can only address the CQ *Who was petting Mavi?*, which means that the focal alternatives can be derived without knowing the CQ. Nevertheless, there are those examples that have an ambiguous focus marking, such as (23), which are not covered by Definition 20.

- (23) Nina was petting MAVI.

The CQ of example (23) could be *Who was Nina petting?*, or *What was Nina doing?*, among other questions. I will discuss this issue in more detail in section 4.5, where it is described how the speaker can accommodate the CQ based on focus marking. In order to avoid the problems that arise with the previously presented definitions of CQ, I favor a definition of CQ that is less restricted and is compatible with different semantic approaches of questions.

Furthermore, there is one important detail of the CQ that previous approaches ignored, namely how CQ comes into existence. How does the speaker know what the CQ is, and how does the hearer know? As mentioned above, I will follow an approach based on Velleman et al. (2012), which assumes that the CQ comes into existence just by being provided by the context. More specifically, it is provided by the speaker in the context. Assume that there are no restrictions on the CQ, as long as it is a question and it is determined by the speaker. This leads to the following rather weak definition of the current question.

Definition 21 (Current Question). *For a discourse move m , the current question (CQ) is the question that a speaker aims to address with m . CQ is determined by the speaker.*

Addressing a question Q means uttering Q or asserting an answer to Q . This definition, in principle, allows the speaker to choose a CQ that is irrelevant to the preceding discourse. However, the CQ will still play an important role in the process of planning a discourse move or accepting a discourse move. We will see that principles of relevance and acceptance conditions will prevent the speaker from choosing an irrelevant CQ. Therefore, those principles need to be revised, which will be executed in the following.

For the revised versions of *relevance* and *accepting a discourse move*, I follow Simons et al. (2017), who introduce the notion of a DISCOURSE QUESTION (DQ), next to the CQ. The discourse question is a discourse-segment relative notion and can intuitively be interpreted as the topic of a (sub-)inquiry. This concept can also be found in van Kuppeveld (1995) who calls it the question that corresponds to the discourse topic.

Simons et al. (2017) refer to only the *current* topic with the notion of DQ. I argue, however, that it is possible that there are several possible topics, and, thus, several DQs in a discourse. Moreover, I assume that the DQ is also given in the context, as the topic that is mutually agreed on by the interlocutors. It will turn out that accepting a question, for instance, will lead to establishing a new DQ in the discourse.

Now, the definitions of *relevance* and *accepting a discourse move* can be revised, based on Definition 21 of CQ and including the DQ. Definition 22 and Definition 23 define relevance for the discourse move of an assertion or a question, respectively.²⁰

Definition 22 (Relevance of an Assertion – revised). *For a given current question Q_c and a given discourse question Q_d , an assertion φ is relevant to Q_d , iff*

- a) φ is focus congruent to Q_c , and
- b) Q_c is either identical to Q_d or a sub-question of Q_d .

Definition 23 (Relevance of a Question – revised). *For a given current question Q_c and a given discourse question Q_d , a question $? \varphi$ is relevant to Q_d , iff*

- a) $? \varphi$ is identical to Q_c , and
- b) $? \varphi$ is a sub-question of Q_d .

Compared to Roberts' (2012) definition of relevance (Definition 17), now both a CQ and DQ must be given for determining relevance of an assertion or a question. Roberts' definition only required a CQ. Relevance is now defined as a relation between a discourse move and the DQ, with the CQ linking the two.

Consider first the definition of relevance of an assertion in Definition 22. Condition a) in Definition 22 accounts for focus congruence between the CQ and the assertion, which was already part of the definition of CQ in other approaches, or defined as the focus principle by Beaver and Clark (2008). This approach, however, makes it a relevance

²⁰These definitions involve the semantic object of assertions and questions, i.e. propositions or sets (of sets) of worlds, and determine relevance conditions for these objects. Accordingly, φ and $? \varphi$ refer to semantic objects, not to speech acts. The same holds for Definition 24 and 25 below.

condition. Intuitively, this can best be understood from the speaker's perspective. The speaker decides which question to address (the CQ) and aims for the addressee to be able to derive this question in the discourse. This is only possible when there is focus congruence between CQ and the speaker's utterance. As we will see in the next section, focus congruence is not sufficient for the addressee to identify every CQ, though.

Condition b) in Definition 22 is equivalent to what Roberts' definition of relevance proposed. Consider the two options of condition b): (i) If CQ and DQ are identical, the discourse question is addressed directly. (ii) If CQ is a sub-question of DQ, it means that CQ is part of a strategy to answer DQ.

For the relevance of a question in Definition 23, the only difference to Roberts' definition is the intermediate step of the CQ between the question φ and the discourse question Q_d . This captures the idea that it is the speaker who determines the CQ, namely by presenting it as an explicit question (condition (a) in Definition 23). Still, the speaker needs to relate it to some commonly shared question of interest, which used to be the CQ in Roberts' system. This commonly shared question of interest is now represented by the DQ. The sub-question relation in condition b), again, expresses that the uttered question φ is supposed to be part of a strategy to answer DQ, as in Definition 22 and as proposed similarly by Roberts' with respect to the CQ instead of the DQ.

A brief note about why I use the sub-question relation in these definitions of relevance: In principle, there are other relations than the sub-question relation that could hold between CQ and DQ, and that would still result in a relevant discourse move. One such case is the *dependent question* relation. Ginzburg (2012), among others, argues that a dependent question should be analyzed as a relevant reply to a question, as well. However, dependent questions are not sub-questions of the question they answer. The following example by Ginzburg (2012), repeated in (24), illustrates the dependency between questions.

(24) A: Who is going to win the race? – B: Who is going to participate?

(Ginzburg, 2012:56)

Who is going to win the race depends on who is participating, for this example. The answer to the second question has a good chance of eventually leading to an answer to the first question. Hence, B's question should be categorized as a relevant discourse move, even though it is not a sub-question of A's question. Ginzburg (2012) describes the dependence between two questions q_2 and q_1 as follows:

Discussion of q_2 will necessarily bring about the provision of information about q_1 .(Ginzburg, 2012:57)

Groenendijk (2015), in contrast, points out that there are cases of dependent questions in which the second question does not lead to an answer to the first question. In this thesis, I will exclude dependent questions as relevant discourse moves for reasons of simplicity, but I acknowledge that they do constitute relevant discourse moves. If necessary, Definition 22 and 23 can be adapted to also account for dependent questions (see Ciardelli (2014) for a formal implementation).

There is yet another category of relevant questions, the so-called potential questions, such as the question in (25).

(25) Maria met a mysterious man yesterday. – Who was that man?

Potential questions play an important role for the analysis of clefts, and will be discussed in detail in section 4.6. The discourse model that I will propose in section 10.3 will then incorporate potential question.

Let us sum up the advantages of this approach to CQ compared to Beaver and Clark (2008) or Simons et al. (2017). The new definition of CQ is compatible with whatever approach one wants to take to the semantics of questions in general. Furthermore, this definition of CQ does not cause any circularity when one assumes a definition of focus that involves the existence of a CQ, since CQ is not defined via focus. The effect is that the CQ itself does not do much in discourse, but instead, the discourse effects of an utterance are modeled by relevance conditions. For instance, focus congruence or the sub-question relation are part of the definition of relevance instead of the definition of CQ.

Some approaches, such as inquisitive semantics, might satisfy some or all of the relevance conditions already in the definition of questions. However, this would just cause some redundancy, but no incompatibilities.

Finally, Definition 24 and Definition 25 revise Definition 18 (accepting a discourse move) for assertions and questions, respectively, based on the revised definition of relevance.

Definition 24 (Accepting an Assertion – revised). *For a given current question Q_c and a given discourse question Q_d , an assertion is accepted iff it is relevant to Q_d . Then, the common ground is updated and Q_c is removed from the QUD stack.*

Definition 25 (Accepting a Question – revised). *For a given current question Q_c and a given discourse question Q_d , a question $?φ$ is accepted iff it is relevant to Q_d . Then, $?φ$ becomes a new discourse question.*

Definition 24 for accepting an assertion says the same as the former definition of accepting a discourse move stated for assertions, with the only differences that relevance is required with respect to the DQ instead of the CQ. However, it is still CQ that is removed from the stack. Of course in the case in which CQ and DQ are identical, both questions are removed from the stack.

Definition 25 for accepting a question is also formulated with respect to the discourse question. This means again that relevance is required with respect to DQ. What is also new in this revised version is that the question $?φ$ becomes a new DQ. From Definition 23 (Relevance of a Question), it follows that actually CQ becomes a new DQ, since Relevance requires $?φ$ to be identical to CQ.

A consequence of this approach to the CQ and its embedding in the discourse rules is that the QUD stack introduced above loses importance, which is not just a coincidence as described below. The definition of CQ, as well as the definitions of relevance made no reference to the QUD stack. Crucially, it is not necessary for the CQ to be the top-most element of the the QUD stack, in Roberts' terms. The definition of accepting a question does not involve the QUD stack, either, and the definition of accepting an assertion does not hinge on CQ being removed from the QUD stack.

On the one hand, it is possible to give up the QUD stack entirely, since it does not seem to do much. On the other hand, the stack is a good means to keep track of strategies. One motivation for Roberts (2012) to introduce the QUD stack in the first place was to prevent discourse participants to move on the next topic before the current QUD is settled. This can be exemplified by comparison to how a newspaper is structured. Assume the first section is concerned with politics, thus addressing *What happened in politics?*. It has the subcategories addressing *What happened in national politics?*, and *What happened in international politics?*, each of them with further subcategories. It will not include a section addressing *What happened in sports?*. The category *sports* is only addressed once politics is covered. A conversation is assumed to be structured in a similar way. When the current QUD is $Q_1: What\ happened\ in\ politics?$ as a sub-question of $Q_2: What\ is\ new\ in\ the\ world?$, it should not be possible to immediately ask $Q_3: What\ happened\ in\ sports?$. Roberts' QUD stack captures this by assuming that the top-most question on the stack, here Q_1 , always has to be addressed first before a new question that is not a sub-question, such as Q_3 , can be added to the stack.

This makes the correct predictions for many cases. The discourse model presented above, in contrast, does not make such strong predictions since it allows there to be different DQs that could each be addressed by the CQ. Those DQs are not ordered. Hence, nothing prevents a discourse participant to already address a new DQ while another DQ is not settled yet. This prediction is, however, absolutely intentional, given that clefts have been observed to frequently address questions that are not the top-most question in the stack. There are further constructions and other linguistic means that have the same effect, e.g. phrases such as *by the way* or changing the word order or the tense. This is exemplified in (26).

- (26) A: What happened in politics in 2000?
 B: Bush won the election. And what happened in sports?
 A:
 a. #Ist Bush gegen Al Gore angetreten?
 is Bush against Al Gore run
 ‘Did Bush run against Al Gore?’

- b. War es Al Gore, gegen den Bush angetreten ist?
 was it Al Gore against whom Bush run is
 ‘Was it Al Gore that Bush ran against?’
- c. War Bush gegen Al Gore angetreten?
 was Bush against Al Gore run
 ‘Had Bush run against Al Gore?’

In this example, speaker A ignores B’s question, which seems much more acceptable using the variants in (26-b) or (26-c) than the variant in (26-a). Speaker A asks a different question that is not a sub-question of B’s question, but that is rather related to the A’s own previous question. Roberts (2012) predicts (26-a), (26-b), as well as (26-c) to be inappropriate because they are not sub-questions of the top-most question on the QUD stack, which would be *And what happened in sports?*. However, this is only correct for (26-a), which is a canonical question without extra marking. Example (26), however, shows that it is possible to address a different question than the top-most question when using a cleft question, as in (26-b) or past perfect instead of present perfect, as in (26-c). Hence, those constructions seem to ignore the order of the QUD stack. If the stack is given up entirely, one would still want to be able to account for the unacceptability of (26-a), though. In section 10.2, I will thus introduce a ranking based on expectedness of speech acts of questions, which replaces the QUD stack. Depending on the construction, different degrees of expectedness of the question are allowed.

If one wanted to keep the QUD stack, as it is, as part of this model, it is worth noticing that this is possible and does not contradict with any of the assumptions of the revised definitions above. Definition 24 even explicitly makes reference to the QUD stack. However, given that the other definitions do not involve the QUD stack, it is unclear how and when the CQ actually enters the QUD stack. According to Roberts (2012), a question enters the stack when the interlocutors agreed it to be a question worth discussing. Now that the speaker determines the CQ, it is not obvious anymore when it enters the QUD stack.

In order to combine the new approach with the model of a QUD stack, two assumptions are necessary. First, we need to assume that the CQ is the top-most element for the

speaker once she/he addresses it. This can also be understood in a way that the speaker tentatively adds the CQ as the top-most element, or tentatively moves a question to the top of the stack. Second, we need to assume that the addressee can accommodate the CQ under appropriate conditions. If the addressee accepts a CQ, she/he adds it to the stack as well (if it was not yet there already for the addressee) before making the update. The case where the addressee has to spontaneously add the CQ determined by the speaker is referred to as accommodation of the CQ, which is discussed in the next section. Such a tentative proposal of a CQ by the speaker has been discussed in *table* model by Farkas and Bruce (2010), mentioned above. According to this model, a proposed discourse move is parked on the table until the interlocutors have discussed whether to accept it.

Hence, with these assumptions, it is possible to combine this approach with the model of a QUD stack. In section 10.3, I introduce an adapted version of the QUD stack of questions ranked by how expected they are.

4.5 Accommodation of the Current Question

In the following, I will explain what it means for the addressee to accommodate the CQ. It is the addressee's task to accept or reject a discourse move. This task requires the identification of the CQ, as required by relevance (Definition 22 and 23). So far, it is not obvious how the addressee manages to do that, given that the CQ is entirely determined by the speaker. For this purpose, I will now discuss relevance and accepting a discourse move from both the speaker's/writer's as well as the addressee's perspective. They turn out to be quite different. The difference arises from the fact that the speaker and the addressee have different access to the identity of the CQ.

From the speaker's perspective, it should always be unambiguously clear what the CQ is. The speaker him-/herself knows which question she/he wants to address with a discourse move. In other words, the speaker decides which CQ to address when planning her/his utterance. Thus, for every discourse move, the CQ is determined by the context through the speaker. However, the speaker will not pick just any question because she/he has

an interest in the addressee accepting her/his utterance. In order to achieve that, the speaker will pick a CQ that leads to a relevant discourse move.

The addressee, in contrast, is in an entirely different situation. She/he is exposed to a discourse move made by the speaker and needs to decide whether to accept it or reject it. For making a choice about accepting or rejecting a discourse move, the addressee needs to identify the CQ, given that both Relevance and Accepting a Discourse Move refer to CQ. Now, there are two situations to distinguish: (i) The addressee knows what the CQ is (usually because the present discourse move is a question), or (ii) the addressee does not know what the CQ is.

In the first situation, the addressee has an easy job since the CQ is provided by the utterance. Hence, she/he just needs to identify the question that was asked as the CQ. Most frequently, however, the addressee finds him-/herself in the second situation, where it is not obvious to her/him what the CQ is. This situation is particularly common in narratives, which usually do not include many explicit questions. In that situation, the addressee needs to accommodate what the CQ is in order to decide whether the move is relevant and can, thus, be accepted.

As a cue for the CQ, the addressee may use the focus marking of the utterance, given the requirement for focus congruence of assertions to the CQ. This does not mean, however, that the CQ is entirely determined by the focal alternatives of the utterance. One complication is focus projection, as introduced in section 3.1. Hence, when the addressee is exposed to the assertion in (27), focus projection leads to different possible foci.

(27) Maria became an ACTRESS.

- a. CQ₁: What did Maria become?
- b. CQ₂: What happened to Maria?

Since focus is ambiguous, there are several possible CQs, for instance those in (27-a) and (27-b). Hence, the addressee cannot conclude from the focus marking what the CQ is. It is rather the other way around. The context (more specifically the speaker), not the focus,

determines the CQ. And a coherent assertion is just required to be focus congruent to this question, by Relevance. In line with the final definition of focus, as presented in (18), focus of an utterance does not determine but *indicate* which CQ the context provides.

To illustrate this, I will use an analogy to how the speaker of an utterance is identified in a context.²¹ When an utterance uses the first person pronoun, the referent that it refers to, namely the speaker or author, is provided by the context. As described for the CQ above, the addressee sometimes knows who the author or speaker of an utterance is and sometimes she/he does not. When she/he knows who the speaker is, it is trivial that the first person pronoun refers to that individual. This corresponds to the case, in which the CQ is known to the addressee.

However, there are contexts where the addressee does not know but can nevertheless accommodate who the speaker of an utterance is. Imagine finding a sticky note with a message, such as (28), on the floor of an office of two colleagues, one female and one male, not knowing who of the two wrote it.

- (28) (Yo) estoy enferma y me fui temprano a casa.
(I) am sick.FEM and PRO.1SG.REFL went early to home
'I am sick and I went home early.'

In a language that has gender agreement between the speaker/author and a predicative adjective, e.g. Spanish, the addressee can accommodate who the author of the note is, even if she/he did not know before who of the two colleagues wrote the note. The grammar of Spanish marks gender on adjectives, which helps the addressee conclude that the female colleague must have written the message. The adjective in (28) is marked as *feminine*, i.e. *enferma* ('sick') instead of the masculine version *enfermo*. The speaker, that *yo* ('I') refers to, however, is still provided by the context and not by the feminine marking on the adjective. If there were two female colleagues, for instance, the identification of the speaker would not work anymore. Hence, it is just a coincidence that the speaker can be identified through the gender marking. The female marking has a syntactic function

²¹Krifka and Musan (2012:10, footnote 4) mention something similar.

in Spanish and only, as a side effect, gives the addressee a hint for the identity of the speaker.

The identification of the CQ, when it is not known to the addressee, can be described analogously. From the focal alternatives of an utterance, the addressee can often conclude what the CQ is (cf. Krifka, 2008:251). However, focus does not determine the CQ. This is indicated by cases in which the addressee must decide on the CQ based on other cues than focus marking, such as in (29).

- (29) Yesterday, I was at Nina's party. She beat Laura in a POKER game.

This example is analogous to (28) in the domain of CQ. The context is represented by the first sentence in (29). In principle, any question could be the CQ of the second sentence. However, the focus marking of the second sentence allows only some questions, those that are focus congruent. Crucially, this is not just one question. This is analogous to the situation with two female speakers. So, for instance, Q_1 : *What did Nina beat Laura in?*, Q_2 : *What did Nina do at the party?*, and Q_3 : *What happened at the party yesterday?* are such that B's reply is focus congruent to them. Hence, when the addressee interprets the second sentence in the context of the first sentence, she/he cannot decide just based on the focus marking what the CQ is. From world knowledge, the addressee will probably not accommodate Q_3 as the CQ, since it is unusual that just one thing happens at a party.²² The addressee will rather accommodate Q_2 as the CQ since it seems most natural. Question Q_1 is probably not accommodated because one does not assume by default that a party host beats others in games.

Concluding, the addressee quite often needs to accommodate the CQ. In some cases, she/he receives unambiguous cues, such as narrow subject focus (as in example (22) mentioned above), which makes it possible to accommodate the CQ without having to include world knowledge. In other cases, such as (29) the addressee needs to rely on world knowledge because visible focus marking is ambiguous. Once the addressee has

²²Of course, this could be a partial answer to Q_3 . In this case, however, I would assume that a sub-question, probably Q_2 , is answered and not Q_3 directly.

accommodated the CQ, she/he only then knows what the focus is. And only then, the addressee can make a decision about accepting or rejecting the discourse move.

4.6 Potential Questions

Recall that the revised definition of relevance only considered asking or addressing a sub-question of the discourse question (DQ) to be a relevant discourse move. It was mentioned above already that there are other possible relations between a discourse move and the DQ which also result in a relevant discourse move. An example of such a discourse move would be asking or addressing a POTENTIAL QUESTION (PQ), as in the two dialogues in (30).

- (30) a. A: How are you doing?
B: My boss called me an idiot.
A: Why did she do that? (asking PQ)
b. A: How are you doing?
B: My boss called me an idiot. She was unhappy about me being late.
(addressing PQ)

The dialogue in (30-a) shows the potential question *Why did he do that?*. This question seems to form a relevant discourse move in this context, even though it is not a sub-question of any question present in the discourse. It is not a sub-question of A's first question, and even if we assume the existence of an implicit discourse-initial super-question, such as *What happened today?*, it is not a natural strategy to ask why did B's boss do what she did. The question only arises because of the answer provided to A's question. The same holds for the dialogue in (30-b) with the only difference that the PQ is not explicitly uttered, but implicitly assumed and then immediately answered.

According to van Kuppevelt (1995), discourses contain many questions that are not formulated explicitly, but arise implicitly, especially in monologues or narratives. He characterizes IMPLICIT questions as questions “which the speaker anticipates to have arisen

with the addressee as the result of the preceding context." (van Kuppeveld, 1995:110) He also includes sub-questions of explicit or implicit questions and follow-up questions, if the answer to a preceding question was unsatisfactory, in his discourse model. My cleft data consists of written texts, mostly narratives. Hence, explicit questions occur rather infrequently. Accordingly, I will use a model that focuses on the questions that the author anticipates the addressee to have, based on the previous text. Hence, such implicit questions need to be included into the analysis.

Moreover, van Kuppeveld (1995) discusses the notion of satisfactoriness of an answer to a question. He concludes that many implicit questions arise due to unsatisfactoriness of the provided answer to a previous question. I will make use of the concept of satisfactoriness in my adapted discourse model in section 10.3 by incorporating it into the concept of expectedness.

Onea (2016) introduces POTENTIAL QUESTIONS (PQs), and focuses on questions that are evoked by the immediately preceding utterance and are explicitly not sub-questions of preceding questions, such as (31-b) and (32-b).

- (31) a. My boss called me an idiot today.
 - b. **PQ:** Why did she call you an idiot?
-
- (32) a. Lena was really fast in the marathon yesterday.
 - b. **PQ:** Did she make it to the podium?

A potential question is triggered or licensed, as Onea (2016) calls it, by the preceding utterance, instead of being part of a strategy to answer a super-question. Potential questions differ crucially from the discourse question, the current question, and sub-questions thereof. Once a PQ is raised by an utterance, there is no need for it to be addressed in the course of the conversation. This is why they are called *potential* questions. With PQs, the focus is more on them being raised by certain utterances, than being answered (Onea, 2016). The other questions, in contrast, contribute to the common discourse goal as issues that need to be settled in the future.

According to Onea (2016), potential questions are formally defined as in Definition 26.

Definition 26 (Potential Question). *A PQ q licensed by some utterance U in some context c is such that: $c + U \models_{def} p$ and $c \not\models_{def} p$, where p is the union of all highlighted alternatives of q, and there is no p' , a highlighted alternative of q, such that $c + U \models p'$.*

(Onea, 2016:10)

Intuitively, this definition says that a potential question depends on the utterance U and is not licensed in the context that is not yet updated with U . Furthermore, it should be an open question. In order to understand how this is formally expressed, I will introduce the concepts contained in the definition. The definition uses the concept of *defeasible entailment*, written \models_{def} . Onea (2016:11) paraphrases $p \models_{def} q$ as “from p one can reasonable conclude q in some usual context”. Defeasible entailment is a pragmatic, weaker version of classical entailment (see Onea (2016) for a detailed discussion of defeasible entailment).

Furthermore, the definition of PQs uses the concept of *highlighted alternatives*, based on Roelofsen and van Gool (2010). Independently motivated, they assume that there are certain alternatives in the meaning of a question that are highlighted while others are not.²³ I will not go into the formal details of highlighting, but instead explain it using the two examples of PQs mentioned above (see Onea (2016:chapter 3) for a formal definition of highlighting). The meaning of the question in (32-b) is enhanced by marking each alternative as highlighted or not highlighted resulting in (33), where the highlighted alternatives are underlined.

- (33) $\llbracket \text{Did Lena make it to the podium?} \rrbracket = \{ \underline{\text{Lena made it to the podium}}, \text{Lena did not make it to the podium} \}$

²³Roelofsen and van Gool (2010) develop the concept of highlighting in order to explain the different response patterns of the two questions in (i).

- (i) a. Did Lena win?
 b. Did Lena lose?

Question (i-a) and (i-b) have exactly the same meaning in inquisitive semantics. However, the reply *yes* and *no* mean different things depending on the question. For (i-a), *yes* means Lena won. For (i-b), however, *yes* means that Lena did not win. This difference is not captured in the inquisitive representation. Other frameworks, however, do distinguish the semantic representations without the need to define highlighting. The functional approach would yield different results when the functional representation of (i-a) or (i-b) is applied to the answer *yes*.

For a *wh*-question, as in (31-b), several alternatives are highlighted except for one. This analysis assumes a question meaning that actually includes the non-highlighted alternative (such as inquisitive semantics). Accordingly, the meaning of (31-b) leads to (34).

- (34) [Why did she call you an idiot?] = { She called you an idiot because you were late,
She called you an idiot because you annoy her, She called you an idiot because...,
..., She called you an idiot for no reason }

Let us now understand how the definition correctly predicts (31-b) and (32-b) to be PQs of the preceding utterance, respectively. According to Definition 26, the union of all highlighted alternatives must be defeasibly entailed. For (32-b), the union of all highlighted alternatives is just the one alternative (*Lena made it to the podium*). This alternative is defeasibly entailed by the context updated with the utterance *Lena was really fast in the marathon yesterday*, although it is not semantically entailed. It is, however, quite probable that Lena made it to the podium given that she ran really fast, hence the defeasible entailment.

Furthermore, it is exactly this utterance *U* that leads to the defeasible entailment of this PQ. This becomes clearer once the context that is not yet updated with *U* is exemplified. Therefore, compare the two discourses in (35) and (36), the former including the licensing utterance, the latter not containing the licensing utterance.

- (35) a. *c*: Lena successfully organized a marathon that took place yesterday.
 b. *c+U*: She herself ran really fast.
 \rightsquigarrow PQ: Did she make it to the podium?

- (36) a. *c*: Lena successfully organized a marathon that took place yesterday.
 b. –
 $\not\rightsquigarrow$ PQ: Did she make it to the podium?

In those examples, *c* represents the initial context, and *c+U* represents the updated context. Only the discourse in (35), which contains the licensing utterance, leads to

the potential question. The discourse in (36), however does not. Hence, Definition 26 correctly predicts (32-b) to be a PQ of the utterance in (32-a).

For the PQ in (31-b), the union of all highlighted alternatives in (34) can be rephrased as meaning *She called you an idiot for some reason*. Based on our world knowledge and the utterance U (*My boss called me an idiot*), we can conclude that bosses usually do not call somebody an idiot for no reason. Hence, the union is defeasibly entailed by $c+U$. Furthermore, none of the highlighted alternatives, by themselves, are semantically entailed by $c+U$. Notice the difference of this PQ to the PQ previously discussed. This time, the union of the highlighted alternatives is not identical to the only highlighted alternative. Instead, each highlighted alternative by itself is required to not be semantically entailed, while only the union is required to be defeasibly entailed. Thus, we conclude that the PQ in (31-b) is licensed in the context of the utterance in (31-a).

Additionally, Onea (2016) presents three special cases of PQs, that I will also refer to in my analysis of *es*-clefts in German. Hence, I will introduce them briefly. Onea defines STANDARD POTENTIAL QUESTIONS as follows.

Definition 27 (Standard Potential Questions). *A PQ q licensed by some utterance U in some context c is a Standard Potential Question (SPQ) licensed by U in c , iff $c + U \models p$, where p is the union of all highlighted alternatives in q .*

(Onea, 2016:14)

The SDP differs from the PQ insofar that the union of its highlighted alternatives is not only defeasibly entailed but even semantically entailed. Hence, this holds only for *wh*-questions. Polar PQs would contradict the requirement for SPQs of the semantically entailed union of highlighted alternatives since the definition of PQs requires this union to not be semantically entailed. This results from the fact that the union of the highlighted alternatives is identical to the only highlighted alternative.

Hence, only *wh*-questions can be SPQs, for example the one in (37).

- (37) a. Somebody has a secret.
- b. SPQ: Who has a secret?

The union of the highlighted alternatives of (37-b) is identical to the utterance in (37-a). Therefore, (37-a) trivially semantically entails (37-b), and is, thus, an SPQ. In (31-b), in contrast, the union of the highlighted alternatives (She called you an idiot for some reason) is really only defeasibly entailed. It is still possible that she had no reason, although this seems unlikely.

Onea (2016) claims that SPQs cannot only be licensed by assertive utterances but also by questions. The question in (38-a), for instance, licenses the SPQ in (38-b).

- (38) a. When did you get home yesterday?
b. SPQ: Why are you asking?

Like other potential questions, this SPQ only arises after the first question has been uttered. This shows that also questions can license PQs. Onea (2016) points out that these SPQs are frequently located on a meta-linguistic level, as it is the case in this example as well.

The second special case is the PRIMARY POTENTIAL QUESTION, as defined below.

Definition 28 (Primary Potential Questions). A *PQ* q licensed by some utterance U in some context c is a Primary Potential Question (PPQ) licensed by U in c , iff the set of highlighted alternatives in q is compositionally derived or made salient by U .

(Onea, 2016:14)

Onea (2016) distinguishes PPQs from other PQs by the examples in (39).

- (39) a. A: Peter was not alone.
 B: Who was he with?
b. A: Peter was with someone.
 B: Who was he with?

(Onea, 2016:15)

The PQ in (39-b) is a PPQ of its preceding utterance, while the question in (39-a) only qualifies as a PQ, not a PPQ. This is surprising because the truth-conditions of A's utterance are the same for (39-a) and (39-b). Onea (2016) argues that in the process of

the composition of the meaning of A’s utterance in (39-b), the highlighted alternatives of B’s question are involved. In inquisitive semantics, indefinite expressions, such as *someone*, are by themselves inquisitive. Hence, the meaning of A’s utterance could be paraphrased as *Peter was with Lena or with Arne or with Nina or...*, which corresponds exactly to the highlighted alternatives of B’s question in (39-b). In (39-a), in contrast, no inquisitive expression is salient and no stage of the composition involves the alternatives highlighted by B’s question.

The last special case of PQs to consider is the LIKELY POTENTIAL QUESTION, as defined in Definition 29.

Definition 29 (Likely Potential Questions). *A PQ q licensed by some utterance U in some context c is a Likely Potential Question (LPQ) licensed by U in c , iff q is ranked highest on in the set of PQs licensed by U in c given a partial order \leq_s determined by salience.*

(Onea, 2016:15)

The definition of LPQs singles out the PQ that is most salient in the context, formally modeled by the partial salience order \leq_s for PQs. This question is considered to be most likely to arise. Consider the example utterance U in (40), and a subset of the available PQs in (40-a)–(40-c).

- (40) U : Somebody has a secret.
- a. Who has a secret?
 - b. What is the secret?
 - c. How many letters does that person’s name have?

All those questions are licensed by U in the context. However, they are not equally salient. They are ordered as they are presented, the question in (40-a) being the most salient and the question in (40-c) the least salient. If the context of (40) contained just those three questions, (40-a) would be the LPQ. It is no coincidence that this is a PPQ. According to Onea (2016), PPQs rather often turn out to be LPQs, which is not surprising given their

definition which requires PPQ to be made salient by U . Another example of LPQ that is not also a PPQ is a causal question, such as the one discussed above, repeated in (41).

- (41) a. My boss called me an idiot!
b. **LPQ:** Why did she do that?

The concept of LPQs is especially interesting for my analysis of clefts and my adapted discourse model, presented in section 10.3. The concept of the likelihood of a potential question, but also of other questions, will play a crucial role and will be covered in connection with the concept of *expectedness* (see section 10.2).

I will not go into more details as far as the formal properties of different kinds of PQs are concerned. For this thesis, it is sufficient to have a rather intuitive understanding of the different kinds of PQs. I will leave aside the fact that the definition of PQs allows a PQ such as (40-c), for instance, which is not at all likely to arise to the addressee given U . I view PQs as questions that arise to the addressee due to previous utterances, and I will be concerned with those PQs that are rather likely to arise to the addressee. For an extensive discussion of potential questions in discourse, I refer the reader to Onea (2013).

4.7 Summary

In this chapter, I provided the basic concepts on which I will build my adapted discourse model in section 10.3. Section 4.2 included a version of *context update* that involved speech acts, since my discourse model will be dealing with potential speech acts. Furthermore, Roberts' (2012) QUD framework was presented in section 4.3, and extended by Simons et al.' (2017) distinction of the *current question* (CQ) and the *discourse question* (DQ) in section 4.4. Crucially, I settled on a definition of the current question, based on Velleman et al. (2012), which is less restricted and requires the CQ to be provided by the context. Based on this definition of CQ, adapted principles of relevance in discourse and of accepting a discourse move were developed. The discourse effects were modeled via these principles with the CQ functioning as a link to the DQ, instead of making the CQ

itself require certain discourse constellations. This approach is compatible with all kinds of semantic approaches to questions in general and it is also compatible with approaches to focus which assume that focus depends on the CQ.

Moreover, I discussed accommodation of the CQ in section 4.5, which is important for this approach given that the CQ is entirely determined by the speaker. In general, for many approaches it is not obvious how the speaker and the hearer know what the CQ is. Section 4.5 provided an informal explanation for how the CQ is identified by both the speaker as well as the hearer. Thereby, it was also explained why it is not focus that determines the CQ. Focus is only an indicator of the CQ, but sometimes an ambiguous indicator.

Finally, the concept of *potential questions* was introduced, which are not yet included into the discourse model. This is the starting point for the adjustments of the discourse model that I will discuss in chapter 10. Here, I want to provide a small outlook on why the analysis of clefts might require an extension of the discourse model presented in this chapter.

The main problem lies in what counts as a relevant assertive discourse move. So far, only assertions that addressed an explicit or accommodated question or a sub-question thereof counted as relevant. It seems, however, that the *es*-cleft in German is much more flexible than a non-clefted sentence. In the discourse model presented above, example (22) from chapter 1, repeated in (42), causes a problem.

(42) Lena hat gestern auf der Party mit einem Typen₁ gesprochen. Die beiden haben viel gelacht und sich direkt für den nächsten Abend verabredet. Dann ist Lena glücklich nach Hause gefahren.

'Yesterday at the party, Lena talked to some guy₁. The two of them laughed a lot and they agreed to meet again the next evening. Then, Lena went home happily.'

- a. Es war Peter₁, mit dem sie gesprochen hat.
it was Peter₁ with whom she talked has
'It was Peter₁ she talked to.'

- b. ?Sie hat mit PETER₁ gesprochen.
 She has with PETER₁ talked
 ‘She talked to PETER₁.’

The current discourse model cannot predict (42-a) to be more acceptable than (42-b). According to the focus marking and the cleft structure, the *es*-cleft and the canonical sentence could have the same CQ: *Which guy did Lena talk to?*. Let us take the speaker’s perspective who decides to address this CQ. And whatever the DQ is in (42), it is the same for both sentences. Let us assume the DQ to be a very broad question, such as *What happened to Lena yesterday night?*. According to the definition of *Accepting an Assertive Discourse Move* (Definition 24), both sentences should be unacceptable since the CQ is not a sub-question of the DQ, it is actually a primary potential question of the first sentence. If we pick a different DQ, such as *Who did Lena talk to?*, the model predicts both sentences in (42) to be acceptable since this DQ is a superset of the CQ. Intuitively, the question that is addressed here seems rather irrelevant, and, nevertheless, it seems acceptable to address this question as long as it is addressed with a cleft.

This reveals three problems: (i) The notion of the DQ seems to be a rather vague one, as the DQ of example (42) was not unambiguously identifiable. This makes it difficult to test the predictions of the model. One might consider giving up the notion of a DQ. (ii) Whatever the CQ and the DQ are, the model does not predict a difference between a cleft and a canonical sentence as long as they are both focus congruent to CQ. (iii) The model cannot predict that there are relevant discourse moves that address PPQs or even other kinds of questions, apart from sub-questions of the DQ. It seems, however, that the cleft addresses a PPQ.

Therefore, I argue to replace the DQ, to take into account how exactly the assertion is formulated, clefted or unclefted, and most importantly to modify relevance with respect to different relations between the utterance and questions present in the discourse. In section 10.3, I will model relevance by assigning expectedness values to different kinds of possible questions that arise in a conversation, such as potential questions and other implicit questions.

One might want to argue that the unclefted sentence in (42-b) could, in principle, also address other CQs, when the focus projects. For the cleft, in contrast, this is impossible since it unambiguously marks the CQ. In case the CQ is indeed the one corresponding to the narrow focus and it needs to be accommodated, this is much easier for the cleft than for the canonical sentence in (42). Hence, the addressee might just be confused about the identity of the CQ of (42-b), which makes it less acceptable than (42-a). Indeed, the canonical sentence in the context of (42) improves slightly when it contains unambiguously marked narrow focus, as in (43-b).

- (43) a. Es war PETER, der Lena angesprochen hat.
 it was PETER who Lena approached has
 '*It was PETER who approached Lena.*'
- b. PETER hat Lena angesprochen.
 PETER has Lena approached
 '*PETER approached Lena.*'

I would still argue that the cleft in (43-a) is slightly more acceptable than the canonical version in (43-b) given the context in (42). However, the difference in acceptability between the two sentences decreases significantly. Those differences and a lot more data are discussed in section 10.1.

Part II

Analysis of Cleft Sentences

Chapter 5

Syntax (and Semantics) of Clefts

The syntactic analysis of clefts is still a matter of debate and has led to many different approaches. Already the definition of a cleft structure causes the first challenge. In section 5.1, I distinguish different kinds of structures that are considered clefts in German. In section 5.2, I present some challenges for the syntactic analysis of *es*-clefts in German. In section 5.3 – 5.6, I discuss different syntactic analyses of clefts that have been proposed.

5.1 The *es*-Cleft and Similar Structures

As explicated in chapter 1, I take sentences of the form in (1) as the standard German *es*-cleft, following Hedberg (2000).

(1)	Es	ist	Arne,	der lächelt.
	CLEFT PRONOUN +	COPULA +	CLEFT PIVOT +	CLEFT RELATIVE CLAUSE
CLEFT MATRIX CLAUSE				

'It is Arne who is smiling.'

However, there are several other constructions that are either considered kinds of clefts, structures that clefts derive from, or equivalents of the cleft. First of all, Hedberg (1990) and Gundel (1977), among others, argue that a cleft can also be realized without the relative clause, such as in (2).

- (2) Es ist Arne.
it is Arne
'It is Arne.'

This version of a cleft is called a TRUNCATED CLEFT (Hedberg, 1990). Hedberg (1990) and Gundel (1977) argue that this kind of cleft can only be used if the omitted relative clause is in the focus of attention of the speaker and the addressee, i.e. easily predictable from the context, as in (3-a) or, for German, (3-b). The cleft relative clause, in brackets, may be omitted.

- (3) a. I guess you're leaving for New York soon. – Yes, it's on Saturday (that I'm leaving). Gundel (1977:553)
- b. Ich glaube es hat eben an der Tür geklopft. – Ja, es war Arne, (der an der Tür geklopft hat).
I believe it has just at the door knocked yes it was Arne (who at the Tür geklopft hat).
door knocked has)
'I believe somebody just knocked at the door. – Yes, it was Arne (who knocked at the door).'

In examples (3-a) and (3-b), the content of the relative clauses is explicitly mentioned in the preceding sentences and is, thus, very easy to infer. Following this line of argumentation, these structures are commonly analyzed as elliptical. However, others disagree, arguing that these structures are not clefts. Mikkelsen (2007) argues that such structures are specifical copular constructions with a pronominal subject which are not elliptical. An observation that supports this analysis is that truncated clefts are well-formed not only with the pronoun *es* ('it') but also with the demonstrative pronoun *das* ('that/this'), which does not hold for full clefts.

- (4) Ich glaube es hat eben an der Tür geklopft. – Ja, das war Arne.
I believe it has just at the door knocked yes that/this was Arne
'I believe somebody just knocked at the door. – Yes, that/this was Arne.'

Hedberg (2000), in turn, argues that in English, the cleft pronoun *it* can, in general, be replaced by the demonstrative pronouns *this* or *that*. This would result in a structure like (5) in German.

- (5) Das/dies ist Arne, der lächelt.
That/this is Arne who smiles
'That/this is Arne who is smiling.'

The *das*-cleft in German, in its non-truncated version, needs a specific context, such as (6-a), in order to be syntactically interpreted as a cleft. In (6-b), the *das*-cleft is felicitous, and seems even more appropriate than the *es*-cleft (at least in colloquial German).

- (6) a. A: Anna ist gestürzt und Nina hat sie ausgelacht.
Anna is fallen and Nina has her made.fun
'Anna fell and Nina made fun of her.'
- b. B: Stimmt ja gar nicht. Das war doch Arne, der gelacht hat.
is.true PRT at.all not that was PRT Arne who laughed has
'That's not at all true. That was Arne who laughed.'

It seems that usually, the truncated cleft would still be preferred in this context. The only reason why we need the full cleft is because the elided material cannot be unambiguously identified. Without the relative clause, B's statement could be interpreted as either referring to Arne being the one who fell, or being the one who laughed.

There are two structures that a cleft has to be distinguished from: a copular construction with a restrictive relative clause, as in (7-a), and a copular construction with a non-restrictive relative clause, as in (8-a). Those are, on the surface level, similar to a cleft, but should not be analyzed as clefts.

- (7) Speaker is pointing towards a person:
- a. Das ist der Mann, der das Geld gestohlen hat.
that is the man who the money stolen has
'That is the man who stole the money.'
- b. Dieser Typ ist der Mann, der das Geld gestohlen hat.
that guy is the man who the money stolen has
'That guy is the man who stole the money.'

(8) Speaker arrives with a friend.

- a. Das ist Arne, den du sicher mögen wirst.
that is Arne whom you certainly like will
'This is Arne who you will certainly like.'
- b. Mein Freund hier ist Arne, den du sicher mögen wirst.
my friend here is Arne whom you certainly like will
'My friend here is Arne whom you will certainly like.'

Those structures are different from cleft structures to the extent that the pronoun can be replaced with a full DP as in (7-b) and (8-b), respectively. This is not possible in ‘real’ clefts. I will discuss the differences between the cleft and such structures in more detail in section 5.6. Here, I just present a test that can be used to identify a cleft and distinguish it from structures like (7-a). The NP in (7-a) can be left-dislocated together with the relative clause without changing the meaning of the sentence, as in (9-a). For the cleft in (1), this is not possible when the intended meaning is supposed to be preserved. When the cleft meaning is intended, (9-b) becomes unacceptable.

- (9) a. Der Mann, der das Geld gestohlen hat, ist das.
the man who the money stolen has is that
'That is the man who stole the money.'
- b. #Arne, der lächelt, ist es.
Arne who smiles is it
'Arne who is smiling is it.'

Having introduced the truncated cleft, structures that cannot be considered clefts, and what I assume the standard cleft structure, I will now present another variant of the cleft structure. Reeve (2011) claims that in English there is also a *there*-CLEFT, such as (10), which could be translated to German as (11).

(10) There is Arne who is smiling.

(11) Da ist Arne, der lächelt.
There is Arne who smiles
'There is Arne who is smiling.'

In German, this structure cannot have the same function as an *es*-cleft, though. It is either a presentational statement, as in (12), or a reminder, as in (13).

- (12) Guck mal, da ist Arne, der schon wieder gute Laune hat!
look PRT there is Arne who already again good mood has
'Look, there is Arne, who again is in a good mood!'
- (13) Da ist noch Arne, der auch mitkommen wollte.
there is in.addition Arne who also come.along wanted
'And there is Arne who also wanted to come along.'

The movement test for clefts shows that (12) is not a cleft.

- (14) Arne, der schon wieder gute Laune hat, ist da.
Arne who already again has good mood is there
'Arne who again is in a good mood is there.'

After the NP and the relative clause have undergone syntactic movement as in example (14), the sentence is still well-formed. Hence, it cannot be a cleft. Instead, it is a presentational statement with a non-restrictive relative clause. A restrictive relative clause would have the same effect. The *there*-cleft in (13), on the other hand, does not maintain well-formedness in the movement test, at least not under the reminding reading, which expresses to not forget about the fact that Arne also wanted to come along.

- (15) #Arne, der auch mitkommen wollte, ist da noch.
Arne who also come.along wanted is there in.addition
'And Arne who also wanted to come along is there.'

According to the movement test, the *there*-cleft in (13) could be a variant of the *es*-cleft in German, when used as a reminder. It seems that this cleft is still different from the 'normal' *es*-cleft in German, which does not have this kind of reminding function.

Besides the variants of the typical *es*-cleft, there are at least three constructions that are considered to be structurally related, or to have a similar function as a cleft: *wh*-clefts (16), definite pseudoclefts (17), and focus fronting (18) or contrastive focus.

Wh-clefts, as those in (16), are claimed to be both structurally similar to *it*-clefts (e.g. Akmajian, 1970), and as far as their discourse function is concerned. Prince (1978), in contrast, argues that the two structures differ slightly with respect to text coherence. *Wh*-clefts will be discussed in more detail in section 5.4.

- (16) a. Wen Laura am meisten mag, ist Juanca.
 who Laura the.most likes is Juanca
 ‘Who Laura likes most is Juanca.’
- b. Wer am schönsten singt, ist Laura.
 who the.most.beautiful sings is Laura
 ‘Who is singing most beautifully is Laura.’

Definite pseudoclefts in German, such as (17), contain the anaphoric element *derjenige* (or *diejenige*, *dasjenige*, etc., depending on gender and number).

- (17) Derjenige, der lächelt, ist Arne.
 The.one who smiles is Arne
 ‘The one who is smiling is Arne.’

The anaphoric element *derjenige* is composed of the definite article *der* and the demonstrative element *-jenige* (‘*that one*’). English definite pseudoclefts, in contrast, only contain a definite article and a pro-NP form *one*. There might be some cross-linguistic differences arising because of this (see De Veauh-Geiss et al., 2018b). Those scholars (e.g. Percus, 1997) who speak in favor of the syntactic similarity of *it*-clefts and definite pseudoclefts (e.g. Percus, 1997) argue that also semantic similarities between the two structures can easily be derived under this assumption. They translate uniqueness of the definite description into exhaustivity in the *it*-cleft. Definite description approaches will be discussed more thoroughly in section 5.3.

Finally, there is (contrastive) focus or focus fronting, as in (18), which on the surface level does not seem to resemble the *it*- or *es*-cleft, but which is frequently argued to have a similar function and even structure (e.g. by É. Kiss, 1998).

- (18) ARNE lächelt.
 ARNE smiles.
 ‘ARNE is smiling.’

É. Kiss (1998) derives the syntactic cleft representation from Hungarian preverbal focus, as in (19).

- (19) Mari PÉTER hívta fel.
Mari.NOM PÉTER.ACC call up
'It is PÉTER Mari called.' (É. Kiss, 1998:256)

In this example, the focused object *Pétert* is moved to the immediately preverbal position out of a lower syntactic position. The result is that *Pétert* is interpreted as an exhaustive focus. Assuming the same analysis for clefts, she predicts also the cleft to be exhaustive. This approach will be discussed in section 5.5. The information-structural implication of treating clefts and focus constructions in a parallel way will be discussed in section 8.6.

5.2 Challenges

Given the observed parallels of the cleft to very different structures and the various functions of clefts (introduced in chapter 1, and discussed in detail below in this chapter), there seems to be a mismatch between the structure of the cleft and its function. Structurally, it appears to be a copular sentence. Functionally, it seems to be a focus fronting construction. The syntactic analysis will ideally incorporate both aspects.

The first challenge in the syntactic analysis of a cleft is to incorporate the cleft relative clause, that does not always behave like a ‘normal’ relative clause, into the structure. It is neither obvious where the cleft relative adjoins, nor which element it modifies semantically, nor whether it undergoes syntactic movement. The movement test for clefts showed that the cleft pivot cannot be moved together with the relative clause. Other relative clauses in German, however, can be moved together with their head noun phrase, as illustrated in (20).

- (20) a. Ich sehe die Frau, die auf den Baum klettert.
I see the woman who onto the tree climbs
'I see the woman who is climbing the tree.'

- b. Die Frau, die auf den Baum klettert, sehe ich.
 the woman who onto the tree climbs see I.
'The woman who is climbing the tree, I see.'

In (20-b), the whole DP including the restrictive relative clause is moved to the prefield.

In a German *es*-cleft, only the pivot without the relative clause can be moved to the prefield, as in (21).

- (21) Arne ist es, der lächelt.
 Arne is it who smiles
'It is Arne who is smiling.'

Hence, any syntactic theory of *es*-clefts in German should account for the fact that the cleft relative clause does not behave like a restrictive or non-restrictive relative clause usually would.

The second challenge arises from the fact that the cleft superficially looks like a copular sentence, be it specifical, predicational, or presentational, but semantically and pragmatically corresponds to a non-cleft sentence with focus on the former pivot, possibly contrastive or exhaustive focus. Hence, a syntactic analysis of clefts needs to, on the one hand, account for the presence of the copula and explain its function, and, on the other hand, it needs to account for focus assignment in clefts.

A third challenge is determining the status of the cleft pronoun. This section showed that the cleft pronoun *es* ('it') can, at least in certain contexts, be replaced by an anaphoric element like *das* ('that') or *da* ('there'), which would indicate that the cleft pronoun must be referential in those cases. Those approaches that compare the cleft to focus structures, which lack the cleft pronoun altogether, must assume the cleft pronoun to be expletive. Here, the challenge is to model how and why the expletive pronoun enters the syntactic structure of the cleft. Furthermore, those approaches need to explain why this expletive element can sometimes be replaced with an anaphoric element.

Finally, different syntactic analyses favor or even include certain semantic features, which are themselves debatable. Those theories that take clefts to have an underlying focus

structure need to explain why the cleft seems to have a stronger existence presupposition than plain focus structures have. Percus (1997), among others, claims that focus does not give rise to an existence presupposition. Seuren (1985) and others argue that it is rather contrastive focus that corresponds to clefts than plain focus. They argue that contrastive focus does have an existence presupposition, and it also shows other parallels to clefts.

5.3 Definite Description Approach

There are several approaches that propose to derive the *it*-cleft from a complex definite description, for instance the DEFINITE PSEUDOCLEFT in (22).

- (22) Derjenige, der lächelt, ist Arne.
 The.one who smiles is Arne
 ‘*The one who is smiling is Arne.*’

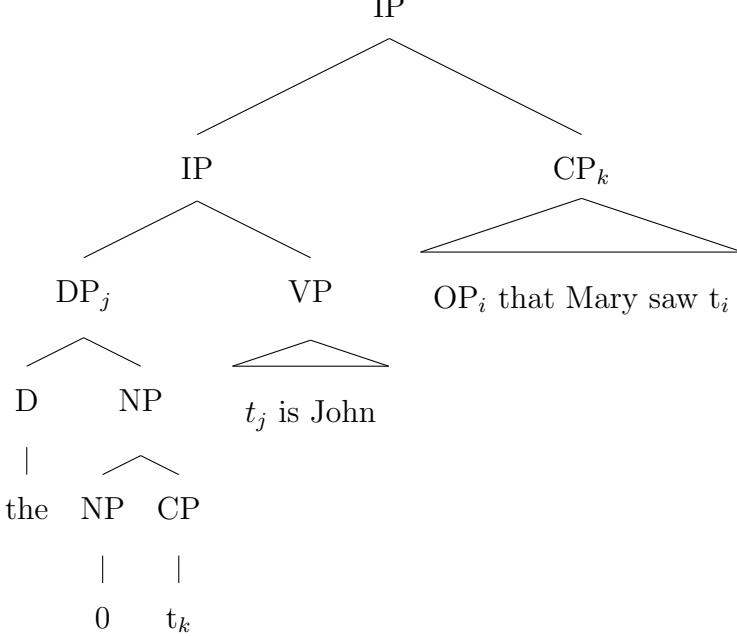
Percus (1997) argues that the *it*-cleft contains a hidden definite description like *the one* or *derjenige* as in (22). This definite description consists of a definite determiner and a null head as indicated in (23-c). Percus assumes the cleft in (23-a) to derive from (23-c), and to be structurally indistinguishable from the definite pseudocleft in (23-b).

- (23) a. It is John that Mary saw.
 b. The one that Mary saw is John.
 c. The 0 that Mary saw is John. Percus (1997:340)

In order to derive the surface structure of the cleft from the underlying structure in (23-c), Percus (1997) assumes two more derivational steps: the extraposition of the relative clause and a spell-out rule for the hidden definite description. In the extraposition step, the cleft relative clause is extraposed to the sentence-final position. This step is necessary since the relative clause surfaces in a different position in the cleft compared to the definite description. The extraposition is realized by adjoining the relative clause to IP, as shown in (24), indicated by the index *k*.²⁴

²⁴ OP_i is an operator that models the abstraction over t_i for integrating the relative clause.

(24)



This approach has the advantage that it correctly predicts that the cleft pivot, *John*, and the relative clause cannot be moved together since the two phrases do not form one constituent in the proposed syntactic structure.

However, the extraposition is problematic. In languages that expand syntactic trees on the left side, such as English or German, it is rather uncommon to assume movement to the right. In more recent syntactic approaches, moreover, every movement needs a semantic feature that motivates this movement. Percus (1997) does not assume a semantic difference between the definite pseudocleft and the *it*-cleft, but still the former does not involve movement and the latter does. Hence, the cleft should involve some semantic feature that distinguishes it from the definite pseudocleft. It is not a trivial question what this semantic feature could be.

One explanation in favor of extraposition could be that the cleft relative clause is moved into a less prominent position by extraposition. Gundel (1985), among others, argues that the sentence final position frequently contains old, already ‘activated’ topics. Given that the cleft relative clause contains the presupposed part (at least in the focus-background cleft), it seems reasonable that it is moved to the sentence final position. This, by the way, also puts the cleft pivot in a more prominent position. If this is a correct explanation for extraposition, the definite pseudocleft, in contrast to the *it*-cleft, would be predicted to

be information-structurally different since it does not involve extraposition. De Veaugé-Geiss et al. (2018b) discuss information-structural differences between German *es*-clefts and definite pseudoclefts involving *derjenige*. They claim that such definite pseudoclefts are odd at the beginning of a story, while the *es*-cleft is appropriate in an out-of-the-blue context. On the one hand, this supports an analysis which assigns different information-structural properties to *es*-clefts and definite pseudoclefts in German. On the other hand, this difference might exist in German just because of the anaphoric nature of the pronoun *derjenige* used in the definite pseudocleft, as De Veaugé-Geiss et al. (2018b) propose.

According to Percus (1997), the DP $[DP \text{ the } 0 t_k]$ in (24) is spelled out as *it* after extraposing the relative clause. This spell-out rule is rather stipulative. Percus does not have an independent motivation for it except for the observed word order in clefts. Furthermore, given the variants of the cleft using *that* or *there* as the cleft pronoun, one would have to adjust the spell-out rule in order to account for those variants. However, this approach would then struggle to predict when the underlying DP is spelled out as *it*, and when it is spelled out as *that* or *there*. As we saw above, the variants involving *that* or *there* are only well-formed under certain conditions depending on the context. Hence, the choice of the correct pronoun cannot be made on the level of syntax.

The parallel analysis of definite descriptions and clefts does not only have syntactic implications but also semantic ones. The cleft will, accordingly, take over semantic features of definite descriptions, such as uniqueness and existence. Percus (1997) presents some semantic arguments that support this analysis of *it*-clefts, which I will present in the following. First of all, he argues that his approach predicts the definite pseudocleft and the *it*-cleft to have an existence presupposition while the non-clefted canonical version of the sentence does not. The reason is that the former two are derived from an underlying definite description, which has been shown to have an existence presupposition (c.f. Strawson, 1950). Therefore, it is plausible that also clefts have an existence presupposition. Since the canonical structure is not derived from a definite description, no existence presupposition is predicted. This difference can be illustrated by using *nobody* as the pivot of the cleft or as the focus in the canonical sentence, as in (25). This should cause

a contradiction between a potential existence presupposition in (25-d) and the content of the sentences in (a)-(c).

- (25) a. Mary saw [NObody]_F.
- b. #The one that Mary saw was [NObody]_F.
- c. #It was [NObody]_F that Mary saw.
- d. Existence presupposition: There is somebody that Mary saw.

However, the use of *nobody* only results in unacceptability of the cleft and the definite pseudocleft while the non-cleft with narrow focus on *nobody* is acceptable. Percus (1997) concludes that only the cleft and pseudocleft have an existence presupposition. Hence, (25-a) does not yield a contradiction. If the cleft and the pseudocleft inherit the existence presupposition from their underlying definite description, this difference are predicted. The existence presupposition is discussed in more detail in chapter 6.

Moreover, Percus (1997) argues that the definite pseudocleft and the *it*-cleft inherit uniqueness from the definite description, which is realized as an exhaustivity presupposition associated with the focus, as illustrated in (26-d). Hence, the cleft is incompatible with focus particles that require other alternatives to hold because that would contradict exhaustivity. The fact that clefts with the focus particles *even*, *also*, and *only* in (26-b) and (26-c) are infelicitous supports Percus' view. The canonical sentence, in contrast, is acceptable with such focus particles, as (26-a) indicates.

- (26) a. It was even/also/only the case that [JOHN]_F saw Mary.
- b. ?It was even/also/only the case that it was [JOHN]_F who saw Mary.
- c. ?It was even/also/only the case that the one who saw Mary was [JOHN]_F.
- d. Exhaustivity presupposition: Only John saw Mary.

Percus' (1997) analysis of clefts would account for these observations by assuming the hidden definite description in clefts. The attested presuppositions of the cleft are simply the existence and uniqueness (=exhaustivity) presupposition of the definite article, for which they have been frequently assumed anyway (see again Russell, 1905, 1957). However, as

I will show later in this chapter, exhaustivity is not necessarily defined in the same way as uniqueness. Furthermore, even definite descriptions are claimed to not always involve uniqueness (e.g. Szabó, 2003).

Pavlović (2019) presents many counter-examples to Percus' judgments on focus particles in the cleft matrix clause. She presents corpus evidence of clefts with the focus particles *auch* ('also'), as in (27), *vor allem* ('mainly'), as in (28), *nur* ('only'), as in (29), and *sogar* ('even'), as in (30) (taken from Pavlović, 2019).

- (27) Es ist **auch** ihre Perspektivlosigkeit, die viele Jugendliche zur Flasche greifen lässt.

'It is also their lack of perspective that makes many teenagers reach for the bottle.'

(RHZ04/APR.20135 Rhein-Zeitung, 23.04.2004; Jugend braucht mehr Chancen)

Pavlović (2019:82)

- (28) Es ist **vor allem** das Wetter, das uns bis jetzt einen Strich durch die Rechnung macht.

'It is mainly the weather that has messed up our plans so far.'

(NUZ06/JUN.00081 Nürnberger Zeitung, 01.06.2006; Umsatzzrückgang beim Einzelhandel im April - Wetter verregnete das Geschäft) Pavlović (2019:82)

- (29) Es ist **nur** die in Udine gebotene Leistung, die momentan so nachdenklich stimmt.

'It is only the performance presented in Udine that makes one thoughtful right now.'

(A98/OKT.64464 St. Galler Tagblatt, 13.10.1998, Ressort: TB-SPO (Abk.); "Spiritus retour") Pavlović (2019:86)

- (30) Es ist **sogar** Sanftmut, die durchdrückt, als im Hintergrund das Töchterchen zu weinen beginnt und Rangelow tröstend eingreift.

'It is even gentleness that carries through when the little daughter starts to cry in the background and Rangelow steps in soothingly.'

(NZZ13/DEZ.02104 Neue Zürcher Zeitung, 16.12.2013, S. 33; Dank Vertrauen anders) Pavlović (2019:94)

Those examples are problematic for Percus' (1997) approach if the underlying definite description indeed requires the cleft to have a uniqueness presupposition. On the other hand, it is not obvious where in Percus' structure the focus particles are generated. Beaver and Clark (2008), for instance, assume that focus particles are always sentential modifiers. In (31), a problem with the uniqueness presupposition only occurs when *John* is focused.

- (31) Even [the 0 that Mary saw is John].

Since *even* is a focus-sensitive particle (as discussed in section 3.2), it associates with the focus in the sentence it dominates. Hence, it is crucial to determine the focus in (31) in order to test if there is a conflict between a potential uniqueness presupposition of the cleft and the felicity conditions for *even*. If the focus particle is generated in the relative clause, as in (32), the pivot *John* is not in the scope of even.

- (32) the 0 [even [that Mary saw]] is John

Here, there is no conflict with the uniqueness presupposition. If we assumed a similar structure for German *es*-clefts in combination with additive focus particles, the acceptability of such clefts would be captured. The cleft sentence in (33-a) would have the reading in (33-b).

- (33) a. Es ist sogar John, den Mary gesehen hat.
it is even John that Mary seen has
'It is even John that Mary saw.'

- b. *Intended reading assuming the underlying structure in (32):*

For the unique individual John, it is unexpected that Mary would see him, in contrast to other things that could have happened to John.

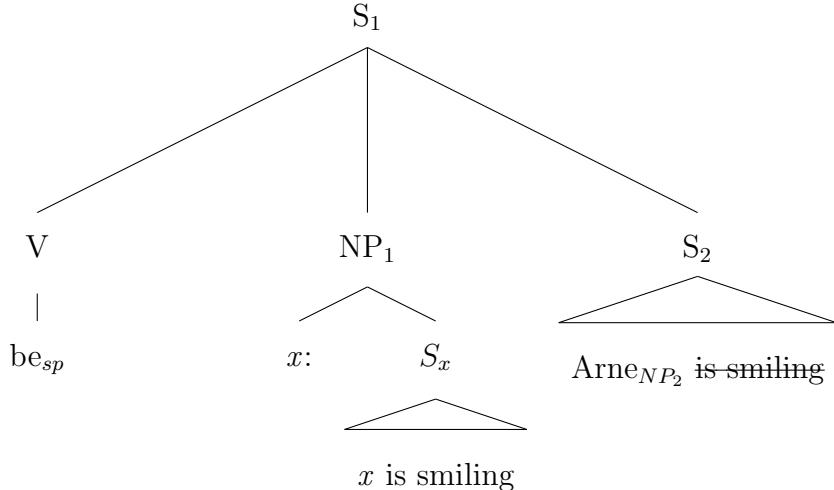
This reading is only available for topic comment clefts. As long as *John* is focused, the sentence in (33-a) predominantly expresses that it is unexpected that Mary saw John, rather than other people she could have seen.

Apart from Percus (1997), also others have analyzed clefts as definite descriptions. Gundel (1985) assumes that the cleft relative by itself is a definite referring expression. According to Gundel et al.'s (1993) givenness hierarchy, which will be presented in section 8.2, this has implications on the activation status of the antecedent of the relative clause in the context. Moreover, Seuren (1985) relates the cleft to a definite description, but in a very different way. He assumes the cleft to have an underlying structure that does not correspond to any surface structure, neither that of the cleft nor that of any other corresponding sentence. This underlying structure expresses the relative clause as a definite description of the form *the x such that x Vs.* As opposed to Gundel (1985), he assumes that the reference of this definite description is difficult to resolve in the discourse. He requires for the referent of the definite description, expressed by the relative clause, to involve ambiguity in the preceding context, which makes it harder to identify the referent. Furthermore, even though his syntactic analysis involves the definite description, he rather treats contrastive focus structures and cleft sentences as parallel structures. In the following, the syntactic-semantic analysis of Seuren (1985) will be presented.

Seuren (1985) proposes that the cleft raises an implicit question and answers it at the same time. He argues that the focused constituent in the cleft typically answers the question that is implicit in the topic constituent, which he takes to be the cleft relative clause (Seuren, 1985:297). The first step of his analysis is concerned with determining which version of the verb *be* is found in the cleft matrix clause. He distinguishes it from locative *be*, identity denoting *be*, and predicative *be*. Instead, he assumes it to be what he calls specifying *be* (*be_{sp}*), which selects two arguments: a question-raising NP and a question-answering sentence. The surface structure of the cleft is derived from an underlying structure based on *be_{sp}* and its two arguments. For example, the cleft in (34-a) is assumed to have the underlying structure in (34-b).

- (34) a. It is Arne who is smiling.

b.



The question-raising NP₁ is roughly of the form *the x such that x is smiling*, which raises the question *What x is such that x is smiling?*. Seuren (1985) describes this NP as a property which is insufficient in the context to pick out a specific referent unambiguously and, thus, gives rise to the question who this referent is. In other words, what Seuren (1985) proposes is that NP₁ is a definite description that does not unambiguously refer, which, in turn, motivates the question about its referent. The question-answering sentence for (34-a) is represented in S₂ (*Arne is smiling*). Seuren additionally assumes deletion under identity between S_x and S₂, which results in identical material in two phrases of the same type to be deleted. Hence, S₂ ends up containing just the NP₂ *Arne*. Material in the question-answering sentence that is identical to material in S_x, i.e. *is smiling*, is deleted. Deletion under identity is motivated for other phenomena independent of clefts (Seuren, 1985).

In order to get the word order correctly, Seuren (1985) assumes two more steps, which are also motivated independently of clefts for other examples: (i) NP₂ is incorporated by V, which moves it in between *be_{sp}* and NP₁ on the surface structure, and (ii) *it*-insertion in the subject position whereby NP₁ becomes a *wh*-clause or a *that*-clause.

Seuren (1985) argues that not only the surface cleft structure but also an unclefted contrastive surface structure may have the underlying structure containing *be_{sp}*. According to his assumptions, the unclefted sentence in (35) is derived from the same underlying structure in (34-b), which was assumed for the cleft.

- (35) ARNE is smiling.

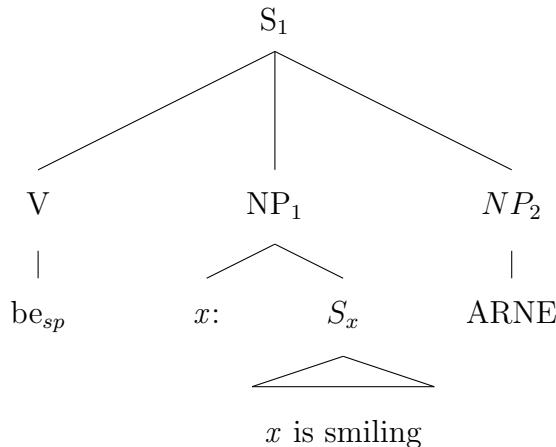
In order to achieve this result, Seuren (1985) uses the optional rule of Predicate Lowering (PL), which he again motivates independently of the cleft. PL is introduced for predicates of the form [*be* + adjective + sentence argument], which turn into sentential adverbials modifying the former sentence argument when they are lowered. In the process of this lowering, they lose their predicative status and the remaining material in the VP is deleted. Example (36) illustrates how the predicate *predictable* is lowered into its sentential argument where it becomes the sentence adverbial *predictably*.

- (36) a. It was predictable that Arne would be smiling.
 b. **Predicate Lowering:** ~~It was~~ Predictably, Arne was be smiling.

In (36-b), the verb *predictable* has lost its predicative status after lowering. Accordingly, what remains in the VP (*it was*) is deleted. Seuren (1985) points out that the transformation from (36-a) to (36-b) usually involves semantic differences between the two as well. While (36-b) entails that Arne actually smiled, (36-a) does not. Seuren incorporates those differences into the lexical entries of *predictable* and *predictably*.

Seuren (1985) treats the transformation of the underlying cleft structure into the surface canonical structure analogously, as indicated in the following tree structure.

- (37)



The intonation of the unclefted sentence determines which part of the sentence will be identified as NP₁ and which part will correspond to S₂. The focused part is now the former

S_2 , which is reduced to just NP_2 (*ARNE*), and the other parts are the backgrounded elements, which are represented identically to the cleft in (34-b). Analogously to the derivation of the cleft surface structure, NP_2 is first incorporated into V . Then, NP_2 is lowered into the sentential complement replacing x , and be_{sp} is deleted (by PL). Seuren (1985) claims that NP_2 receives a strong rise-fall accent in the process of PL. To my knowledge, he does not motivate where this accent originates, but given this assumption, he concludes that those canonical sentences that are derived from an underlying cleft structure always involve a contrastive focus.

Concluding, Seuren (1985) treats the cleft and the contrastive canonical as equivalents, instead of the cleft and the plain canonical sentence. In both these constructions, he assumes the sentence S_x to be a definite description that, in a way, represents an unanswered question, similar to the functional representation of questions, mentioned in section 2.2. However, it is not transparent how the focused NP receives the contrastive accent in the derivation of its structure. Apart from that, this approach is interesting for the analysis of clefts in discourse (the main purpose of this thesis), because it relates the underlying structure to a question. Subsection 8.6 will elaborate on Seuren's analysis shedding more light on its relation to discourse.

5.4 *Wh*-cleft Approach

Akmajian (1970) formulates an argument against the definite description approach, which will be presented in this section. He takes *it*-clefts to be syntactically derived from *wh*-clefts. According to his assumptions, the *it*-cleft in (38-b) derives from the *wh*-cleft in (38-a), and (39-b) derives from (39-a).

- (38) a. Who Laura likes most is Juanca.

- b. It is Juanca who Laura likes most.

- (39) a. Who is singing most beautifully is Laura.

- b. It is Laura who is singing most beautifully.

Akmajian (1970) argues against the definite description approach by presenting *it*-clefts for which it is not possible to find a definite description equivalent that is grammatical, as in (40).

- (40) a. It was to Boston that I went.
b. *The place I went was to Boston.
c. *The place I went to was to Boston.

While the *it*-cleft in (40-a) is grammatical with a prepositional pivot, such as *to Boston*, both possible definite description equivalents, (40-b) and (40-c), are ungrammatical when the prepositional phrase appears postverbally. The only grammatical definite description would be the one in (41-a), where the postverbal component is a noun phrase instead of a prepositional phrase.

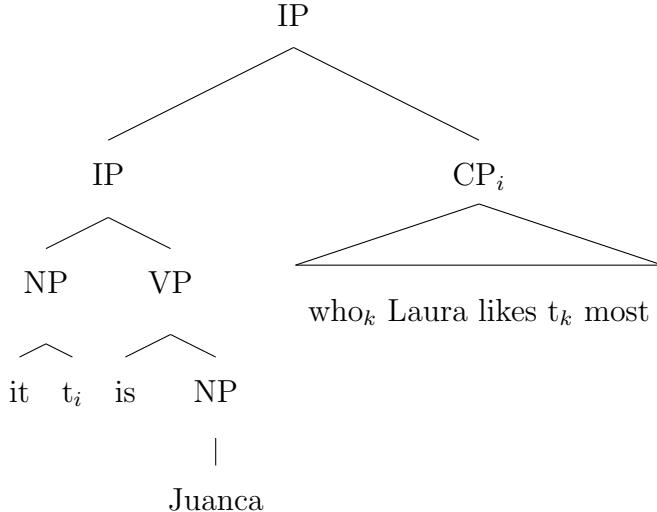
- (41) a. The place that I went to was Boston.
b. It was Boston that I went to.

However, the definite definite description in (41-a) is not structurally equivalent to (40-a), but to (41-b), instead. The *wh*-cleft (42), in contrast, does constitute a good equivalent to (40-a).

- (42) Where I went was to Boston.

Hence, Akmajian (1970) argues that *it*-clefts cannot derive from copular sentences including a definite description but rather derive from *wh*-clefts. He analyzes *wh*-clauses in *wh*-clefts as headless clauses. In order to derive the correct word order for the *it*-cleft or *es*-cleft, he posits the CLEFT EXTRAPOSITION RULE (Akmajian, 1970:166), which right-adjoins the *wh*-clause to the sentence (to the IP to be precise). Translated to a more recent syntactic framework, Akmajian (1970:166) proposes a structure like (43) for (38-b).

(43)



The *wh*-clause is moved to the right by cleft extraposition leaving only the pronoun *it* in subject position. Just like in Percus' (1997) approach, the extraposition rule is stipulated with no other obvious motivation than getting the correct word order.

In German a cleft like (40-a) is not well-formed, as (44-a) and (44-b) indicate.

- (44) a. *Es war nach Boston, das ich gegangen bin.
 it was to Boston that I gone am
 ‘It was to Boston that I went.’
- b. ?Es war nach Boston, wohin ich gegangen bin.
 it was to Boston whereto I gone am
 ‘It was to Boston where I went.’
- c. Es war Boston, wohin ich gegangen bin.
 it was Boston whereto I gone am
 ‘It was Boston where I went to.’

Both the cleft with the *d*-relative pronoun *das* (‘that’), as well as the one with a *wh*-relative pronoun are ungrammatical in German. The cleft is only well formed when the NP *Boston* is clefted without the preposition, as in (44-c). Hence, such a cleft does not provide a counter-example to the definite description approach for clefts in German. To my knowledge, there is no example of a German *es*-cleft that can only be translated into a *wh*-cleft but not into a definite description. Thus, Akmajian’s (1970) argument does, until further evidence, not apply to German clefts.

Prince (1978) presents some use-conditional differences of *it*-clefts and *wh*-clefts, which can be taken as counter-evidence against the assumption that *it*-clefts derive from *wh*-clefts. Example (45-a) and (45-b), which are based on the corresponding English version in Prince (1978:894), show that in some contexts the *es*-cleft is appropriate, but not the *wh*-cleft.

(45) A towards B (eating chocolates): Aren't those good?...

- a. Es war nur meine unglaubliche Selbstbeherrschung, die mich davon abgehalten hat gestern schon alle aufzuessen.
It was only my unbelievable self-control that me of.that kept has yesterday already all eat.up
It was only my unbelievable self-control that kept me from eating them all yesterday already.'
- b. ?Was mich davon abgehalten hat gestern schon alle aufzuessen, war nur meine unglaubliche Selbstbeherrschung.
What me of.that kept has yesterday already all eat.up was only my unbelievable self-control
'What kept me from eating them all yesterday already was only my unbelievable self-control.'

Example (45-b) with the *wh*-cleft is not entirely unacceptable, but is at least degraded compared to the *es*-cleft in (45-a). More specifically, the *wh*-cleft makes A's statement less coherent compared to the *es*-cleft in this context.

In example (46), which is also adapted from examples in Prince (1978), the opposite is the case.

(46) At the beginning of an "Intro to Syntax" class:

- a. Was wir heute besprechen wollen, ist die Syntax von Spaltsätzen.
what we today talk.about want is the syntax of clefts
'What we are going to look at today is the syntax of clefts.'
- b. #Es ist die Syntax von Spaltsätzen, die wir heute besprechen wollen.
it is the syntax of clefts that we today talk.about want
'It is the syntax of clefts that we are going to look at today.'

Here, the *es*-cleft is degraded compared to the *wh*-cleft. The teacher might utter the *wh*-cleft in (46-a) towards the students at the beginning of the class. Uttering the *it*-cleft in (46-b) in that context would be odd, although not completely infelicitous. Given these differences between the *es*-cleft and the *wh*-cleft it seems less plausible that one is derived from the other. If it is, one would have to explain how these differences arise in the process of the derivation, which I will get back to below.

Prince (1978) argues that these use-conditional differences arise due to the status of the existence presupposition in the respective cleft structure. According to her, the *wh*-cleft requires this presupposition to be on the addressee's mind, while the presupposition must only be a 'known fact' for the *it*-cleft. In particular, she argues that by using an *it*-cleft the speaker marks the presupposition as a known fact for herself, but not necessarily for the addressee. For the *wh*-cleft, the existence presupposition must be on the addressee's mind. In example (45), the fact that something kept A from eating all of the chocolates yesterday is not on B's mind in this context. It is, however, a known fact since there are still some chocolates left for B to eat. Accordingly, Prince correctly predicts the *es*-cleft to be more acceptable than the *wh*-cleft in this context.

Prince (1978) also correctly predicts why (46-a) is appropriate at the beginning of a class given that students usually wonder what the topic of the class will be. Hence, the fact that they will look at something today can be assumed to be on their minds. Prince's approach, however, cannot explain why (46-b) is not appropriate. The moment the teacher starts speaking, it seems to be a known fact that they will look at something today. Hence, the cleft should also be appropriate. Nevertheless, example (45) and (46) show that, contra Akmajian (1970), *it*-clefts and *wh*-clefts cannot be treated as equivalents. I will discuss those examples in more detail in chapter 8.

One could still argue that even though the two clefts are not used in the same situations, one is still derived from the other. However, one would have to motivate the movement of the relative clause in the case of the *it*-cleft, given that the movement has information-structural implications. For example, it could be motivated, as mentioned for

extraposition above, by assuming that the content is backgrounded, and is, thus, moved to the sentence-final position (c.f. Gundel, 1985).

5.5 Focus Approach

An approach that differs fundamentally from the approaches presented so far is what I call the FOCUS APPROACH. It differs insofar from the other approaches that it does not consider *it*-clefts or *es*-clefts to be derived from a copular construction. The cleft pronoun and the copula just arise due to syntactic well-formedness conditions, and they have no semantic content.

In the current research, it is widely accepted that clefts mark focus (e.g. De Veaugh-Geiss et al., 2015; Tönnis et al., 2018). Some assume that it is identificational focus (É. Kiss, 1998), or contrastive focus (Seuren, 1985). The pivot of the cleft in (47-a) is considered to represent narrow focus just as intonation on the subject in the canonical sentence in (47-b) does.

- (47) a. Es ist Arne, der lächelt.
It is Arne who is smiling.
b. ARNE lächelt.
ARNE is smiling.

É. Kiss (1998) makes it more specific by saying that the cleft does not just correspond to focus but to IDENTIFICATIONAL FOCUS, which she distinguishes from INFORMATIONAL FOCUS. This distinction is inspired by Hungarian data, which, according to É. Kiss (1998), shows two focus positions. Informational focus is marked in-situ, as in (48), while identificational focus involves movement of the focused element to the immediately preverbal position, as in (49).

- (48) Mari fel hívta PÉTERT.
Mari.NOM up called PÉTER.ACC
'Mari called PÉTER.'

- (49) Mari PÉTER hívta fel.
 Mari.NOM PÉTER.ACC call up
 ‘It is PÉTER Mari called.’ (É. Kiss, 1998:256)

For expressing identificational focus, the focused object *Pétert* is moved to the preverbal position immediately preceding the verb *hívta* (‘called’). É. Kiss (1998) takes Hungarian preverbal focus to be parallel to cleft constructions as indicated by the translation of (49). According to her, English realizes identificational focus by an *it*-cleft. Besides structural definition, identificational focus has a semantic/pragmatic effect as well, which É. Kiss describes as follows.

An identificational focus represents a subset of the set of contextually or situationally given elements for which the predicate phrase can potentially hold; it is identified as the exhaustive subset of this set for which the predicate phrase actually holds. (É. Kiss, 1998:245)

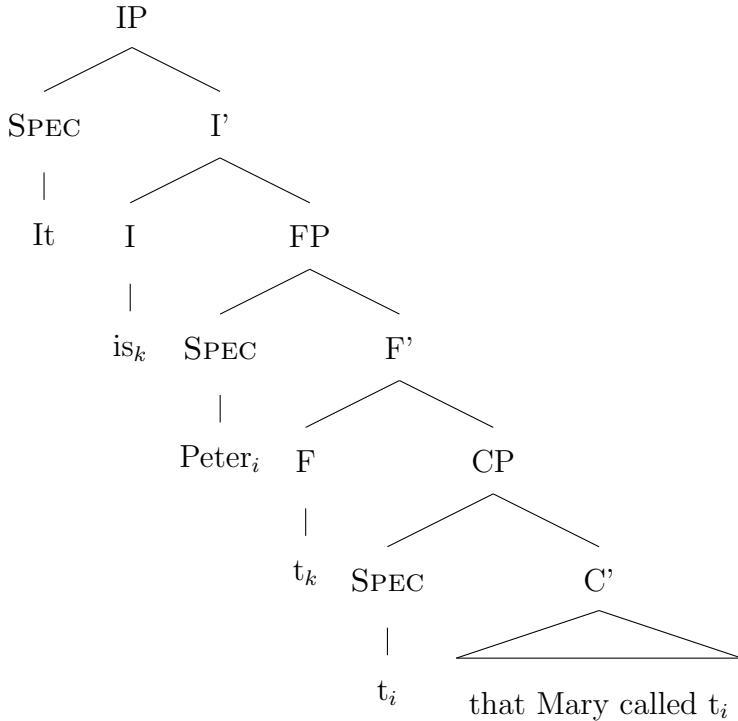
In the case of (49), this implies that the preverbal focus, which expresses identificational focus in Hungarian, provides the exhaustive answer to the question *Who did Mari call?*. The same effect is predicted for cleft constructions. Plain information focus as in (48), in contrast, does not provide an exhaustive answer in Hungarian.

Syntactically, identificational focus involves movement into the specifier of a focus projection. Information focus, on the other hand, does not involve a specific kind of movement. This explains the difference between (48) and (49), where the focused element *Pétert* stays in-situ in the former, and is moved to a higher projection in the latter.

For an English *it*-cleft, as in (50), É. Kiss (1998) proposes the syntactic tree structure in (51). This structure translates to German *es*-clefts without any crucial changes.

- (50) It is Peter that Mary called.

(51)



This structure is only slightly different from the structure É. Kiss (1998) would assume for Hungarian preverbal focus, as in (49), namely with respect to what the head of the focus phrase selects as its complement. While F selects a CP for the English or German cleft, it selects a VP for preverbal focus in Hungarian. This structure explicates the assumed parallel between preverbal focus and the cleft structure. The cleft pivot corresponds to the focused constituent in Hungarian preverbal focus. Like the focused constituent in (49), the pivot in (50) moved from a lower position inside the CP to the higher specifier of FP. The structure is assumed to be derived from the canonical informational focus sentence, as in (52), in which the focused constituent is in the lower position.

(52) Mary called PETER_F.

Given that the F head is empty, there is V-to-F-movement. In Hungarian preverbal focus, the verb moves into this F head. In English, however, this movement is blocked by the relative pronoun *that*. Therefore, É. Kiss (1998) claims that F is filled by expletive *be*. Furthermore, she assumes the cleft pronoun to be an expletive element because the specifier of IP must be filled. These two expletive elements can be interpreted as a ‘dummy’ verb and ‘dummy’ subject, which are used to not leave required positions empty

(following the Extended Projection Principle by Chomsky, 1982:10). É. Kiss needs to assume those elements to be expletive since, otherwise, the parallel to preverbal focus, which does not contain those elements, cannot be maintained. Hence, those examples in which the cleft pronoun can be replaced with an anaphoric pronoun, such as *das* ('that'), are a challenge for this approach.

With respect to the semantics, É. Kiss (1998) analyzes identificational focus as an operator which takes the complement of F as its scope. This means that the cleft relative clause is the predicate with respect to which the pivot is exhaustified, in example (51) with respect to *{Mary called x}*.

É. Kiss' (1998) approach runs into problems once the cleft has a different function than expressing identificational focus. Hedberg (2000), for example, argues that the cleft can have a topical instead of a focal function. She claims that it can even introduce a new topic, as indicated in her example repeated in (53).

- (53) It was the Greeks who first made wine around 1500 B.C. They then took this unique art to all the corners of the ancient world, including Italy, Spain, Russia, and, in about 600 BC, France.

(Liquor menu from 'It's Greek to Me' restaurant, Minneapolis)

(Hedberg, 2000:915)

In this case, É. Kiss' (1998) analysis is problematic since, in her approach, the information-structural function of the cleft is determined by the syntax via the focus phrase. A topical pivot, however, cannot be realized in the specifier of that focus phrase. One would have to assume a different syntactic structure for the topic comment cleft.

5.6 Referential Pronoun Approach

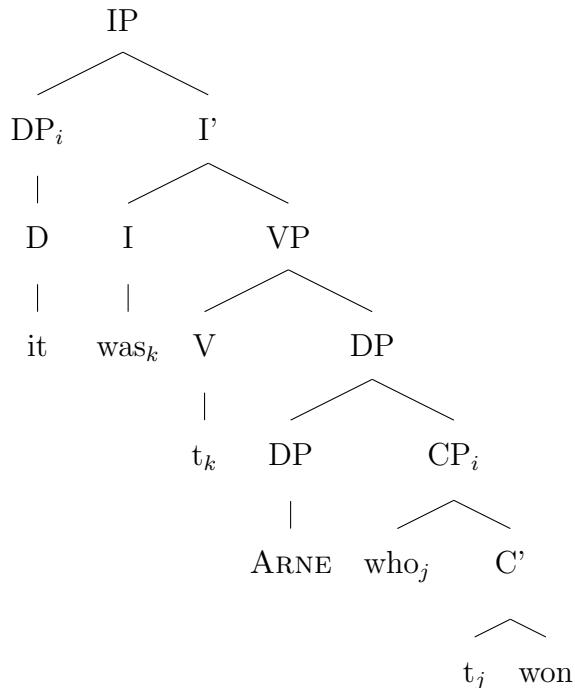
In the following, I present an approach that is based on the assumption that the cleft pronoun is referential. This contrasts with the focus approach, in which it is essential

that the cleft pronoun and the copula are expletive. Proponents of the REFERENTIAL PRONOUN APPROACH have shown that the cleft pronoun *it* can be replaced by pronouns that are usually not used as expletive pronouns. Recall example (5), repeated in (54).

- (54) Das/dies ist Arne, der lächelt.
 That/this is Arne who smiles
 ‘That/this is Arne who is smiling.’

Based on this observation, Hedberg (2000) takes the cleft pronoun and the cleft relative clause to function like a discontinuous definite description. The cleft pronoun corresponds to the definite article, and is referential in that sense. She assumes the pronoun *it* to be an allomorph of *the*, which occurs whenever the descriptive NP does not immediately follow the D head. It is the intransitive version of the definite article *the*, representing just the indexical component of a DP, and lacking the descriptive component. Hence, the cleft pronoun just refers in the sense that it relates the descriptive content of the discontinuous DP, i.e. the content of the relative clause, to the previous context. Based on these assumptions, Hedberg (2000) suggests the syntactic structure in (55) for an English *it*-cleft.

(55)



The cleft relative is adjoined to the pivot by extraposition, but semantically co-indexed with the cleft pronoun *it*. In the other extraposition approaches presented here (definite description approach and *wh*-approach), the cleft relative was adjoined higher – on the level of IP – than it is the case here. In Hedberg’s analysis, however, the pivot and the cleft relative clause form a syntactic constituent, and the cleft pronoun and the relative clause form a semantic unit. This approach struggles to account for the fact that the pivot cannot be moved together with the relative clause. In (55), however, the pivot and the relative clause form a constituent (*Arne who won*), and nothing in this approach speaks against moving this constituent. The semantic co-indexation captures the concept of the discontinuous DP.

According to Hedberg, the cleft pronoun behaves according to the accessibility hierarchy by Gundel et al. (1993), which ranks different degrees of givenness of referents for referring expressions (and which will be discussed in detail in section 8.2). She claims that different cleft pronouns can be used based on the degree of givenness, comparing *it* to *this* or *that* in the full cleft, for instance. Hedberg (2000) uses the truncated cleft as an example in favor of her approach. She claims that in truncated clefts, such as (56), the cleft pronoun is fully referential since the descriptive content is elided.

- (56) Somebody knocked at the door. It was Arne.

I argue that in the interpretation process of (56) the addressee does not resolve the reference of the pronoun *it*. Instead, I suggest that she/he recovers the elided descriptive material in order to interpret the full cleft structure. The cleft pronoun would then not differ from a cleft pronoun in a full cleft. The difference seems to lie in the givenness of the cleft relative clause rather than in the givenness of the pronoun.

Reeve (2011), as well, argues that the cleft pronoun cannot be expletive. According to his approach, it patterns with fully referential pronouns. He observes that the cleft pronoun cannot be dropped in neither English, German, Danish, nor Icelandic. However, for each possible environment for an expletive *it*, at least one of those languages drops the expletive or makes it optional. This would predict the cleft pronoun to be optional in least one

of the languages. Which language that would be, depends on which kind of expletive the cleft pronoun would be. Only referential pronouns are never dropped in any of these languages. Hence, the cleft pronoun must be referential. However, Reeve (2011) does not explicate what exactly the cleft pronoun is referential to.

Furthermore, Reeve (2011) elaborates on Hedberg's (2000) analysis by saying that the cleft relative clause must be a restrictive relative clause. He presents a lot of syntactic parallels between the restrictive relative clauses and cleft relative clauses. The only huge problem with this approach is that the cleft relative clause semantically does not behave like a restrictive relative clause. This is illustrated by example (57).

(57) In einer Bar waren vier Personen. Ein Mann, eine Frau, der Inhaber und ein Kellner. An dem Abend wurde alles Geld aus der Kasse gestohlen. Die vier konnten nicht klären, wer von ihnen das Geld genommen hat. Um 22 Uhr hat eine Nachbarin gesehen, dass eine der vier Personen die Bar verlassen hat.

'There were four people in a bar. A man, a woman, the owner, and a waiter. That night all the money was stolen from the register. The four couldn't sort out who of them took the money. At 10 o'clock a neighbor saw one of the four persons leave the bar.'

- a. Es war der Mann, der das GELD gestohlen hatte.
it was the man who the MONEY stolen had
'It was the man who had stolen the money.'
- b. Es war der MANN, der das Geld gestohlen hatte.
it was the MAN who the money stolen had
'It was the MAN who had stolen the money.'
- c. Es war der Mann, der das GELD gestohlen hatte, und zwar der
it was the man who the MONEY stolen had namely the
Inhaber selbst.
owner himself
'It was the man who had stolen the MONEY, namely the owner himself.'
- d. #Es war der MANN, der das Geld gestohlen hatte, und zwar der Inhaber
it was the MAN who the money stolen had namely the owner
selbst.
himself
'It was the MAN who had stolen the money, namely the owner himself.'

(57-a) is a presentational statement with a restrictive relative clause, while (57-b) is a cleft. Superficially, they look very similar. However, the first difference occurs in the intonation (indicated in capital letters) and in the prosodic phrasing (c.f. Garro and Parker, 1982). In (57-a), there is no phrase boundary in between *der Mann* ('the man') and the relative clause. For (57-b), there is a boundary tone (c.f. Doetjes et al., 2004).

Example (57-c) and (57-d) show that the two structures are referential in different ways. The NP+restrictive relative clause, as in (57-c), can in principle refer to any of the three men mentioned in the context, i.e. the guest, the owner, or the waiter. The cleft in (57-d) can only refer to the man mentioned as such. For the cleft, the reference of *der Mann* ('the man') must be resolved without using information in the relative clause. This explains the unacceptability of the cleft (57-d) since *the man* cannot at the same time refer to the owner and to the man referred to as *a man* in the context. Example (57-c), in contrast, is acceptable since *the man who stole the money* can still refer to any of the men in the context, i.e. also the owner. Hence, this provides counter-evidence against the cleft relative clause being a restrictive relative clause.

5.7 Summary

This chapter showed many different syntactic approaches to the structure of clefts. Several questions are still not settled. First of all, there is the question whether the cleft relative clause is extraposed or not. In general, an extraposition rule is problematic because it is always stipulative, and when it is realized as a right-dislocation it constitutes a very unusual movement in languages such as German or English. If the cleft relative clause is extraposed by adjunction to IP, as in Akmajian (1970), the pivot and the relative clause do not form a constituent. This is a desirable result given that the two cannot be moved as one constituent. Another question concerned with the relative clause is whether it is a restrictive or a non-restrictive relative clause. I argue that this issue is settled since it cannot be a restrictive relative clause, based on semantic arguments presented in section 5.6.

Another question is concerned with the cleft pronoun, which is assumed to be either referential (in one way or the other) or expletive. Depending on the other assumptions of the respective approach, the answer to this question turns out differently. For É. Kiss (1998), the pronoun necessarily needs to be expletive, while an analysis such as Hedberg (2000) requires the pronoun to be referential.

Furthermore, we saw that the cleft structure has been derived from or identified to many different constructions ranging from copular constructions to focus constructions. Each parallel had different implications for the semantics of the cleft. On the one hand, it is very convincing, syntactically, that the cleft resembles a copular construction, such as a definite pseudocleft or a *wh*-cleft. However, such approaches make too strong predictions concerning uniqueness, or cannot account for the difference between *it*-clefts and *wh*-clefts. On the other hand, deriving the cleft from an underlying focus canonical sentence seems to be a bigger derivational effort, which involves quite a complex underlying structure. Nevertheless, this approach can account for the many observed parallels between focus and clefts. As mentioned above, now the problem is that the representation is too tightly bound to focus, and, therefore, topic comment clefts cannot be accounted for. An approach which combines both the parallel to a copular construction and the relation to focus is the approach by Seuren (1985), if we interpret the question that is addressed in the structure he proposes as the question that focus indicates. He explicitly only covers focus background clefts, and proposes a maybe too strong contrastivity for clefts. It might be possible to extend this approach to also cover topic comment clefts, if one assumed that they pose and answer a different question than the focus background cleft. I will leave this extension to future research.

Chapter 6

Existence Presupposition of Clefts

This chapter gives a brief overview about the existence presupposition of clefts. The literature on cleft sentences in many different languages agrees that clefts have an existence presupposition (e.g. Percus, 1997; Halvorsen, 1978; Prince, 1978). This includes German *es*-clefts, illustrated in example (1).

- (1) a. Es ist Maya, die Gitarre spielt.
it is Maya who guitar plays
'*It is Maya who is playing the guitar.*'
- b. Existence presupposition: Somebody is playing the guitar.

Different embeddings of the cleft under non-veridical predicates, as in (2), show that the existence inference projects, which supports the assumption that it is indeed a presupposition. As we will see below, this only holds for focus background clefts, when *Maya* is focused in (2).

- (2) a. It is possible that it is Maya who is playing the guitar.
b. It is not the case that it is Maya who is playing the guitar.
c. If it is Maya who is playing the guitar, I'll be happy.
 ~ Somebody is playing the guitar.

Percus (1997) argues that the existence presupposition of the cleft explains why the cleft is incompatible with *nobody* or *nothing* in the pivot, as in (3).

- (3) a. #Es ist niemand, der Gitarre spielt.
it is nobody who guitar plays
'It is nobody who is playing the guitar.'
- b. #Es ist nichts, das Maya spielt.
it is nothing that Maya plays
'It is nothing that Maya is playing.'

According to Percus (1997), such examples are infelicitous because the prejacent of the cleft (*Nobody is playing the guitar*) contradicts the existence presupposition. He derives the existence presupposition from an underlying definite description, as in (4).²⁵

- (4) The one who is playing the guitar is Maya.

According to Percus (1997), the cleft inherits the presupposition from the definite description. Those have been assumed to have an existence presupposition independently of clefts (c.f. Strawson, 1950). Truth value judgment studies, in contrast, have shown that the existence presupposition of definite description is ignored by English speakers in certain contexts. Abrusán and Szendrői (2013) found that under certain conditions a sentence containing the definite description *the king of France* is judged true, even though there is a presupposition failure.²⁶ Example (5) presents some of their examples.

- (5) a. ?The king of France isn't bald.
b. The king of France is not on a state visit to Australia this week.
c. The king of France is not married to Carla Bruni.

(Abrusán and Szendrői, 2013:85)

Abrusán and Szendrői (2013) found that (5-b) and (5-c) were judged true significantly more often than (5-a). Those participants who did not choose '*true*' chose the option '*can't*

²⁵The syntactic analysis is described in more detail in section 5.3.

²⁶The participants did have the option to choose '*can't say*', which would be the more expected option in case of a presupposition failure.

say' because of presupposition failure. Hence, it seems that the presupposition failure was ignored in (5-b) and (5-c) but not in (5-a). Following von Fintel (2004), Abrusán and Szendrői conclude that sentences that involve a non-referring definite description, such as *the king of France*, can be judged true more easily if the sentence otherwise contains information that can be evaluated in the context. In (5-c), for instance, the participants know that Sarkozy is married to Carla Bruni, and whoever the king of France is, it is not Sarkozy. Hence, the truth can be judged without identifying a referent for *the king of France*.

This translates to the definite description, proposed by Percus (1997), as in (6-a), which corresponds to the cleft in (6-b).

- (6) a. ?The one who destroyed the Eiffel Tower was not some man from the street.
 b. ?It was not some man from the street, who destroyed the Eiffel Tower.

The non-referring part in (6-a), corresponding to *the king of France* in (5-a), is the definite description *the one who destroyed the Eiffel Tower*. We know that this definite description does not fulfill its existence presupposition since nobody destroyed the Eiffel Tower. Hence, it does not refer to anybody, like *the king of France* did not refer. The part corresponding to *isn't bald* is *was not some man from the street*. The cleft is derived, as Percus (1997) proposes, in (6-b). In line with (5-a), these examples seem to involve a presupposition failure, too, and I hypothesize participants to choose the 'can't say' option, since they cannot evaluate the sentence in the context because they know nothing about some guy from the street.

If the content is changed to one that can be evaluated more easily, as in (5-c), it still does not seem to change the impression that the truth of the sentence cannot be determined, as illustrated in (7-a) and (7-b).

- (7) a. ?The one who destroyed the Eiffel Tower one minute ago was not this little girl
 playing in front of us.

- b. ?It is not this little girl playing in front of us who destroyed the Eiffel Tower one minute ago.

The definite description and the cleft still lead to a presupposition failure, it seems. I hypothesize that participants would still choose the ‘can’t say’ option even if they could evaluate that this little girl who was playing here one minute ago cannot have destroyed the Eiffel Tower one minute ago. As long as the addressee does not accommodate that the Eiffel Tower must have been destroyed one minute ago, the existence presupposition cannot be ignored, and the effect observed by Abrusán and Szendrői (2013) does not seem to hold for the existence presupposition of clefts or definite pseudoclefts. This might have to do with the fact that there is much more descriptive content in the part of the definite description of the cleft as for *the king of France*. The description of the referent in the cleft always involves an event, such as destroying the Eiffel Tower, which is maybe more difficult to ignore. This speaks in favor of the existence presupposition in clefts to be particularly strong. This claim, as well as the judgments for (7) need to be investigated more thoroughly in future research.

Contrary to this claim, there are some cases in which the existence presupposition is not as clearly identifiable, or possibly not present at all. One such case is the topic comment cleft, such as in (8).

- (8) Unser Auftritt musste beinahe abgesagt werden, weil unser Gitarrist krank geworden ist. Aber Maya hat den Abend gerettet. Sie hat uns beruhigt. Und sie war es, die am Ende die GITARRE gespielt hat.

‘Our gig almost had to be canceled because our guitar player got sick. But Maya saved the night. She calmed us down. And it was her who played the GUITAR in the end.’

In this context, it is not obvious that the inference *Somebody played the guitar in the end* is presupposed. On the one hand, one could argue that it is easy to accommodate the

presupposition given that Maya saved the night. On the other hand, embedding under negation in this case shows that the presupposition does not project, illustrated in (9).

- (9) Our gig almost had to be canceled because our guitar player got sick. But Maya saved the night. She calmed us down. **And it was not the case that it was her who played the GUITAR in the end.** But she joined the band with her cello.

In this example, it is implied by the last sentence that nobody played the guitar in the end, which does not seem to conflict with any inference of the cleft. Hence, it seems that the existence presupposition is either not existent for the topic comment cleft or it has to be formulated in a different way. Prince (1978), for instance, argues that the relative clause of a topic comment clause is marked as a known fact at least to the author, though not necessarily to the addressee. This explanation, however, cannot account for the acceptability of the last sentence in (9), either. The speaker would contradict what she herself considers a known fact.

Another challenging case is the cleft in (10-a), which can neither be argued to have the existence presupposition in (10-b) nor in (10-c) (c.f. Büring and Križ, 2013).

- (10) a. It is $MAYA_F$ who invited $PAUL_F$, not the other way around.
b. Somebody invited Paul.
c. Maya invited somebody.
d. Somebody invited somebody.

The only plausible existence presupposition is (10-d), as suggested by Velleman et al. (2012). They argue that the existence presupposition depends on the focus in the cleft, and that this focus is not always identical to the cleft pivot. In example (10-a), for instance, there are two foci, and, thus, a presupposition involving two existential expressions.

Hence, the existence presupposition seems to depend on the focus in the cleft and cannot just be determined by the underlying definite description. Those approaches that derive

the cleft from an underlying focus structure, in contrast, need to account for why the existence presupposition is much stronger in clefts than it has been observed for focus. Compare (11-a) and (11-b).

- (11) a. #It is nobody who is playing the guitar.
 b. NOBODY is playing the guitar.

Focus does not lead to a contradiction when *nobody* is focused, as opposed to the clefted *nobody*.

Chapter 7

Exhaustivity of Clefts

In this chapter, I provide a summary of the theoretical approaches on the exhaustivity inference of clefts, comparing their predictions to the empirical findings. The experimental findings will be discussed in detail in 9.2. In section 7.1 of this chapter, I introduce attissueness and the semantics/pragmatics distinction as relevant topics for exhaustivity in clefts. Furthermore, I distinguish exhaustivity from uniqueness and maximality. In section 7.2, the exhaustivity inference is discussed as an entailment. Section 7.3 presents different presuppositional accounts, and section 7.4 proposes to treat cleft exhaustivity as a scalar implicature. In section 7.5, different pragmatic approaches that derive a generalized implicature are subsumed .

7.1 Introduction

Generally, the *it*-cleft in (1-a) is associated with the exhaustivity inference in (1-b), which is the same for *es*-clefts.

- (1) a. It is Maya who danced salsa.
 b. Exhaustivity inference: Nobody other than Maya danced salsa.

There is not much debate about whether this inference exists or not. However, it is much less clear what the SOURCE of this inference is, i.e. whether it is a semantic inference or

a pragmatic inference. Furthermore, the STATUS of the inference is a point of debate, i.e. whether it is an at-issue or a not-at-issue inference.²⁷

For the distinction between semantic and pragmatic inferences, I use the diagnostic of cancellation, as in (2). Semantic inferences cannot be canceled while pragmatic inferences may be canceled without leading to a contradiction.

- (2) It is Maya who danced salsa, but Paul danced salsa as well.

The diagnostic shows that the exhaustivity inference may be canceled since (2) is felicitous. Hence, according to this test, the inference seems to be a pragmatic inference. For identifying the at-issue-content of a sentence, I will use the following diagnostic, taken from Tonhauser (2012).

- (3) A: It was Maya who danced salsa today.

B:

- a. No, that's not true, she did not dance salsa.
- b. #No, that's not true, Paul also danced salsa.

This diagnostic uses the effect that when the addressee disagrees with an utterance by saying *No, that's not true*, the correcting statement (*she did not dance salsa/Paul also danced salsa*) must address the at-issue content, as in (3-a). Otherwise, it is infelicitous, as in (3-b). Hence, the diagnostic for identifying at-issue content of a test sentence works as follows. It identifies the content of the correcting sentence as the negated at-issue content of the test sentence, as long as it is felicitous in combination with *No, that's not true*. According to this diagnostic, the exhaustivity inference of clefts is not-at-issue.

These diagnostics serve as a good starting point. However, they do not capture every facet of the semantics/pragmatics distinction with respect to exhaustivity in clefts. If it was a pragmatic inference, one would still need to identify the exact nature of that inference. On the one hand, it could be an implicature derived in the Gricean way (Horn, 1981; De

²⁷The terms SOURCE and STATUS are established by Destruel et al. (2015), and I will use their terminology.

Veaugh-Geiss et al., 2018a). On the other hand, it could be derived by reasoning about the process of resolving anaphoric material in the cleft given a QUD model (Onea, 2019a; Pollard and Yasavul, 2014; Velleman et al., 2012). All of these derivations, in one way or the other, involve pragmatic reasoning.

It has been shown that the exhaustivity inference of clefts differs in strength from other exhaustivity inferences. Exhaustivity in clefts is often analyzed in comparison to exclusives, as in (4-a), and plain focus, as in (4-b), which are both claimed to give rise to some kind of exhaustivity inference.

- (4) a. Only Maya danced salsa.
 - b. MAYA danced salsa.
- \rightsquigarrow Nobody other than Maya danced salsa.

The exhaustivity inference of clefts seems to be stronger than the exhaustivity inference of focus, and weaker than the inference of exclusives (see, among others, De Veaugh-Geiss et al., 2018b; Onea and Beaver, 2009; Destruel, 2013). As Onea (2019a) points out, there is a discrepancy between the theoretical predictions and the experimental findings with respect to cleft exhaustivity. While most of the theoretical work on clefts analyzes their exhaustivity inference as part of their semantics, most experimental findings speak against a semantic analysis because the inference is not very robust in experimental settings.

The at-issueness of the exhaustivity inference of clefts is less debated, but different positions have been discussed in previous literature, which I will summarize in this chapter. Hence, there are four possible options how to classify the exhaustivity inference of clefts, as summarized in Table 7.1.

Source	Status	Classification of the inference
semantic	at-issue	entailment
semantic	not-at-issue	presupposition
pragmatic	at-issue	scalar implicature
pragmatic	not-at-issue	generalized implicature or other pragmatic inference

Table 7.1: Possible classifications for the cleft exhaustivity inference

Based on the source and the status of this inference, it can be classified, as displayed in the third column in Table 7.1. In the following sections, I will discuss each of these classifications in more detail. Before, I want to distinguish the concepts of MAXIMALITY and UNIQUENESS from exhaustivity. Many approaches propose uniqueness or maximality inferences for the cleft, which might lead to exhaustivity, but which do not describe the same inference.

UNIQUENESS with respect to some predicate P requires there to be just one individual satisfying P . An example for uniqueness is provided in (5), where the definite description can be argued to presuppose uniqueness.

- (5) The one who danced was Maya.

Uniqueness requires that there is exactly one dancer in the context of (5). This entails the existence of a dancer. It is important to note here that uniqueness is determined independently of the main predicate of the sentence that contains the definite description. The uniqueness presupposition of the definite description in example (5) is determined just based on the predicate *one who danced*, independently of the main predicate (*be Maya*). This aspect of uniqueness will turn out to be crucial when the exhaustivity inference of clefts is expressed via uniqueness.

For a cleft, such as (6-a), uniqueness is paraphrased as in (6-b). Exhaustivity, in contrast, is represented in (6-c).

- (6) a. It is Maya who danced.
 b. There is a unique individual that danced.
 c. $\forall x[danced(x) \rightarrow (x = Maya)]$

“All x that dance must be equal to Maya.”

Uniqueness again entails the existence of a dancer, and is not dependent on the predicate *be Maya*. The exhaustivity inference in (6-c) can be formulated as a disjunction stating that for all x , x does not dance or x is identical to Maya. This inference does not

entail anything about the existence of a dancer. It is compatible with nobody dancing, which describes the case in which both disjuncts are true. Furthermore, the exhaustivity inference depends on the main predicate *be Maya*. In a context in which there is just one dancer, namely Paul, uniqueness, as described in (6-b), is satisfied, and exhaustivity, as defined in (6-c), is not satisfied.

Now, consider the difference between exhaustivity and MAXIMALITY. Maximality can be defined with respect to individuals satisfying a predicate, as in (7), or with respect to true answers to a question, as in (8).

- (7) $\sigma x [A(x)]$, where σ picks out the maximal individual that fulfills the predicate A .
 (c.f. Link, 1983:131)

- (8) $\text{MAX}_C(p) = \lambda w. \forall q \in C Q_C[(q >_C p) \rightarrow \neg q(w)]$
 “No true answer to the question CQ_C is strictly stronger than p .”
 (Velleman et al., 2012:451)

In (7), x can be a single individual or a sum individual, and the predicate A is the singular or plural version, respectively. The formalization in (8) will be discussed in detail in section 8.6. For now, we stick to the paraphrase. For our cleft example, maximality predicts the following two inferences, (9-b) for maximality with respect to predicates, and (9-c) for maximality with respect to questions.

- (9) a. It is Maya who danced.
 b. There is a maximal (sum) individual that danced.
 c. No true answer to the question *Who danced?* is strictly stronger than *Maya danced*.

Büring and Kriz (2013) point out that maximality, as formulated in (9-b), trivially follows from the existence presupposition of clefts. As long as there is at least one individual α that dances, there is necessarily a maximal individual of dancers. It is either α itself, or a

sum individual containing α . It will depend on the assertive content whether this version of maximality leads to exhaustivity or not, as we will see below.

Exhaustivity and maximality, as described in (9-c), describe the same effect in (9), given that if A *danced* is true and A *and* B *danced* is false, it follows that also B *danced* must be false.²⁸

7.2 Exhaustivity as Entailment

Especially the earlier approaches towards cleft exhaustivity claimed that the inference is an entailment, which is an at-issue, semantic inference (c.f. Bolinger, 1972; Atlas and Levinson, 1981; Szabolcsi, 1981). They treat exhaustivity as one component of the assertive meaning of a cleft.

Support for this approach comes from examples that involve negation. Negation only interacts with at-issue content, as we have seen in the diagnostics for at-issue content above. If the exhaustivity inference is part of the at-issue content, it should be possible to negate that inference with sentential negation of the cleft sentence, such as in (10).

- (10) Es war nicht Laura, die geschlafen hat. Es waren Laura und Lea.
it was not Laura who slept has it were Laura and Lea
'It wasn't Laura who slept. It was Laura and Lea.'

Since (10) is well-formed, the exhaustivity inference seems to be at-issue. The canonical sentence, which is usually assumed to not entail exhaustivity, shows a different behavior, as indicated by (11).

- (11) #Laura hat nicht geschlafen. Es waren Laura und Lea.
Laura has not slept it were Laura and Lea
'Laura did not sleep. It was Laura and Lea.'

In German, the same effect as with the cleft can be achieved when negation only takes scope over the subject *Laura*.

²⁸This only follows for distributive predicates. For a collective predicate P , it does not follow from $P(A)$ and $P(B)$ that $P(A+B)$.

- (12) Nicht LAURA hat geschlafen. Es waren Laura und Lea.
 not LAURA has slept it were Laura and Lea
 ‘LAURA did not sleep. It was Laura and Lea.’

There are very few contexts in which (10) and (12) are felicitous. I argue that those are cases of metalinguistic negation (Horn, 1985), which can target not-at-issue contents, as opposed to sentential negation. Meta-linguistic negation can, for instance, negate presuppositional content, as illustrated in the following example.

- (13) The king of France isn’t bald, because there is no king of France.

In this example, only the existence presupposition is negated in the first sentence, not the part about the king of France being bald. Something similar can be assumed for examples (10) and (12), and possibly even for (11) with the right intonation. If negation is only possible as long as we are dealing with metalinguistic negation, the examples do not show that we are dealing with a non-at-issue inference.

In general, the entailment approach has been discarded on the basis of Horn’s (1981) cleft examples, which showed that exhaustivity can be canceled. This is not compatible with the assumption that exhaustivity is part of the cleft’s semantics. Horn’s approach will be discussed in section 7.5.

7.3 Exhaustivity as Presupposition

Several researchers have argued that exhaustivity in clefts is presuppositional, hence a not-at-issue, semantic inference. The exact formulation of this presupposition varies, though. A challenge for nearly all of the presuppositional approaches is the observation that the exhaustivity inference does not seem to project, as indicated in (14).

- (14) It is not the case that it is Natalio who purred.

↗ Nobody other than Natalio purred.

This example shows that the exhaustivity inference does not survive when it is embedded under negation. Other presuppositions do survive embedding under non-veridical operators, though.

Percus (1997), as mentioned in section 5.3, assumes *it*-clefts to have a uniqueness presupposition, which they inherit from an underlying definite description. The uniqueness presupposition can be defined as in (15-b) for the cleft in (15-a).

- (15) a. It is Natalio who purred.
- b. $x : \forall y.\text{PURR}(y) \rightarrow (x = y)$

Percus (1997) does not specify what he assumes as the at-issue meaning of the cleft. His approach could be interpreted, though, as asserting that the presupposed unique individual is identical to Natalio (i.e. $x = \text{Natalio}$). This would be a semantic representation based on the underlying definite description assumed by Percus. Hence, in this example, the assertion and the uniqueness presupposition together yield exhaustivity because the unique individual is identified as the pivot.

For the cleft embedded under negation, Percus (1997) would assume the following components.

- (16) It is not the case that it is Natalio who purred.
- a. Presupposed content: $x : \forall y.\text{PURR}(y) \rightarrow (x = y)$
- b. Asserted content: $\neg(x = \text{Natalio})$

Correctly, this representation does not imply that nobody other than Natalio purred, hence no exhaustivity. It does still presuppose that there is a unique individual that purred, though.

For plural clefts, such as (17), however, exhaustivity cannot be correctly predicted in this approach.

- (17) It is Natalio and Felipa who purred.

We do not want to assume that the plural cleft presupposes a unique cat. Moreover, assuming uniqueness of a plural sum individual, such as $Natlio \oplus Felipa$, would not be a good alternative, either, as Szabolcsi (1994) points out. If Natalio and Felipa purred, it follows that also Natalio purred, and Felipa purred. Hence, it is simply impossible that the sum of Natalio and Felipa is the unique referent that purred. This holds at least for distributive predicates.

According to Szabolcsi (1994), it is rather maximality that would make the correct predictions for plural clefts. Link (1983) defines maximality using the σ operator, as illustrated for the plural clefts in (17) as follows.

- (18) $\sigma x [PURR^*(x)]$, where σ picks out the maximal individual that fulfills the predicate PURR*.

PURR* is the plural predicate corresponding to PURR, which includes all entities that purr, and all sums of thereof. Accordingly, the variable x in (18) can represent a single individual as well as a sum individual. Szabolcsi (1994) proposes the following presupposition for the plural cleft in (17).

- (19) $x : PURR^*(x) \wedge \forall y.PURR^*(y) \rightarrow y \subseteq x$

This solves the problem of plural clefts. As for Percus (1997), the assertive content of the cleft ($x = Natlio \oplus Felipa$) and the maximality presupposition together yield exhaustivity. If the prejacent of the cleft ($PURR^*(Natlio \oplus Felipa)$) was assumed as the assertive content of the cleft, the maximality presupposition would be equivalent to the existence presupposition, at least if we assume cumulativity or a structure for plural predicates that is closed under sum formation \oplus (Link, 1983). Whenever there exists any individual, there must be a maximal individual as well. Either it is exactly that individual, or it is a sum containing that individual. Hence, maximality does not say more than the existence presupposition. Crucially, it is possible that this presupposed maximal sum individual contains more individuals than Natalio and Felipa. Hence, exhaustivity is not predicted.

Pollard and Yasavul (2014) assume that clefts presuppose a discourse referent, and that they further specify this discourse referent as their at-issue content. This approach does not assume clefts to presuppose uniqueness or maximality. For Pollard and Yasavul, it depends on the context how this reference is resolved. Crucially, they consider both maximal and non-maximal referents to be possible. Accordingly, the cleft is predicted to only sometimes give rise to an exhaustivity inference. They argue that the exhaustivity inference only arises when the cleft answers an explicit *wh*-question, as in (20).

(20) Who went to CLS?

- a. Greg and Dan. I don't know if anyone else did. / Scott did, too.
- b. It was Greg and Dan. #I don't know if anyone else did. / #Scott did, too.

(Pollard and Yasavul, 2014:3)

Pollard and Yasavul (2014) argue that questions introduce a maximal discourse referent which the cleft addresses. This leads to exhaustivity. In other examples, such as (21), there is no exhaustivity inference.

(21) A: Did you hear, Bob got an NSF grant!

B: Well, actually, it was Rob. And Mike got one, too!

(Pollard and Yasavul, 2014:2)

In this context, the cleft anaphorically refers a non-maximal referent (*Bob*), and does, thus, not lead to exhaustivity.

Büring and Križ (2013) argue that the cleft has a homogeneity presupposition. They claim that a cleft of the form *it is x that P* presupposes that *x* is not a proper part of the sum of all elements in *P*. Furthermore, they assume that the cleft asserts the prejacent *x P*. For the cleft in (15-a) (*It is Natalio who purred*), this means that Natalio is presupposed to not be a proper part of the sum of all individual that purred. This presupposition is fulfilled when either (i) Natalio is the only individual that purred, or (ii) he did not purr.

The second option is ruled out by the assertion made by the cleft, which is *Natalio purred*. Hence, in this case, it leads to just an exhaustivity inference.

Büring and Križ (2013) predict that exhaustivity does not arise for a cleft which is embedded under negation, as in (22), repeated from above. Their presupposition, in contrast, does project.

- (22) It is not the case that it is Natalio who purred.
 ↗ Nobody other than Natalio purred.
 ↘ Natalio is not a proper part of the individuals that purred.

According to Büring and Križ, the negated cleft in (22) asserts that Natalio did not purr, and it presupposes that Natalio is not a proper part of the sum of all individuals that purred. This time, option (i) of the presupposition (Natalio is the only individual that purred) is ruled out by the assertive content of the cleft. Hence, it is only presupposed that Natalio is not in the set of individuals that purr, as in option (ii). Hence, the presupposition does project but without maintaining exhaustivity.

Büring and Križ (2013) admit themselves that their proposed presupposition is not very natural, and that it does not seem plausible that discourse participants have previous knowledge of the form of their presupposition. Instead, it seems that the addressee is informed about the exhaustivity through the utterance of a cleft and is hardly ever previously aware of the exhaustivity. In other words, the presupposition must be accommodated in most cases. This does not speak in favor of it being a presupposition. Usually, presuppositions are only accommodated under specific conditions depending on the context.

Similar to the entailment approaches, Büring and Križ (2013) cannot explain why the exhaustivity inference may be canceled, as in (23).

- (23) It was Natalio that purred, and Felipa purred, too.

When processing the first clause of (23), the presupposition that Natalio is the only individual that purred must be accommodated (if not present in the context). This contradicts with the second clause in (23). Therefore, (23) would be incorrectly predicted to be infelicitous.

Finally, Velleman et al. (2012) analyze exhaustivity as a maximality presupposition with respect to answers to the current question (CQ). This approach is discussed in detail in section 8.6. Here, I will just provide the gist of it. Velleman et al. (2012) claim that the cleft involves a focus-sensitive cleft operator that presupposes that no true answer to a given CQ is strictly stronger than the prejacent of the cleft. Furthermore, it asserts that there is a true answer to this CQ that is at least as strong as the prejacent. Those two components taken together yield exhaustivity. What distinguishes the approach by Velleman et al. from all the other approaches to exhaustivity is that they can account for the different exhaustivity inferences depending on which element of the pivot is focused, as in (24-a) and (24-b).

- (24) a. It was JOHN'S eldest daughter who attended the party.
 ~> Nobody else's eldest daughter attended the party.
 b. It was John's ELDEST daughter who attended the party.
 ~> No other daughter of John's attended the party.

The presupposition of (24-a) is focus-sensitive with respect to the question *Whose eldest daughter attended the party?*, and the presupposition of (24-b) is focus-sensitive to the question *Which of John's daughters attended the party?*, which would yield different exhaustivity inferences.

7.4 Exhaustivity as Scalar Implicature

Scalar implicatures are an at-issue, pragmatic inference, such as implicatures arising from scalar adjectives, as in (25).

- (25) a. The soup is warm.
 b. Scalar implicature: The soup is not hot.

The *yes, but...*-test by Onea and Beaver (2009), illustrated in (26), shows that such a scalar implicature is at-issue.

- (26) Context: *A and B are eating really hot soup.*
 A: The soup is warm.
 a. B: #Yes, but/and the soup is hot.
 b. B: No, the soup is hot.

Destruel et al. (2015) showed in an empirical study that the answer in (26-b) was chosen significantly more often than the other responses (92% *no*-responses). Hence, the scalar implicature is at-issue. Furthermore, it is a pragmatic inference, which is indicated in (27).

- (27) The soup is warm, even hot.

In this example, the inference is canceled without creating a contradiction. Therefore, it must be a pragmatic inference. According to Horn (1972), scalar implicatures are derived on the basis of contextually provided scales, such as the scale of temperature predicates for (25).

De Veaugh-Geiss et al. (2015) argue that the exhaustivity inference of clefts can be derived on the basis of the scale of ordered focus alternatives. For the cleft in (28-a), this scale is illustrated in (28-b), in a very simplified version.

- (28) a. It is JOE who solved the problem.
 b. <Joe, Mira and Anna solved the problem, Joe and Mira solved the problem,
 ..., Joe solved the problem>

The scale contains the focus alternatives of the cleft in (28-a) ordered by their strength. Hence, the higher alternatives entail the lower alternatives. De Veaugh-Geiss et al.

(2015) analyze exhaustivity in clefts as a focus-driven scalar implicature. The implicature is derived as follows: Based on the scale in (28-b), the speaker could have made the stronger statement *Joe, Mira and Anna solved the problem* or *Joe and Mira solved the problem* on the scale. Since the speaker is assumed to follow the maxim of quantity the addressee assumes that the speaker would have made the stronger statement if she/he had evidence for it. The addressee concludes that the speaker did not have evidence, and the alternatives which are higher on the scale are excluded. Hence, the addressee derives the scalar implicature that the other individuals in the context, except for Joe, did not solve the problem. This derivation is identical to how the exhaustivity inference is usually derived for plain focus, as in (29).

- (29) JOE solved the problem.

As shown in chapter 9, many empirical studies found that plain focus gives rise to a significantly weaker exhaustivity inference than clefts (see, e.g., De Veaugh-Geiss et al. (2018b) for German, or Destruel et al. (2015) for French). De Veaugh-Geiss et al. (2015) provide an explanation for this, which is based on focus projection. As mentioned before, focus marking for plain focus can be ambiguous. For clefts, however, focus is always marked unambiguously, and focus projection is not possible. Example (30) shows that the focus cannot project from the pivot to a bigger constituent, as in (30-a). It can only mark narrow focus, as in (30-b).

- (30) a. What did Maria do yesterday? – #It was a PIZZA that she ate.
b. What did Maria eat yesterday? – It was a PIZZA that she ate.

Based on this observation, De Veaugh-Geiss et al. (2015) argue that for plain focus the pragmatic scales for deriving the implicature are ambiguous and, thus, complicate pragmatic enrichment. For clefts, the scales are always easily accessible, and lead to a stronger exhaustivity inference.

There are some problems with this explanation, though. First of all, I argue that when a sentence with an ambiguous focus marking is interpreted, its focus needs to be accom-

modated anyway, independently of the exhaustivity of focus. Otherwise, the addressee cannot make a decision about whether she/he can accept the sentence as an appropriate discourse move. See section 10.3, where I define the accommodation of the CQ based on discourse expectedness (Definition 35). If this accommodation has to take place anyways, the scale of the focus alternatives is automatically disambiguated as well. Hence, exhaustivity should be as easy to derive as for clefts.

Another problem for this approach is that it predicts an asymmetry between the exhaustivity inference of focus marked objects and focus marked subjects. The former is an ambiguous focus marking, while the latter is not ambiguous. The focus marking in (29), for instance, may only answer the question *Who solved the problem?*. Focus projection is not possible, just as for the cleft. Hence, De Veaugh-Geiss et al. (2015) would predict that (29) would have a stronger exhaustivity inference as a sentence with ambiguous focus marking, and they would predict (29) to give rise to an equally strong exhaustivity inference as clefts. De Veaugh-Geiss et al. (2018b) tested plain subject focus, though, and did find a weaker exhaustivity inference as for clefts.

7.5 Exhaustivity as Generalized Implicature or Other Pragmatic Inference

Finally, I will get to the approaches that can account for the cancelability of the exhaustivity inference in clefts. Accordingly, exhaustivity is derived as a generalized implicature. The approaches differ with respect to how this implicature arises. In the following, I discuss different pragmatic derivations of a potential exhaustivity inference of clefts, and I also include approaches that propose a pragmatic reasoning in the wider sense. The exhaustivity inference is treated as a not-at-issue, pragmatic inference. The main motivation for these accounts is Horn's (2016) famous example.

(31) As we go marching, marching unnumbered women dead,
go crying though our singing their ancient call for bread.
Small art and love and beauty their drudging spirits knew.

Yes, it is bread we fight for, but we fight for roses too!

poem by James Oppenheim (1911)²⁹

This poem, cited from Horn (2016), contains a cleft, immediately followed by the cancellation of its exhaustivity inference. Furthermore, Horn (1981) showed that the exhaustivity inference is not at-issue by comparing it to the at-issue inference of exclusive particles like *only*.

- (32) #Ich wusste, dass Laura geschlafen hat, aber ich wusste nicht, dass es Laura war,
I knew that Laura slept has but I knew not that it Laura was
die geschlafen hat.
who slept has
'I knew that Laura slept, but I didn't know that it was Laura who slept.'
- (33) Ich wusste, dass Laura geschlafen hat, aber ich wusste nicht, dass nur Laura
I knew that Laura slept has but I knew not that only Laura
geschlafen hat.
slept has
'I knew that Laura slept, but I didn't know that only Laura slept.'

As observed above, sentential negation can only target at-issue content. In (32), this causes a contradiction since the negated at-issue content of the cleft in the second part of the sentence contradicts with the first part of the sentence. For the exclusive in (33), exhaustivity is at-issue, and there is no contradiction. Hence, Horn (1981) concludes that exhaustivity is not-at-issue. The diagnostic in section 7.1 came to the same result.

Horn (1981) formulates the pragmatic principle in (34), based on which the exhaustivity implicature of clefts is derived.

²⁹https://web.archive.org/web/20160216133611/https://en.wikipedia.org/wiki/Bread_and_Roses#Words, accessed 01/15/2021.

- (34) The utterance in context C of any sentence which entails $F\alpha$ and conventionally implicates $\exists x Fx$ will induce a generalized conversational implicature to the effect that $\neg\exists x(x \neq \alpha \& Fx)$, where the variable x ranges over entities in a set determined by the context C . (Horn, 1981:8)

This principle relies on the assumption that the cleft *It is α who Fs* (i) entails $F\alpha$, and (ii) has an existence presupposition, or, in Horn's (1981) terms, it *conventionally implicates* the existence of an entity x such that Fx . The requirement (i) is generally assumed (see Beaver and Clark (2008) for an explication of how this entailment arises). As mentioned above, also the existence presupposition is widely agreed upon. Hence, Horn's principle stands on solid grounds. The only step in the argument that is not obvious is why (i) and (ii) give rise to an exhaustivity implicature. It is not very transparent how exhaustivity relates to the entailment of the prejacent and the existence presupposition of the cleft.

I will now take a closer look at how Horn (1981) suggests to derive the exhaustivity implicature. Horn claims that the exhaustivity implicature is derived when the addressee interprets why the speaker made the additional effort of using a cleft construction, such as (35-a), instead of the much simpler canonical sentence in (35-b).

- (35) a. It is Mavi who is smiling.
 b. Mavi is smiling.

Horn's (1981) reasoning for deriving the exhaustivity implicature for (35-a) can be summarized as follows. The speaker could have used (35-b), which does not presuppose that there is somebody that is smiling. However, the speaker used a more complex structure, which does presuppose exactly this. The speaker must have had a reason for making an extra effort since she/he is assumed to follow the Gricean maxim of manner. Based on Horn's principle, the reason for choosing the cleft must have been exhaustivity.

Onea (2019a), among others, pointed out that the last step of this derivation, and thereby also Horn's (1981) principle, seems ad hoc. Horn does not elaborate on the role of the

existence presupposition but makes it the crucial difference between the two competing structures. However, it is not very plausible that this existence presupposition leads to exhaustivity.

If the inference rather arises because of exploiting the maxim of manner, it is, again, not obvious why the extra effort of a complex structure implicates exhaustivity, and not something else. In principle, there could be several other reasons why the speaker could have chosen a more complex structure than the canonical sentence. Expressing emphasis, contrast, or unexpectedness are only some of the possible reasons. All of these reasons seem equally plausible as exhaustivity, given Horn's (1981) reasoning.

Another point of criticism is that Horn's principle seems to only hold for the cleft, and maybe variants of clefts such as the definite pseudocleft.

- (36) Diejenige, die lächelt, ist Mavi.
the.one who is.smiling is Mavi
'The one who is smiling is Mavi.'

Hence, it does not seem to be a general principle of pragmatic reasoning, but a very specific rule that only holds for a very limited set of structures. This rather speaks in favor of the principle describing a conventional exhaustivity inference encoded through the grammatical structure.

Onea (2019a) presents a way of reinterpreting Horn's (1981) principle in the context of a QUD framework. Following Velleman et al. (2012), he assumes that clefts presuppose an issue instead of an entity. More precisely, he interprets their approach as assuming clefts to anaphorically refer to an issue. He adjusts Horn's (1981) principle by (i) replacing the existence presupposition of an entity by the existence presupposition of an issue and (ii) assuming that the presupposed issue is addressed in a maximal way. He translates the principle into (37).

- (37) If an expression E refers to an issue Q while answering it, the addressee should infer that Q is answered in a maximal true way by E .

(Onea, 2019a:415)

The predictions of this principle do not conflict with those that follow from Horn's (1981) principle. Conceptually, however, it differs quite a lot. First of all, Onea (2019a) categorizes the cleft as a question-answering device. In Horn's analysis, it is conceptually quite unclear what the cleft is. Secondly, it makes a difference to assume a question instead of an entity. This difference gets particularly apparent when the existence of an entity is given via an explicitly mentioned individual as in (38).

(38) Nina lächelt.

'Nina is smiling.'

a. #Es ist Nina, die lächelt.

'It is Nina who is smiling.'

b. #Es ist Mavi, die lächelt.

'It is Mavi who is smiling.'

c. Nein! Es ist Mavi, die lächelt.

'No! It is Mavi who is smiling.'

In such a context, the existence presupposition is fulfilled, but the cleft is nevertheless infelicitous. In (38-a), the cleft is uninformative. In (38-b), it could be a correction, but it is still not felicitous. Only with additional marking, the cleft can function as a correction, as in (38-c) preceded by the response particle *nein* ('no'). The unacceptability of (38-a) and (38-b) indicate that the presupposition that Horn (1981) requires does not arise from (38), but from (39).

(39) Jemand lächelt. Es ist Mavi, die lächelt.

'Somebody is smiling. It is Mavi who is smiling.'

The difference between *Somebody is smiling* and *Nina is smiling* is, conceptually, that in case of the former the addressee does not know who smiled, as opposed to the latter case. Theoretically, the difference amounts to whether the issue about who is smiling is settled or not. Following Onea (2016), van Kuppevelt (1995) and others, the context in (39) raises the potential question *Who is smiling?*, while the context in (38) does not.

Hence, it seems that the crucial factor is that there is an open issue. According to Onea (2019a), the existence of an entity follows from the existence of an issue. He argues that the presupposed issue Q implies the existence presupposition via existential closure of the alternatives in Q . In other words, each alternative in Q describes at least one entity satisfying the required predicate, which means that the existence presupposition holds when the issue is presupposed.³⁰

It seems that Onea's (2019a) principle, just like Horn's principle, only applies to a limited set of constructions. For Onea, it is those structures that refer to an issue while answering it. So far, these conditions only apply to clefts, different kinds of pseudoclefts, and maybe other constructions in other languages (e.g. preverbal focus in Hungarian). However, Onea does not formulate a general pragmatic principle but one that is intentionally tied to a grammaticalization of a construction which refers to a question and answers it, such as the cleft.

Onea (p.c.) argues that his principle is less ad hoc than Horn's (1981) principle. He considers it plausible that the relationship between a question and its answer is grammaticalized in the cleft structure since question-answer pairs are an important ingredient of human communication anyway. Furthermore, it is often claimed that answers are considered to be exhaustive (e.g. by deriving a quantity implicature as illustrated above). Assuming Horn's explanation, however, it does not seem equally plausible that the cleft grammaticalizes the relationship between an existence presupposition and the actual assertion of an entity. These do not seem to be naturally related for reasons of communication. This is why Onea takes Horn's explanation to be less plausible. Nevertheless, Horn's principle does not make any predictions conflicting with those by Onea's (2019a) principle.

Another aspect of Onea's (2019a) principle, that I want to address here, is the assumed anaphoricity of the constructions that it applies to. In general, the reasoning of his principle is reminiscent of how exhaustivity implicatures are commonly derived for plain

³⁰This only holds if the issue does not include negative answers, involving nobody, for instance. This is discussed in section 4.6, where it is explained how indefinite expressions give rise to potential questions.

focus. Consider example (40-a) with a narrow focus on *Mavi*, compared to the cleft equivalent in (40-b).

- (40) a. MAVI was smiling.
b. It was MAVI who was smiling.

It has been argued that also plain focus has a fairly strong exhaustivity implicature. Moreover, focus is also used to answer a question, and, as focus was defined in chapter 3, it at least indicates a question. Therefore, one could argue that Onea's (2019a) principle should equally apply to plain focus. It seems, however, that the issue that focus refers to must be a more salient issue than the one the cleft can refer to. It is like the difference between an anaphoric pronoun like *her* compared to a definite description like *his mother*. The former corresponds to plain focus, while the latter corresponds to the cleft. This difference is illustrated in the following example.

- (41) Tom's mother gave him a ride to school. He got off the car and immediately started chatting with his friends in the school yard.
a. ?He did not even see her wave.
b. He did not even his mother wave.

In this context, the discourse referent given by *Tom's mother* is not particularly salient anymore. Hence, it is more appropriate to refer to her using a definite description, as in (41-b), instead of a pronoun, as in (41-b). Analogously, the cleft can be argued to refer to a less salient issue, compared to focus. This approach is quite close to what Gundel (1977) and Hedberg (1990) are describing, but formulated in a different framework, based on questions. Moreover, it naturally leads to the analysis that I will propose in section 10.2.

With respect to exhaustivity, however, we either have to adjust Onea's (2019a) principle to distinguish between salient issues and less salient issues, or we predict that focus and clefts are both exhaustive, which is not what is observed. Furthermore, even for clefts the principle is too strong since it predicts every cleft to provide a maximal answer. Hence,

those cases in which the cleft is not exhaustive are not covered. The principle could be reinterpreted as only describing the default case of clefts. The non-default uses of a cleft would refer to a question while answering it for a different reason than answering the question in a maximal true way. One possible reason could be reminding the addressee of a question. In such a non-default case, the principle would not apply.

De Veaugh-Geiss et al. (2018a) present experimental evidence which they argue to speak in favor of exhaustivity being a pragmatic inference in German *es*-clefts, which is still stronger than for plain focus, though. They used a mouse-driven picture verification/falsification task, in which participants incrementally received more information about the scenario until they decided that they had enough information to make a truth value judgment. In a post-hoc analysis of their experimental results, they found that the participants split up into two groups. Group 1 interpreted the cleft exhaustively and group 2 did not, which speaks in favor of a pragmatic analysis. The experiment is discussed in more detail in section 9.2.

De Veaugh-Geiss et al. (2018a) follow the approach by Pollard and Yasavul (2014) in assuming that clefts must be resolved to a presupposed discourse referent. They derive an implicature based on the observation that German *es*-clefts vary in number. In (42), the cleft copula agrees with respect to number with the singular pivot, and in (43) it agrees with the plural pivot.

- (42) Es **war** Mavi, die gelächelt hat.
it be.SG Mavi who smiled has
'It was Mavi who smiled.'

- (43) Es **waren** Mavi und Nina, die gelächelt haben.
it be.PL Mavi and Nina who smiled have
'It was Mavi and Nina who smiled.'

De Veaugh-Geiss et al. (2018a) argue that the presupposed discourse referent inherits the number feature from the cleft sentence. They derive the exhaustivity implicature as follows. If the speaker chose the singular cleft, she/he implies that there are no sum-individuals which the predicate in the relative clause applies to. If there were, the ad-

dressee would not be able to resolve the singular reference. Hence, the speaker must have meant that there are no sum-individuals in the context that serve as antecedent. Exhaustivity follows. They argue that those participants that treated the cleft exhaustively, restricted the domain in a way to only include the referent that the clefted constituent refers to, since the context did not provide much information to rely on.

De Veaugh-Geiss et al. (2018b) suggest a different interpretation of the same results. They claim that the choice of interpretation for each group depended on which context the participants imagined in the experiment. At the beginning of the experiment, four flatmates were introduced who undertake various activities together (it was not specified which activities). Then, each stimulus was presented without context. Following Pollard and Yasavul (2014), De Veaugh-Geiss et al. claim that the participants needed to accommodate the discourse referent that the cleft refers to since the context did not provide it. There were two options that either led to an exhaustive or a non-exhaustive interpretation. Consider the cleft (44), which was actually used in the experiment.

- (44) Es ist Ben, der einen Pullover angezogen hat.
it is Ben who a sweater put.on has
'It is Ben who put on a sweater.'

Interpreting this cleft without context, the participant needs to accommodate the discourse referent. She/he can do this by assuming the question *Who put on a sweater?*, or by assuming an indefinite statement, such as *Somebody put on a sweater*. The former leads to an exhaustive inference, the latter leads to a non-exhaustive inference. This claim needs to be tested, e.g. by varying the context in an experiment, such as the one by De Veaugh-Geiss et al. (2018b).

7.6 Summary

We have seen that there is a large amount of analyses of the exhaustivity inferences of *it*-clefts and *es*-clefts. Some of these approaches can be ruled out because of striking counter-evidence. The entailment approach, for example, can be discarded since it has

been shown by corpus examples and experimental results that the exhaustivity inference of clefts is not particularly robust. It is frequently canceled or ignored, at least in English and German. The presuppositional accounts also struggled with the problem that the exhaustivity inference of clefts can be canceled. The scalar implicature approach had problems predicting the difference between plain focus and clefts.

The empirical results, which are discussed in more detail in section 9.2, speak in favor of a pragmatic analysis. I conclude that it is safe to assume that clefts have a non-at-issue, pragmatic exhaustivity inference. The remaining challenge is to explicate how this inference is derived. It could be by maximally resolving a presupposed issue (Onea, 2019a), by deriving a number implicature (De Veaugh-Geiss et al., 2018a), by making assumptions about the context (this explanation is closely tied to the experimental set-up in De Veaugh-Geiss et al. (2018b)), etc. It needs further empirical evidence in order to decide which approach makes the best predictions.

Chapter 8

Discourse Approaches

A large amount of different discourse functions of clefts have been proposed in the literature. Most claims are made about English *it*-clefts, some about French *c'est*-clefts, and very few about German *es*-clefts. Section 8.1 presents a brief and informal overview of most of the discourse functions that have been assigned to the cleft. The following sections will discuss in more detail how clefts are related to givenness (section 8.2), discourse coherence (section 8.3), contrast (section 8.4), focus (section 8.5), and questions (section 8.6).

8.1 Introduction

As mentioned above, a lot has been said about clefts in discourse. Prince (1978) argues that the cleft marks the material in the cleft relative clause as a known fact. She distinguishes, however, whether it is known to the speaker and the addressee or just to the speaker.

Frequently, the cleft has been argued to express contrast with respect to an entity or proposition in the preceding context (e.g. Destruel and Velleman, 2014; Destruel et al., 2019; Prince, 1978). Others phrase the role of clefts in discourse in terms of unexpectedness (e.g., Hedberg, 1990; Sheil, 2016). Hedberg states that it “is particularly common in

mystery novels for clefts to be used to highlight the role of an unexpected participant in an expected event” (Hedberg, 1990:120). Her example illustrates what she describes.

- (1) Beginning at the top of the list, I went along the landing and tapped at Ruskin’s door. When it was opened, it was Webber who stood there. We stared at each other for a moment, both of us taken aback.

[Lucille Kallen, The Piano Bird, p. 95]

(Hedberg, 1990:120)

Fischer (2009) claims that clefts are used to build up suspense, which he illustrates with an example taken from Harry Potter.

- (2) “*Little tyke*” chortled Mr Dursley as he left the house. He got into his car and backed out of number four’s drive.

a. It was at the corner of the street that he noticed the first sign of something peculiar

b. An der Straßenecke fiel ihm zum ersten Mal etwas
at the corner.of.the.street occurred to.him for.the first time something
Merkwürdiges auf
peculiar

‘At the corner of the street, he noticed the first sign of something peculiar’

Fischer (2009:185)

He argues that the German translation, which is the official German translation, and which does not use a cleft, loses the effect of suspense. Delin and Oberlander (1995) claim that replacing an original cleft in a text sometimes leads to a different interpretation of temporal relations between the events introduced in the context.

Sheil (2016) observes that clefts often provide unexpected information, namely by realizing a discourse move that opens a new line of inquiry instead of sticking to the current inquiry as expected. She terms the cleft a narrative device used to progress the story or to signal the termination of a sub-plot.

Dufter (2009) comes to the conclusion that *es*-clefts in German fulfill among other functions also ”discursive goals such as enhancing cohesion, marking transitions, or differenti-

ating between propositional content that is under discussion and information that is not” (Dufter, 2009:115).

Other approaches analyze the discourse effect of clefts in a question-based framework. Tönnis et al. (2018), for instance, analyze clefts as devices to mark the current question in order to disambiguate focus. Velleman et al. (2012) treat clefts as inquiry terminating constructions that terminate the ongoing inquiry about a current question under discussion.

These claims are quite diverse, which raises the question whether they either each describe only one aspect of the discourse function of clefts, or whether they can be understood as different manifestations of a broader discourse function.

8.2 Clefts and Givenness

This section discusses those approaches which argue that the main effect of clefts is not highlighting some constituent but marking some other constituent as given. The focus background cleft in (3), for instance marks the relative clause as given.

- (3) It was ARNE who invited all his friends.

This is in line with the existence presupposition of (3) (*Somebody invited all his friends*), which must be given in the sense that it must be part of the common ground. It follows naturally that the cleft relative clause must be given. Grubic (2015) describes the backgrounding of the cleft relative clause using the concept of a *topic situation* (c.f. Kratzer, 2020). The topic situation is the situation the sentence is about and it can be paraphrased as in (4) for the cleft in (3).

- (4) The situation that the question ‘Who invited all his friends?’ is about is a situation in which Arne invited all his friends. (c.f. Grubic, 2015:132)

The topic situation, described as *the situation that the question ‘Who invited all his friends?’ is about*, is a definite description which Grubic (2015) suggests to be the backgrounded part of the cleft. Notably, this definite description makes reference to a question, which Grubic identifies as the presupposed question under discussion, and which will be discussed in more detail in section 8.6.

Noticing the relevance of definite descriptions in clefts, Hedberg (1990) analyzes clefts based on the GIVENNESS HIERARCHY by Gundel et al. (1993), as in (5).

- (5) in focus < activated < familiar < uniquely identifiable < referential < type identifiable

Gundel et al. (1993:275)

The categories in the hierarchy describe the cognitive status that can be assumed for a referent by a speaker in a context. The cognitive status of a referent describes how this referent is represented in the mind of the addressee, e.g. in long-term or short-term memory. Gundel et al. (1993), among others, noticed that the kind of referring expression that can be used to refer to an antecedent depends on this cognitive status of that antecedent in the context. The higher a antecedent ranges on the hierarchy, the more activated it is. Depending on the category of a referent, different referring expression may be used to refer to that referent. Consider example (6), inspired by Gundel et al. (1993).

- (6) a. *Type identifiable*: I couldn’t sleep last night. **A dog** (next door) kept me awake all night.
- b. *Referential*: I couldn’t sleep last night. **This dog** (next door) kept me awake all night.
- c. *Uniquely identifiable*: I couldn’t sleep last night. **The dog** (next door) kept me awake all night.
- d. *Familiar*: I couldn’t sleep last night. **That dog** (next door) kept me awake all night.
- e. *Activated*: I couldn’t sleep last night. **That** kept me awake all night.

- f. *In focus*: I couldn't sleep last night because of the neighbor's dog. **It** kept me awake all night. (c.f. Gundel et al., 1993:276–280)

These examples show expressions (printed in boldface) that are sufficient for the respective level of the hierarchy. In (6-a), the referent of *a dog* is *type identifiable*, which means that the addressee can only access the type of object that *a dog* refers to but not a specific dog. A type identifiable referent is sufficient to be referred to with an indefinite description.

For the *referential* referent, the addressee must be able to identify a specific object either by retrieving an existing representation or accommodating the representation intended by the speaker. A referential referent is sufficient to be referred to by an indefinite *this N* phrase, as in (6-b). In this example, the speaker refers to a specific dog, which the addressee might know or can accommodate based on world knowledge.

The *uniquely identifiable* referent needs to be identified from an existing representation from the addressee's memory, or, if there is enough descriptive content, from the referring expression used. A uniquely identifiable referent is sufficient for using a definite description with the article *the*, as in (6-c). In this example, the referent of *the dog* cannot be accommodated based on world knowledge. Instead it must be retrieved either from an existing representation or from the definite description alone. For the latter case, the descriptive content must be rather elaborate. If the addressee did not know about the dog in (6-c), the more elaborate description *the dog next door* would be more suitable to refer to a uniquely identifiable referent than just *the dog*.

The *familiar* referent must be retrieved from the addressee's memory. Such a referent is sufficient to use a demonstrative *that N* phrase, as in (6-d). In this example, the addressee needs to know already that the neighbor has a dog, and cannot accommodate the referent which *that dog* refers to.

An *activated* referent, Gundel et al. (1993) state, must be on the current short-term memory, therefore *activated*. It may either be retrieved from long-term memory or must have been introduced in the immediate context. It is then sufficiently activated in order

to use pronominal *that* to refer to it. In (6-e), *that* may refer to the barking of a dog only if the addressee knows about the barking or it was mentioned before.

A referent that is *in focus* must be in the center of attention, which is most likely the topic of the preceding sentence or an expression that is likely to become the topic of the next sentence. A referent in focus is sufficient for using an unstressed pronoun, as in (6-f). The pronoun *it* can only be used here since the dog is explicitly mentioned.

In the givenness hierarchy, the higher statuses entail the lower ones in the hierarchy. Hence, a referent that is familiar is also uniquely identifiable, referential and type identifiable. It is, however, not necessarily activated or in focus. This means that a referent with a certain status in the hierarchy can, in principle, be referred to by the expression(s) of its status or lower ones. The maxim of quality, however, will make the speaker use the strongest expression. Otherwise, unintended implicatures will be derived by the addressee, as in (7).

- (7) Yesterday, I played with the neighbor's dog all day. And then, later, a dog kept me awake all night
 ~ The dog was not the neighbor's dog.

In this example, the antecedent is in focus, which allows to use an expression for type identifiable referents such as the indefinite description used in (7). The addressee concludes that the dog is not uniquely identifiable because the speaker could have used an expression that marks that. Therefore, the dog cannot be the neighbor's dog.

Based on Gundel's (1977; 1985) work, Hedberg (1990) and Hedberg (2000) investigate the activation status of the antecedent of the existence presupposition in focus-background clefts. She observes that depending on the cognitive status of this antecedent, different versions of the cleft may be used. Therefore, she argues that a truncated cleft may only be used when the antecedent of the existence presupposition is in focus, i.e. already in the addressee's center of attention, as in example (8).

(8) My heart beat fast, for I had thought that as the discoverer of the body I would be the first to be called; **but to my surprise, it was Marcel.** He stepped forward, neat, dark, debonair...

(Mary Fitt, Death and the Pleasant Voices, p. 156)

(Hedberg, 2000:899)

In this example, the antecedent is of the form *x was called*, which is explicitly mentioned in the sentence preceding the truncated cleft, and, thus, in the center of the addressee's attention. When the antecedent is only activated, however, it is not possible to use the truncated *it*-cleft, as Hedberg (2000) illustrates with the following example.

(9) I wasn't surprised by the massacre in China. [pause]

This is not Iowa we're talking about – This is a different society.

(Eric Sevareid, interview on CSPAN by Brian Lamb, 12/31/89)

(Hedberg, 2000:899)

Hedberg (2000) claims that the acceptability of (9) decreases when the *this*-cleft is replaced with the truncated cleft *It is not Iowa*. She argues that the antecedent of the existence presupposition (*we're talking about x*) is not in focus but just activated. She concludes that a truncated cleft cannot be used with an activated antecedent. Furthermore, she explains that *this* instead of *it* is used here to indicate a topic shift.

Hedberg (1990) proposes that for the full *it*-cleft, the activation status differs depending on whether the cleft is a focus background cleft or a topic comment cleft. The antecedent of the focus background cleft must be at least activated, while the antecedent for the topic comment cleft is at least uniquely identifiable.

Consider first the focus background cleft. Hedberg (1990) claims that focus-background clefs require their antecedent to be at least activated. Furthermore, she notes that they are sometimes *directly* activated and most frequently *indirectly* activated. The former is exemplified in (10).

- (10) JM: I want to ask this question: Why is this agreement so bad? I ask you.
 JG: Because our whole intention was to bring some form of democracy there; our intention was to make the Sandinistas cry uncle. **It is the contras who have cried uncle.** (Hedberg, 1990:112)

In this example, the antecedent (*x has cried*) is retrieved from the immediately preceding context where the crying is mentioned explicitly. This antecedent is activated, but not really in focus. Accordingly, a truncated cleft *It is the contras uncle* is less felicitous than the full focus background cleft.

Example (11) presents an example of an indirectly activated antecedent of a focus background cleft.

- (11) When prices go up, people come in the store and they throw the items on the counter and they blame us. Eggs go up ten cents a dozen and they act like **it's us that raised them.** Actually, we make two cents on a gallon of milk. You can't tell them that...
 [Terkel, p. 548, neighborhood merchant] (Hedberg, 1990:119)

The antecedent (*x raised the prices*) is also activated but this time indirectly as the cause of the explicitly mentioned event of prices going up. Again, the cleft cannot felicitously be replaced by the truncated cleft *It's us.* For those cases in which the antecedent is indirectly activated, Hedberg (1990) singles out different relations of the explicitly mentioned situation or event and the antecedent, which lead to different discourse functions. The cleft can represent a causal relation, as in (11), a consequent relation, as in (12), or a superlative relation, as in (13).

- (12) At nine o'clock, I was thinking of calling up the household at Lady Dormer's to ask when he was to be expected home, when the 'phone rang.'
 'At nine exactly?'
 'About nine. I might have been a little later, but not more than a quarter-past at latest. **It was a gentleman (who) spoke to me.** [...]
 [Murder at the Bellona Club, p. 38] (Hedberg, 1990:120)

(13) M: This is brandy. I want to make some brandy balls.

G: Well, you might as well just keep it, I don't really want it.

M: I thought it was something that you LIKed, better than m-, other things?

G: That?

M: Oh, **it's cognac you like the best.**

[Frederickson tapes, 1988]

(Hedberg, 1990:121)

In example (12), the antecedent *x spoke to me* is the consequent of the explicitly mentioned event of the phone ringing. In example (13), the antecedent (*you like y best*) is the superlative of an event explicitly mentioned in the context. In this case, it is already an event in the comparative, namely liking something better than other things. It can also be activated by a ‘normal’ event such as just liking something, though.

Getting back to the topic comment cleft, as in (14), Hedberg (2000) considers it to require at least a uniquely identifiable antecedent which is two levels below the sufficient activation for the focus background cleft.

(14) The federal government is dealing with AIDS as if the virus was a problem that didn't travel along interstate highways and was none of its business. **It's this lethal national inertia in the face of the most devastating epidemic of the late 20th century that finally prompted one congressman to strike out on his own.**

Gerry Studds, D-Mass, has begun to treat his constituents like foreigners – in the best way.

(Ellen Goodman, op-ed column, 5/25/87)

(Hedberg, 2000:903)

In this example, the antecedent of the existence presupposition (*x finally prompted one congressman to strike out on his own*) is retrieved from the description alone, i.e. from the cleft relative clause. This is apparent given that the cleft relative clause provides new information in topic comment clefts, such as (14). This also means that the antecedent of

the topic comment clause cannot be retrieved from the addressee's memory, which Gundel et al. (1993) claimed to be possible for other uniquely identifiable referents. Furthermore, the observation that the relative clause of a topic comment cleft always represents new information, and is presupposed at the same time seems contradictory. Hedberg (2000) adopts Prince's (1978) view that the information is new to the hearer but not to the speaker. Accordingly, the topic comment cleft is used in order to mark information as a known fact (at least to the speaker). I will get back to this issue below.

In example (14), the truncated cleft is not appropriate. Obviously, if the relative clause provides new information, it cannot be omitted. Moreover, also a focus background cleft is inappropriate in (14). This can be tested by giving the cleft in (14) the focus background intonation with the deaccented cleft relative clause, which turns out to be unsuitable. The unacceptability of the truncated cleft and the focus background cleft is correctly predicted because the antecedent is not sufficiently activated for neither of them.

Given the definition of the hierarchy, both topic comment and focus background clefts can be used when the antecedent is in focus or activated. However, Hedberg's (2000) examples show that the less activated the antecedent of the existence presupposition is, the more likely we are dealing with a topic comment cleft as opposed to a focus background cleft. Moreover, a cleft in the context of an *in focus* or *activated* antecedent is most likely a focus background cleft.

Summing up Hedberg (1990) and Hedberg (2000), truncated clefts require the antecedent of their existence presupposition to be in focus, focus background clefts require at least an activated antecedent, and topic comment clefts need at least a uniquely identifiable antecedent. Givenness seems to play a role for the acceptability of clefts. Hedberg (1990) and Hedberg (2000) derive, among many other discourse functions, the function of marking information as a known fact, the function of expressing a cause, a consequence or a superlative of a previous event. Actually, they identify even more discourse functions of clefts. I will not discuss further of them in this section, but I want to emphasize that all of them are based on different statuses of cognitive activation in the context.

There are some problems with this approach based on the givenness hierarchy. As noticeable in the description of the different activation statuses above, those are rather vague categories. In many of Hedberg's (1990) corpus examples, it is unclear or debatable what the status of the antecedent is. Hence, it is not obvious how to precisely determine the least activation on the hierarchy for the focus background cleft and the topic comment cleft. At what point is a referent in the current short-term memory in order to count as activated and not only familiar? Consider the constructed example in (15).

- (15) I couldn't sleep last night. This time, it was Mavi who didn't want to sleep.

The antecedent of the existence presupposition (*x didn't want to sleep*) can only be retrieved if the addressee knows that the speaker has baby twins that sometimes keep her/him awake at night. However, is this knowledge current short-term knowledge? It does not seem clear to me. It would be desirable to make predictions about the discourse function of clefts that is based on more clearly defined categories, or based on empirically elicited categories.

Another point of critique is that Hedberg (2000) only compares different variants of clefts to each other. I argue, however, that it is necessary to also compare clefts to unclefted sentences. In Hedberg's example in (9), for instance, it seems appropriate to replace the cleft with an unclefted sentence with focus on the former cleft pivot, as in (16).

- (16) I wasn't surprised by the massacre in China. [pause]

We're not talking about IOWA – This is a different society.

There are other examples, in contrast, in which replacing the cleft with an unclefted sentence is not appropriate, at least not with the same meaning. Hence, this approach cannot account for the difference between cleft sentences and canonical focus sentences. Given the many observed parallels between focus constructions, such as preverbal focus on Hungarian (É. Kiss, 1998), and clefts, an approach to clefts should be able to account for the differences between clefts and focus marking in canonical sentences.

Finally getting to Prince' (1978) work, her main contribution is distinguishing different kinds of clefts with respect to their discourse function. She was the first to distinguish the focus background cleft from the topic comment cleft. She states that in the focus background cleft, the pivot represents new, often contrastive, information, while the relative clause “represents known or old information, which is not marked as assumed to be in the hearer’s consciousness and which is not the theme” (Prince, 1978:896).

According to Prince (1978), the informative-presupposition *it*-cleft, which I call topic comment cleft, contains new information in the relative clause. In contrast to the focus background cleft, its relative clause is not deaccented. In such sentences, it is assumed that the addressee does not know the information conveyed by the cleft relative clause. Prince, however, points out that the cleft still marks that information as a known fact at least to the author. She uses the following example to illustrate that.

- (17) **It was just about 50 years ago that Henry Ford gave us the weekend.**

On September 25, 1926, in a somewhat shocking move for that time, he decided to establish a 40-hour work week, giving his employees two days off instead of one. (Prince, 1978:898)

This example is taken from a newspaper. Prince (1978) argues that by using the cleft, author of the article expresses that she/he was aware of the fact that Henry Ford gave us the weekend, but that she/he does not assume the reader to have that knowledge. If the author had used the non-clefted version (*Just about 50 years ago Henry Ford gave us the weekend.*), he/she would have suggested that this fact was only just discovered. Prince describes this effect as taking responsibility away from the speaker.³¹ Presenting an information as a known fact means that the speaker is not expected to defend the statement she/he is making.

Another discourse function of a topic comment cleft is “bringing the reader up to date so that s/he can appreciate the actual ‘news’ ” (Prince, 1978:902). In other words, the

³¹Whorf (1938) argues that other languages, such as Hopi, mark the responsibility of speakers towards their statements overtly. Hence, Prince (1978) considers *responsibility* as a category of grammar, for which the cleft is yet another way of marking.

topic comment cleft can function as a reminder. Moreover, it sometimes indicates a subordinating discourse relation, such as cause-effect or background, which is in line with Hedberg (1990)'s (1990) observations. In more general terms, Prince states that the cleft indicates that the conveyed information is concerned with a previous issue, instead of starting a new issue, which will be explored in more detail in the next section.

8.3 Clefts and Discourse Coherence

This section discusses the effect of clefts on discourse coherence, especially compared to the unclefted canonical sentence. Delin and Oberlander (1995) refer to this process as *de-clefting*. They mostly analyze written discourse, and investigate how de-clefting changes the discourse coherence. For them, de-clefting means taking an occurrence of a cleft in its context and replacing it with an unclefted sentence in its canonical word order, such as in (18).

- (18) a. Der Bus fuhr Schlangenlinien. Es war Peter, der am Steuer saß.
the bus drove zigzag.lines it was Peter who at.the wheel sat
'The bus zigzagged along the road. It was Peter who sat behind the wheel.'
- b. Der Bus fuhr Schlangenlinien. Peter saß am Steuer.
the bus drove zigzag.lines Peter sat at.the wheel
'The bus zigzagged along the road. Peter sat behind the wheel.'

I argue that the cleft should rather be compared to its OPTIMAL UNCLEFTED VERSION (OUV), as in (19) (c.f. Tönnis, 2018).

- (19) Der Bus fuhr Schlangenlinien. Am Steuer saß Peter.
the bus drove zigzag.lines at.the wheel sat Peter
'The bus zigzagged along the road. Behind the wheel sat Peter.'

OUV does not involve any discourse markers or other additional material that is not contained in the original cleft, but it is not required to have the canonical word order. This comparison is motivated by the assumption that the cleft competes with other non-complex sentences. With respect to discourse coherence, it should only be chosen if the

text coherence cannot be established by changing the word order since, generally, syntactic movement is less costly than using a syntactic construction such as the cleft. Therefore, if movement in the unclefted version makes the sentence more coherent and the author still chose a cleft instead, this non-canonical unclefted sentence should be the candidate which is compared to the cleft. Admittedly, the canonical sentence and the OUV will differ more frequently in German than in English since German has a more flexible word order. Hence, the OUV might not play such an important role for English clefts.

Taking the simpler version of de-cleaving, Delin and Oberlander (1992) and Delin and Oberlander (1995) provide many examples of clefts, both corpus examples as well as invented examples, which show that de-cleaving leads to incoherence. Consider the focus background cleft in example (20).

- (20) 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.** (Delin and Oberlander, 1995:471)

The de-clefted sentence in (20-b) confuses the reader. It creates the impression that the second sentence elaborates on the first sentence, which does not make sense. Delin and Oberlander (1995) claim that the focus background cleft is used to express contrast. Supporting this claim, the cleft in (20-a) is interpreted as contrasting with the first sentence, which is a more reasonable discourse relation than elaboration in this context. Therefore, the cleft can establish discourse coherence while the de-clefted sentence cannot. Hence, it seems that clefts are used to express other discourse relations than canonical sentences. Delin (1990) observed that clefts cannot express the discourse relation of *Narration*, which has been claimed to be the default discourse relation (c.f. Polanyi, 1988; Lascarides and Asher, 1993). Accordingly, clefts can be assumed to express less frequent discourse relations, for which the reader would not as easily be able to establish discourse coherence if

it was not linguistically marked. This view on discourse relations is closely related to discourse expectedness, which is discussed in section 8.4 and in subsection 10.2.1.

Delin and Oberlander (1992) derive the cleft's effect on discourse coherence from its *aspect-switching* function. They argue that the cleft has an underlying eventuality (a state or an event) and introduces a new state induced by the cleft copula, the created state. For the topic comment cleft in example (21), de-clefting illustrates how this effects the temporal relations between the eventualities introduced in the discourse.

- (21) Mr. Butler, the Home Secretary, decided to meet the challenge of the 'Ban-the-Bomb' demonstrators head-on. Police leave was canceled and secret plans were prepared. **It was Mr. Butler who authorized action which ended in 32 members of the Committee of 100 being imprisoned.** The Committee's president and his wife were each jailed for a week.

(Delin and Oberlander, 1992:282)

- (22) Mr. Butler, the Home Secretary, decided to meet the challenge of the 'Ban-the-Bomb' demonstrators head-on. Police leave was canceled and secret plans were prepared. **Mr. Butler authorized action which ended in 32 members of the Committee of 100 being imprisoned.** The Committee's president and his wife were each jailed for a week.

Delin and Oberlander (1992) claim that the temporal order of events is perceived differently comparing (21) and (22). In the cleft example (21), the event of authorizing action is interpreted as preceding the event of the cancellation of the police leave and the preparation of the plans. In the de-clefted example in (22), in contrast, the event of authorizing action is interpreted as being temporally located after those events. According to Delin and Oberlander, the cleft is used as a stativizing device, and that it is this state that is now temporally related to the preceding discourse. For the topic comment cleft, this means that the underlying event (here Mr Butler authorizing action) is backgrounded. It is not forced to be temporally related to the current reference time.

A similar effect can be observed for the following topic comment cleft.

- (23) a. Victoria turned over the body. She knew the killer's identity.
b. Victoria turned over the body. It was she who knew the killer's identity.

(Delin and Oberlander, 1995:491)

Delin and Oberlander (1995) point out that the unclefted sentence in (23-a) is very likely to be interpreted as a consequence of turning the body over. The cleft in (23-b) seems to imply that Victoria knew the identity of the killer anyway, not because of turning over the body. In the last two examples, text coherence is not disrupted by de-clefting, but de-clefting changes the preference for the interpretation of the discourse relation in the process of deriving discourse coherence.

Delin and Oberlander (1995) model the discourse effects of clefts as a product of resolving the existence presupposition of clefts. They follow van der Sandt and Geurts (1991), who propose that presuppositions are anaphoric and need to be resolved as bound to a propositional antecedent, or by accommodation. They argue that binding and accommodation take place at different places, respectively. Delin and Oberlander identify the cleft's existence presupposition as its underlying eventuality. Furthermore, they argue that the existence presupposition for the topic comment cleft usually has to be accommodated, and the presupposition of the focus background cleft is mostly bound to an antecedent. Based on these assumptions, they explain that the underlying eventuality of the cleft is integrated at different levels depending on whether it is resolved by binding or accommodation. This leads to different temporal relations for underlying eventuality of the focus background cleft compared to the topic comment cleft. The focus background cleft updates the reference time with the time of the underlying eventuality whereas the topic comment cleft does not. Focus background clefts, in contrast, can attach to higher discourse segments because anaphora binding can take place at higher levels than attachment of new discourse segments.

Hence, Delin and Oberlander (1995) derive discourse effects from the aspect-shifting feature of clefts and the integration of the existence presupposition. This leads to different

possible discourse functions. However, Asher and Lascarides (2003) showed extensively that different discourse relations have different implications for the temporal relations between eventualities. Therefore, one could take the opposite perspective of Delin and Oberlander (1995), claiming that clefts preferably express a specific set of discourse relations, excluding some of the standard discourse relations, and the temporal effects follow from those discourse relations. I will argue something along those lines in section 10.2, where I argue that clefts only express those discourse relations that are less frequent, thus less expected.

8.4 Clefts and Contrast

Seuren (1985:296) claims for English *it*-clefts that they correspond to an unclefted sentence with a contrastive accent on the former pivot. Hence, the cleft in (24-a) corresponds to the contrastive focus in (24-b). The same could be claimed for German, as indicated by the German variants in example (24).

- (24) A: I think John won the race.
- a. B: I believe it was Peter who won.
B: Ich glaube es war Peter, der gewonnen hat.
 - b. B: I believe PETER_{CF} won.
B: Ich glaube PETER_{CF} hat gewonnen.

In German, contrastive focus is in most cases marked with the same means as non-contrastive focus, namely by intonation. There is only a gradual difference between the prosodic marking of information focus and contrastive focus. This difference has been attested by measuring the pitch contour, for instance, but it is not necessarily perceived by German speakers in a sentence that is presented out of context (c.f. Hartmann, 2008).

In example (24), the contrastive intonation is marked with the index CF.

More generally, contrast is a topic that has been discussed a lot, and it is still not settled how to define it. Conditions that have been proposed for contrast are, for instance, the

existence of explicit alternatives in the context, or a closed well-defined set of alternatives to choose from (see Repp (2010) for an overview on contrast). I will refrain from defining contrast. Instead, I will concentrate on Zimmermann's (2011b) understanding of contrastive focus, which he formulates in his CONTRASTIVE FOCUS HYPOTHESIS (CFH), given in (25).

- (25) *Contrastive Focus Hypothesis:*

Contrastive focus marking on a focus constituent α is required if the speaker has reason to believe that the hearer will not consider the content of α , or the information-structural status of α as the focus of the utterance as likely to be(come) part of the Common Ground. (Zimmermann, 2011b:1167)

The CFH, obviously, does not define contrast but focuses on when contrast should be marked. The understanding of contrast that is involved in the CFH includes both the semantic as well as the pragmatic level. Contrast marking is defined as a marking of information that is unlikely or unexpected for the hearer. According to Zimmermann (2011b), contrastive focus marking is needed in order to guide the addressee towards the intended mismatch between the new information and her/his expectations. This should facilitate integrating the information into the common ground. On the semantic level, this amounts to the semantic content of this information being unlikely to be added to the common ground. A typical example in which contrast plays a role on the semantic level is correction, as in (26), or simply an unexpected reply to a question, as in (27).

- (26) A: Peter bought a car.

B: No, Peter bought a BICYCLE_{CF}.

- (27) A: What did Peter buy?

B: Peter bought a KANGAROO_{CF}.

The semantic content of B's reply in (26) is unlikely to be added to the common ground from the perspective of A, given that it contradicts with what A had just said. The answer

in (27) is simply unexpected since it is unlikely that Peter bought a kangaroo, and other alternatives would be much more expected.

On the pragmatic level of (25), Zimmermann (2011b) proposes contrastive focus to be necessary if the choice of focus, given the preceding discourse, is unexpected to the addressee. According to Zimmermann, this explains argument asymmetries between subjects and non-subjects in many West Chadic languages. In those languages, subject focus is required to be marked morphologically while non-subject focus is not. Zimmermann considers this focus marking to be a contrastive marking because it marks an unexpected development of the discourse. Usually, subjects are topics, as opposed to focus, in such languages. Hence, in terms of (25), the information-structural status of subjects as a focus is unlikely, and is, thus, predicted to be marked as contrastive. Again, this marking helps the addressee to integrate the information into the common ground even though it was unexpected. However, this does not mean that subjects are always contrastive on a semantic level in West Chadic languages. In section 10.2, I will introduce *expect-edness*, which is closely related to the pragmatic side of contrastive focus, according to Zimmermann. I will elaborate on the “information-structural status of α ”, as in CFH, investigating (un)expected discourse moves more generally, based on a question-based analysis.

Destruel and Velleman (2014) pursue a similar approach, and formulate the following claims about the contrastive component of *it*-clefts in English, based on Zimmermann’s (2011b) contrastive focus hypothesis.

- (28) *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 and Velleman, 2014:199)

Part a. corresponds to the semantic level of unlikely semantic content in the CFH. In part b., Destruel and Velleman (2014) describe what they call *meta-linguistic* expectations. They extend the unexpected focus marking in CFH to unexpected discourse developments of the discourse in general. In particular, they focus on how at-issueness affects the expectations about the further development of the discourse. They assume that the speaker is expected to address those propositions that are currently at-issue. Addressing not-at-issue propositions is unexpected. Their claims are based on the differences they observe in example (29) and (30) for corrective clefts.

- (29) a. A: This bean dip is fantastic. I really want to get the recipe. I am sure that Shannon brought it.
- b. B: It is Tim who brought it.

- (30) a. A: This bean dip is fantastic. I really want to get the recipe. I can't believe that Shannon brought it. She's normally not a very good cook.
- b. B: It is Tim who brought it. (Destruel and Velleman, 2014:207)

Destruel and Velleman (2014) hypothesize that the answer in (30-b) is more acceptable compared to (29-b). They argue that the answer in (29-b) only conflicts with the expectations about the world. The addressee A does not expect that Tim made the dip, A has just proposed to be sure that Shannon made it. The answer in (30-b), in contrast, does not only conflict with the expectations about the world but also with the expectations about the discourse because A presented the content that B corrects as not-at-issue content, by presupposing it.

Let us take a closer look at the analysis of example (30). Destruel and Velleman (2014) assume the following (not-)at-issue content for the relevant sentence in (30-a).

- (31) I can't believe that Shannon brought the dip.
- a. At-issue content: The speaker can't believe that...
 - b. Not-at-issue content: Shannon brought the dip.

The *yes, but...* test, introduced by Onea and Beaver (2009), and used to test at-issueness, shows that different replies are felicitous for disagreeing with at-issue or not-at-issue content. According to this test, a correction of not-at-issue content is felicitous with a *yes, but...*-reply while at-issue content is correctable using a *no,...-reply*. I modified the test and the example slightly in order to show the difference more clearly, as illustrated in (32).

- (32) Speaker A can't believe that Shannon brought the dip.
- a. No, speaker A does not have any trouble believing that Shannon brought the dip.
 - b. #Yes, but speaker A does not have any trouble believing that Shannon brought the dip.
 - c. Yes, I understand, but Shannon did not bring the dip.
 - d. # No, Shannon did not bring the dip.

The replies in (32-a) and (32-b) disagree with the at-issue content of the first sentence, the modified version of Destruel and Velleman' (2014) example. In this case, a reply using *no* is felicitous, and a reply using *yes, but* is not. For disagreeing with the not-at-issue content, as in (32-c) and (32-d), it is the opposite. The disagreement using *no* is infelicitous, and using *yes, but* is felicitous. Hence, the at-issueness status seems to differ as Destruel and Velleman (2014) assumed.

I argue that at-issueness is not the only feature that plays a role here. Comparing (29) and (30), another difference catches the eye, namely that there is one additional sentence in (30) (*She is normally not a very good cook*), intervening between the sentence that the reply targets, also called the antecedent, and the reply itself. I argue in my analysis of clefts in chapter 10 that the distance between the cleft and its antecedent plays an

important role as well. I suppose that this effect is relevant in example (30), too. Hence, it is not clear how much of the reduced infelicity in (29-b) is due to at-issueness and how much is due to the missing intervening sentence. I argue that this sentence makes the cleft in (29) felicitous again, as in (33).

- (33) a. A: This bean dip is fantastic. I really want to get the recipe. I am sure that Shannon brought it. She is an amazing cook.
- b. B: It is Tim who brought it.

The improvement of the felicity of (33-b) speaks against at-issueness having such a strong effect, contra Destruel and Velleman (2014). In section 9.3, I will present experimental studies testing Destruel and Velleman's claims, and empirical evidence by Destruel et al. (2019) for the hypothesis that the acceptability of French *c'est*-clefts improves, the more they indicate that an utterance runs contrary to a doxastic commitment of the interlocutor.

Finally, I want to point out that there is a limit to the strength of conflict with the discourse expectations until which the cleft is still felicitous.

- (34) a. A: Lena arrived at the party yesterday and first of all she had a Bloody Mary. Thereafter, she danced with her friend Andrea and the two of them had a lot of fun. Then, Lena went home happily.
- b. B: ?It was a Cosmopolitan she drank.

Without any additional marking, the cleft seems degraded in (34-b), even though it does express a conflict with A's expectations about the world, as well as A's expectations about the discourse. I argue, therefore, that the cleft cannot be used felicitously if the antecedent it targets is too unexpected to constitute the next move in the discourse. I will explicate this issue in chapter 10.

8.5 Clefts and Focus

The cleft is frequently compared to different kinds of focus, such as information focus, as in (35-b), contrastive focus, as in (35-c), or identificational focus as defined by É. Kiss (1998).

- (35) a. It is Maria who has a good idea.
b. MARIA_F has a good idea.
c. MARIA_{CF} has a good idea.

The cleft pivot corresponds to the focused constituent in the canonical sentence. Of course, this transformation is only applicable to focus background clefts since the pivot does not represent the focus in a topic comment cleft. Why are clefts frequently compared to focus structures? One reason is that there are contexts in which a cleft and its corresponding canonical equivalent can be used interchangeably. In the previous section, such an example was mentioned, repeated here.

- (36) A: I think John won the race.
a. B: I believe it was Peter who won.
B: Ich glaube es war Peter, der gewonnen hat.
b. B: I believe PETER_{CF} won.
B: Ich glaube PETER_{CF} hat gewonnen.

This is an example of contrastive focus, which is marked by a gradually different intonation, compared to information focus, though not distinguishable when perceived out of context (c.f. Hartmann, 2008). Example (37) presents a non-contrastive context in which the cleft and the focused canonical sentence can be used interchangeably.

- (37) Gestern hatte Maria ein heimliches Rendezvous.
'Yesterday Maria went on a secret rendezvous.'
a. Es war Linus, mit dem sie sich getroffen hat.
it was Linus with who she herself met has

'It was Linus she met.'

- b. Sie hat sich mit LINUS getroffen.
she has herself with LINUS met
'She met LINUS.'

For German, these parallels are much stronger in spoken language where focus is mostly marked by intonation. In written German, clefts have been proposed to disambiguate focus in some contexts (c.f. De Veaugh-Geiss et al., 2015; Tönnis et al., 2018). I will illustrate this with an example, based on a corpus finding on German *es*-clefts. Tönnis et al. (2018), Hartmann and Zimmermann (2007), and Lambrecht (2001), among others, argue that focus disambiguation, in general, has led to argument asymmetries with respect to focus marking in different languages. Tönnis et al. (2018) found an argument asymmetry for German clefts in a corpus study, discussed in section 9.1. They found that subject clefts, as in (38-a), are much more frequent than object clefts, given in (38-b).

- (38) a. Es ist MARIA, die Linus unterrichtet.
it is MARIA who.NOM Linus teaches
'It is MARIA who is teaching Linus.'
- b. Es ist LINUS, den Maria unterrichtet.
it is LINUS who.ACC Maria teaches
'It is Linus who Maria is teaching.'

Tönnis et al. (2018) suggest that this difference arises due to the lack of intonation as a means to mark focus in written German. Consider example (39).

- (39) a. Maria unterrichtet LINUS.
Maria teaches LINUS
'Maria is teaching LINUS.'
- b. MARIA unterrichtet Linus.
MARIA teaches Linus
'MARIA is teaching Linus.'
- c. Es ist MARIA, die Linus unterrichtet.
it is MARIA who Linus teaches
'It is MARIA who is teaching Linus.'

One means to mark focus in spoken German is intonation, as marked in capital letters in (39), which is however not available in written German. If the context is ambiguous with respect to the current question (CQ)³², the different foci of (39-a) and (39-b) cannot be distinguished without intonation. In such a situation, the writer might still want to mark focus in some other way in order to help the reader. One option would be the cleft, which marks focus unambiguously in most cases. Some exceptions will be discussed in the next section. In example (39-c), however, the focus is not ambiguous even if intonation is missing.

Tönnis et al. (2018) further argue that the cleft structure is only necessary if the intonation would otherwise diverge from the default intonation. The default intonation is exemplified in (39-a), where the main accent falls on the object. Tönnis et al. assume that without a cue from the context, the reader will assume the default intonation. As (39-a) indicates, this default intonation coincides with object focus marking. Subject focus intonation, in contrast, diverges from the default intonation, as (39-b) indicates. Therefore, if the writer wants to make sure that the reader interprets the focus correctly, she/he can use a cleft structure instead. Focus is now again marked unambiguously by default, as in (39-c). Given that clefting only makes a difference with respect to the default intonation of subjects, it is expected that there are more subject clefts than object clefts. This provides evidence in favor of clefts marking focus in German.

Focus marking cannot be the only purpose of using a cleft in German, though. If it was, we could not explain why a cleft sentence would ever be used in spoken German. It seems that intonation should always be the better, because less complex, option. Therefore, clefts should always be degraded in spoken German. However, there are occurrences of clefts in spoken German as well, which need to be accounted for.

Another aspect that the focus approaches need to explain is why the cleft has a stronger existence inference than focus. There are some approaches that do assume focus to give rise to an existence presupposition (e.g., Geurts and van der Sandt, 2004). However, this inference has been shown to be cancellable. For the cleft, in contrast, it seems nearly

³²Focus was defined as indicating the CQ, given in Definition 8 in section 3.4.

impossible to cancel the existence inference. This is illustrated by the following examples, which violate existence.

(40) a. NOBODY had a good idea.

b. #It was nobody who had a good idea.

(41) a. Not even MARIA_F had a good idea.

b. #It was not even Maria who had a good idea.

These examples show that phrases involving a negative quantifier cannot be clefted while they can be focused, as mentioned already in section 5.3. Thus, the existence presupposition of clefts seems to be much stronger than it is for focus, which needs to be explained when clefts and focus are treated as equivalents.

8.6 Clefts and Questions

It has frequently been claimed that the function of clefts is (related to) implicitly marking which question is addressed (e.g., Delin and Oberlander, 1995; Fischer, 2009; Hedberg, 1990; Seuren, 1985; van Kuppevelt, 1991; Velleman et al., 2012). This description is quite similar to what was proposed for the function of focus marking in chapter 3. Furthermore, the previous section discussed the relation of clefts and focus in detail. This section is concerned with how different approaches integrate the relation of the cleft to questions in discourse.

One of the first approaches which related the cleft to a question was the one by Seuren (1985), who provides a syntactic-semantic analysis of the cleft, which assumes that a cleft raises a question that it also answers. The details of the syntax he proposes are presented in section 5.3. Fischer (2009) claims that the cleft provides the missing information to an open proposition expressed by the relative clause. He assumes this to happen on a second semantic layer that he calls *specifying structure*. On this layer, the missing information of

the open proposition is specified. Seuren includes exactly this open proposition into the syntactic representation of the cleft.

Seuren (1985) argues that the focused constituent in the cleft typically answers the question that is implicit in the topic constituent, which he takes to be the cleft relative clause (Seuren, 1985:297). As discussed in section 5.3, he assumes the cleft to involve a question-raising NP and a question-answering sentence that is related by a specifying *be*. For a canonical sentence involving a contrastive focus, Seuren assumes the same underlying structure as for the cleft.

Seuren's (1985) approach takes into account the context of the current discourse, by requiring that the cleft answers a question about an unidentifiable discourse referent. This is basically the requirement for an appropriate context for a question in general. According to Seuren's analysis, the question which the cleft answers is always raised by the cleft itself. I argue that the cleft is rather marking than raising the question that it answers. Whenever a cleft is felicitous, the question that it addresses must have been raised in some way in the discourse. The question does not have to be raised explicitly, but, for instance, as a potential question, as will be discussed extensively in section 10.2. Besides these syntactic-semantic approaches, there are several discourse-related approaches, which analyze clefts with respect to how they relate to questions that are currently discussed in the discourse. My own analysis will eventually follow this line of research (see section 10.2).

One of the most important question-based analyses is Velleman et al. (2012). They conceptualize clefts as devices to terminate an ongoing inquiry (IT-constructions), as in example (42) (a slightly adapted version of Velleman et al., 2012:449).

- (42) 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.

The cleft, uttered by C, is assumed to terminate the ongoing inquiry about the question of what Mary ate. Other IT-constructions are exclusives, according to Velleman et al.'s (2012) account. Their analysis of clefts is embedded in the Question Under Discussion (QUD) framework, as proposed by Roberts (2012). This framework is presented in section 4.3 of this thesis. Velleman et al. argue that the cleft maximally answers the Current Question (CQ), and thereby terminates the ongoing inquiry. Clefts and exclusives differ only with respect to the at-issueness of their meaning components. The cleft presupposes maximality whereas exclusives assert maximality.

Velleman et al. (2012) assume that the cleft introduces a focus-sensitive operator (in the sense of Beaver and Clark, 2008) into the logical form of the sentence, which accounts for the exhaustive component of the cleft. The focus sensitivity of exhaustivity is indicated by the following example.

- (43) a. It was JOHN'S eldest daughter who attended the party.
 \sim Nobody else's eldest daughter attended the party.
- b. It was John's ELDEST daughter who attended the party.
 \sim No other daughter of John's attended the party.

Depending on the focus in the cleft pivot, the sentence evokes a different exhaustivity inference. Other approaches to clefts cannot account for this difference because they derive the exhaustivity inference merely based on the cleft pivot, no matter where the focus inside the cleft pivot lies. Hence, they would incorrectly predict for both (43-a) and (43-b) that nobody else than John's eldest daughter attended the party, which would be too strong.

In order to solve this problem, Velleman et al. (2012) define a cleft operator that depends on the focus in the pivot. In particular, they follow Beaver and Clark (2008) by assuming that focus sensitivity means being related to the current question under discussion (CQ). Moreover, they assume that the *it*-cleft is congruent to the same CQ as the canonical counterpart. I made the same assumption in section 4.4. There, the CQ was defined as the question that the speaker aims to address with an utterance. An utterance is

focus congruent to this CQ if the focus alternatives of this utterance contain the question alternatives of the CQ³³, assuming an alternative semantics approach to questions (as described in 2.3). Accordingly, Velleman et al. assume both (44-a) and (44-b) to be focus congruent to the question in (44).

(44) Which of John's daughters attended the party?

- a. It was John's ELDEST daughter that attended the party.
- b. John's ELDEST daughter attended the party.

Crucially, the cleft does not address the more general question *Who attended the party?*, in contrast to the predictions of other approaches. The definition of the cleft operator is based on the adapted focus-sensitive operators MIN_C and MAX_C , as in (45-a) and (45-b), originally defined by Coppock and Beaver (2011:199).

(45) a. $\text{MIN}_C(p) = \lambda w. \exists q \in CQ_C [q(w) \wedge (q \geq_C p)]$

“There is a true answer at least as strong as p .”

b. $\text{MAX}_C(p) = \lambda w. \forall q \in CQ_C [(q >_C p) \rightarrow \neg q(w)]$

“No true answer is strictly stronger than p .” (Velleman et al., 2012:451)

The subscript C indicates the current context, which contains a current question CQ_C and a partial ordering over the alternatives in CQ_C , indicated by \geq_C and $>_C$. In principle, the alternatives could be ordered with respect to various scales. However, it is most likely an entailment scale. Hence, $q >_C p$ means q entails and is distinct from p in C . The operators in (45-a) and (45-b) each apply to a proposition p , which will turn out to be the *prejacent* of the cleft. The prejacent of (44-a) is the following.

(46) John's eldest daughter attended the party.

Velleman et al. (2012) argue that the cleft asserts MIN_C and presupposes MAX_C , which leads to their definition of the cleft operator CLEFT_C in (47).

³³See the definition of focus congruence in section 3.4.

$$(47) \quad \text{CLEFT}_C = \lambda w. \lambda p : \text{MAX}_C(p)(w). \text{MIN}_C(p)(w) \quad (\text{Velleman et al., 2012:452})$$

This operator takes a world and the underlying prejacent p , presupposes that there is no true answer in the alternatives of CQ_C that is strictly stronger than p , and asserts that there is a true answer in the alternatives of CQ_C which is at least as strong as p .

Since MIN_C and MAX_C are dependent on the alternatives in CQ_C , the difference between (43-a) and (43-b) is predicted. Due to different narrow focus in the cleft pivots, their focal alternatives differ. Therefore, they are focus congruent to different questions, as indicated in (48-a) and (48-b).

- (48) a. Context C1: *It was JOHN'S eldest daughter who attended the party.*

CQ_{C1} : Whose eldest daughter attended the party?

- b. Context C2: *It was John's ELDEST daughter who attended the party*

CQ_{C2} : Which of John's daughters attended the party?

One could wonder why the CQs in (48) are not cleft questions, i.e. *Whose eldest daughter was it who attended the party?* and *Which of John's daughters was it who attended the party?*, respectively. Such CQs would be possible, but it is entirely unclear what the meaning and the function of cleft questions is. Therefore, Velleman et al. (2012) make the assumption that the focus alternatives of a cleft are derived in a similar way as those of exclusives. Beaver and Clark (2008) and Coppock and Beaver (2011) assume that the alternatives for a sentence containing an exclusive are calculated on the basis of the prejacent without the exclusive. Velleman et al. do the same for the cleft by assuming that the focus alternatives are calculated on the basis of the underlying canonical sentence.

Thus, the focus congruence to the CQs in (48) is given.

The cleft in (48-a) is represented as in (49).

$$(49) \quad [\![\text{It was JOHN'S eldest daughter who attended the party}]\!]$$

$$= \text{CLEFT}_{C1}(w)([\![\text{John's eldest daughter attended the party}]\!])$$

This means that it is presupposed, by MAX_{C1} , that no true alternative in CQ_{C1} (*Whose eldest daughter attended the party?*) is stronger than the prejacent *Johns eldest daughter attended the party*. In other words, all proposition of the form $x \text{ eldest daughter attended the party}$ for which x includes John and at least one more individual are false in $C1$. Furthermore, the cleft in (49) asserts that the alternatives of CQ_{C1} contain at least the prejacent *John's eldest daughter attended the party*.

Eventually, Velleman et al. (2012) derived exhaustivity of the cleft sentence depending on the current question. Strictly speaking, their account does not derive an exhaustivity inference but the inference that the cleft provides a maximal answer to the CQ, which is the same in many cases but not all the cases. Based on this analysis, Velleman et al. derive the discourse functions of clefts depending on the current question. They propose that a speaker who uses a cleft and thereby provides a maximal answer to the CQ, aims for terminating the inquiry about this CQ. They interpret the cleft as one of possibly many constructions that are used to mark the end of an inquiry. As mentioned above, exclusives are analyzed as IT-constructions, as well.

Velleman et al. (2012) point out a problematic example for their own approach, which is illustrated in example (50) and (42), the latter repeated in (51).

(50) A: What did Mary eat?

B: ?It was a PIZZA that Mary ate.

(51) 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.

As pointed out above, the approach by Velleman et al. (2012) correctly predicts the acceptability of the cleft uttered by C in example (51). The unacceptability of the cleft in example (50), however, poses a problem for their approach. Nothing in their analysis predicts when an inquiry is allowed to be terminated, and when it is not allowed. Velleman

et al. suggest as a solution that an inquiry may only be terminated once it has extended to a certain degree. If it is too short, the marking of the termination is not necessary, and the cleft is predicted to be infelicitous. According to their intuitions, the inquiry in (50) is not extended enough. This is not a very convincing argument, though. One could equally well argue that an extended inquiry is more likely to terminate than a short one, and, therefore, the latter needs more marking than the former. This would make the opposite prediction. Thus, the extent of the inquiry per se cannot be the crucial factor, and neither can maximality of the provided answer to CQ.

As mentioned above, Grubic (2015) also incorporates the QUD into the analysis of clefts via the concept of a topic situation. She identifies the cleft relative clause as backgrounded part of the sentence, which refers to the topic situation. She paraphrases the cleft in (52-a) as in (52-b), repeated from section 8.2.

- (52) a. It was ARNE who invited all his friends.
- b. The situation that the question ‘Who invited all his friends?’ is about is a situation in which Arne invited all his friends.

The question *Who invited all his friends?* is identical to the CQ that Velleman et al. (2012) assume to be involved in the composition of the cleft meaning. Grubic (2015) assumes that this question is presupposed, while Velleman et al. suggest a CQ provided by the context. This might amount to the same predictions but seems to also lead to similar problems. If we assume the cleft to in (50) and (51) to be described as in (53), why is this allowed to refer to this situation in example (51), but not in example (50)?

- (53) The situation that the question ‘*What did Mary eat?*’ is about is a situation in which Mary ate a pizza.

I will get back to the examples (50) and (51) in section 10.4, and will provide an alternative explanation for the differences. Anticipating somewhat, I will address the question how the CQ that is addressed by a cleft relates to the preceding discourse.

In her dissertation, Sheil (2016) discusses the Scottish Gaelic PROPOSITIONAL CLEFT (PC). In (54), there is an example of a propositional cleft.

- (54) 'S ann a bhios iad a' toir cuideachadh is comhairlean COP in.3MSC.SG that be.FUT 3PL PROG give help and advice seachad 'na rígheachd.
away in.3MSC.SG.POSS kingdom
'*They give out help and advice in the kingdom.*' (Sheil, 2016:3)
lit. *It is that they give out help and advice in the kingdom.*

This construction does neither exactly match the *es*-cleft in German nor the *it*-cleft in English since it lacks the clefted constituent. Moreover, it expresses broad focus, which the *es*-cleft never does.

Sheil's (2016) approach is, nevertheless, relevant to this thesis given that she proposes an analysis which is very much related to what I am proposing in section 10.1. She observes that the propositional cleft frequently provides unexpected information, it may be used contrastively, as a narrative device to progress the story, or to signal the termination of a sub-plot (the latter is reminiscent of Velleman et al., 2012). Her analysis subsumes all of these observed functions as resulting from the requirement for the PC to create a new line of inquiry. Her approach is embedded in a model of discourse trees as proposed by Büring (2003), where a new line of inquiry means moving to a different branch in the discourse tree. This move implies that it cannot be the immediate QUD that is addressed by the PC which is similar to what I propose for the *es*-cleft in German.

Sheil's hypothesis, however, differs with respect to what questions are allowed to be addressed by the cleft. She mentions super-questions or a different sub-question of a super-question (i.e. a sister question). For the contrastive PC, she argues that a sister question is addressed and the current line of inquiry is closed. For the non-contrastive narrative PC, she argues that it addresses the question *What happened (next)?*. The German *es*-cleft cannot address a super-question. The *es*-cleft has a narrow focus on the cleft pivot, which makes it incompatible with the rather broad question *What happened next?*. Addressing a sister question might be possible for contrastive uses of the *es*-cleft.

8.7 Summary

This section showed that a lot of observations were made about clefts in discourse. They were rather made for English *it*-clefts, but it seems that equivalent observations can be made for German *es*-clefts. Those observations lead to a range of hypotheses about the discourse function or discourse effect of clefts. These range from marking given material (Destruel and Velleman, 2014) to marking the clefted material as dissociated from the main timeline (Delin and Oberlander, 1992). These are quite diverging functions.

Nevertheless, it is not convincing to me that clefts have a range of unrelated functions depending on the context. In the process of interpretation, how is the addressee supposed to know which function the cleft should have in each situation? I suggest to reduce those functions to ideally one or at least to fewer functions. The different observed effects could be just different manifestations of one or a few discourse functions of clefts. Hence, this thesis aims to trace back those effects to ideally one source. In section 10.2, I will identify this source (or one of the main sources) as marking unexpectedness discourse moves, i.e. addressing unexpected questions given the previous discourse.

Chapter 9

Empirical Approaches to Clefts

This chapter will provide an overview of various empirical approaches towards cleft sentences, covering different languages including several studies on German. Section 9.1 presents a corpus study on German clefts, which my colleagues and I conducted. Furthermore, some complementary insights from other corpus studies on clefts in different languages are included. In section 9.2, I will summarize the main experimental studies on exhaustivity in clefts, and will also discuss a study by De Veauh-Geiss et al. (2018b), which tests German clefts, in more detail. Section 9.3 presents studies that investigate contrastivity as the main feature of clefts, focusing on French and English.

9.1 Corpus Studies

A good starting point for empirically investigating cleft sentences in German is via a corpus study. Wedgwood et al. (2006) pointed out that clefts have often been analyzed in unnatural contexts (or even without a context), which failed to capture clefts as part of a discourse. I want to add that more generally most of the theoretical approaches to clefts use rather unnatural instances of clefts, such as (1).

- (1) Es ist John, der getanzt hat.
it is John who danced has
'It is John who danced.'

The cleft in (1) is not at all unacceptable. It is rather such a cleft is hardly ever found in natural conversations or texts in German. As we will see below, clefts usually have much more complex pivots than just a proper name (in case of a focus background cleft) or much more complex relative clauses (in case of a topic comment cleft). Also experiments tend to use unnatural clefts, such as the stimuli used in De Veaugh-Geiss et al. (2018b) (the experiment will be discussed in section 9.2 below). Example (2) shows a stimulus used in the experiment, which was also presented with almost no context.

- (2) Es ist Ben, der einen Pullover angezogen hat.
it is Ben who a sweater put.on has
'It is Ben who put on a sweater.'

In an experimental setting, participants might have trouble forming a judgment on (2) because of the missing context and the unusual kind of cleft. Especially, it might turn out very difficult to find an appropriate context for those clefts. Building a theoretical analysis of clefts on constructed, more or less unnatural, examples of clefts, does not have to be problematic per se. However, it still raises the question what kinds of clefts we do find in German texts or conversations. Do they have specific features that distinguish them from those examples frequently used in the literature or experiments?

Some of the first corpora of clefts were build by Prince (1978), Delin (1990), and Hedberg (1990). They, first of all, provided those missing examples of naturally occurring clefts for English *it*-clefts. Furthermore, their explorative studies analyzed clefts in a wider context taking into account the discourse function of clefts in English. Prince (1978) established the distinction between topic comment clefts and focus background clefts based on examples from her corpus. Hedberg (1990) provides many examples for which she analyzes the activation status of their antecedents in the discourse. These corpus studies provide valuable examples of clefts and interesting observations that lead to many hypotheses on clefts in discourse. However, those claims have neither been tested empirically, nor formally implemented.

I will exemplarily present some observations made in those corpus studies, most of which are discussed in more detail in chapter 8, above. Prince (1978) observed that *it*-clefts mark the information conveyed by the cleft relative clause as a known fact. It must be known to the speaker but not necessarily to the addressee. Example (17), repeated in (3), illustrates how this is possible.

- (3) **It was just about 50 years ago that Henry Ford gave us the weekend.**

On September 25, 1926, in a somewhat shocking move for that time, he decided to establish a 40-hour work week, giving his employees two days off instead of one.

(Prince, 1978:898)

Prince (1978) argues that the author of this example, which was taken from a newspaper, wants to suggest to the reader that it is a known fact to the author of that article that Henry Ford gave us the weekend. It is not something the author her-/himself discovered. The clefted utterance just serves as a reminder to the readers. She further observes that clefts are capable of establishing a cause-effect relationship, where an unclefted sentence could not. An example for such a case is the following.

- (4) Here ... were the ideas which Hitler was later to use ... His originality lay in his being the only politician of the Right to apply them to the German scene after the First World War. **It was then that the Nazi movement, alone among the nationalist and conservative parties, gained a great mass following and, having achieved this, won over the support of the army, the president of the republic and ... big business – three “long-established institutions” of great power.** The lessons learned in Vienna proved very useful indeed.

(Prince, 1978:902)

Prince claims that the cleft clearly establishes a link between Hitler's originality and the “success” of the Nazi movement (cause-effect). She argues that when the cleft was replaced by the canonical sentence *Then the Nazi movement...*, this effect is lost. She did not make it a general claim, though. I would consider it a promising endeavor to

investigate the effect of replacing natural occurrences of clefts with their unclefted version and investigate the effect in the discourse or in text coherence. More generally, Prince's examples show that clefts can be quite complex, as opposed to (1) or (2).

Corpus studies have, furthermore, found counter evidence against theoretical claims made in the literature. One such claim is the one made by Percus (1997) that clefts are incompatible with additive focus particles (this claim was discussed in section 5.3). Pavlović (2019), to the contrary, presents a corpus study on German clefts showing that there are occurrences of German *es*-clefts that contain additive focus particles such as *auch* ('also') and *vor allem* ('mainly'), as illustrated in (5) and (6).

- (5) Es ist **auch** ihre Perspektivlosigkeit, die viele Jugendliche zur Flasche greifen lässt.
'*It is also their lack of perspective that makes many teenagers reach for the bottle.*'
(RHZ04/APR.20135 Rhein-Zeitung, 23.04.2004; Jugend braucht mehr Chancen)

Pavlović (2019:82)

- (6) Es ist **vor allem** das Wetter, das uns bis jetzt einen Strich durch die Rechnung macht.

'*It is mainly the weather that has messed up our plans so far.*'

(NUZ06/JUN.00081 Nürnberger Zeitung, 01.06.2006; Umsatzzrückgang beim Einzelhandel im April - Wetter verregnete das Geschäft) Pavlović (2019:82)

Pavlović (2019), furthermore, annotated exhaustivity of the clefts in her study and found that those clefts with non-exclusive focus particles were much less often exhaustive than those with exclusive particles, such as *nur* ('only'). However, she also points out that exhaustivity is a particularly difficult category to annotate. Tönnis et al. (2018) came to the same conclusion, which will be discussed below.

In the following, I will present a corpus study that I conducted with my colleagues and that is published in Tönnis et al. (2018). The aim of the study was to identify factors that facilitate the use of an *es*-cleft in German. For this purpose, both the cleft itself as well as the context in which the cleft appears were analyzed. The study was based

on a random sample of 300 clefts taken from a sub-corpus of the DeReKo.³⁴ Some texts needed to be excluded because they were not fully accessible. The annotation of some of the properties required a lot of context before and after the cleft sentence, which was not accessible for all texts in DeReKo. Furthermore, *Wikipedia* articles were excluded because of the possibility of different authors for adjacent paragraphs of a text. In case the cleft was written by a different author than the preceding context, the interaction of the cleft and this context cannot be properly analyzed since it is not anymore a coherent text written by one author.

It was an explorative study in which a large number of features were annotated by three annotators.³⁵ The categories that were annotated aimed at exploring three broad questions: (i) Which formal features are common and which are less common for natural occurrences of clefts? (ii) Are naturally occurring clefts exhaustive? (iii) Do naturally occurring clefts express contrast?

Question (i) was approached by annotating syntactic features, such as the number, case, and grammatical function of the cleft relative pronoun. Also, the tense of the cleft matrix clause and the cleft relative clause were annotated, as well as several features of the cleft pivot, such as definiteness or anaphoricity. For most of the clefts from the cleft sample, those formal categories were unproblematic in the annotation process and the inter-annotator agreement was high.

The main finding of this corpus study was that subject clefts are significantly more frequent than object clefts. I will now present the main result of the study, and will then get back to question (ii) and (iii).

Example (7) is a subject cleft from our cleft sample, and example (8) is an object cleft from our sample.

³⁴Das Deutsche Referenzkorpus DeReKo (The Mannheim German Reference Corpus)
<https://www1.ids-mannheim.de/k1/projekte/korpora/>, Leibniz-Institut für Deutsche Sprache, Mannheim.

³⁵The annotators were my colleagues Lea Fricke, Alexander Schreiber, and myself. I thankfully appreciate their support and contribution to this corpus study. My research has greatly benefited from their input.

- (7) Es war Alt-Bundespräsident Roman Herzog, der zum 50-jährigen
 It was former.president Roman Herzog, who.NOM.SG on.the 50th
 Jubiläum eine internationale Neuorientierung der Stiftung anregte.
 anniversary a international re-orientation of.the foundation suggested
'It was the former president Roman Herzog who suggested an international re-orientation of the foundation on the 50th anniversary.'

(Die Zeit [Online-Ausgabe], 19.07.2007; Noble Töne, enttäuschter Nachwuchs)

- (8) Es ist der Aufsteiger, den Balzac mit immer neuen charakterlichen
 It is the climber who.ACC.SG Balzac with constantly new character
 Merkmalen porträtiert, [...] features portrays
'It is the (social) climber who Balzac portrays with constantly new character features.'

(Frankfurter Rundschau, 15.05.1999, p.3, Ressort: ZEIT UND BILD; Zum 200. Geburtstag von Honoré de Balzac)

The cleft in (7) is a subject cleft because the relative pronoun is the subject of the cleft relative clause. The cleft in (8) is an object cleft because the relative pronoun is the object.

Tönnis et al. (2018) found mostly subject clefs, a few object clefs and even fewer adverbial clefs. The latter were excluded from the analysis. We found that 91% of the cleft occurrences were subject clefs, and 9% were object clefs. However, this result cannot be interpreted on its own because it depends on the general frequency of subjects in contrast to objects. If subjects were in general nine times more frequent than objects, it would not be surprising to find nine times more subject clefs than object clefs. Therefore, we built a comparison corpus of 200 randomly chosen sentences again from the DeReKo³⁶, and counted all subjects and objects in those sentences. The absolute numbers are displayed in Table 9.1, taken from Tönnis et al. (2018:235).

³⁶<https://www1.ids-mannheim.de/kl/projekte/korpora/>

	n _{cleft}	n _{comp}
Subjects	249	192
Objects	24	93

Table 9.1: Absolute numbers n_{cleft} for cleft sample and n_{comp} for the comparison corpus.

Tönnis et al. (2018) analyze the data in two ways. For the first analysis, we calculated the relative frequencies from the absolute numbers in Table 9.1 in order to test them for significant difference. This method assumes that every grammatical argument is equally likely to be clefted if present, independent of the sentence it belongs to. It ignores the fact that various grammatical arguments are unevenly distributed in sentences. For instance, there might be just one subject but two objects in a sentence, but never the other way around.

The second analysis solves this problem by calculating the probabilities to be clefted for each subject and object in the comparison corpus. This time, it was this probability that was compared to the relative frequencies of subjects and objects in the cleft corpus. This probability is now dependent on the number of arguments in the respective sentence.

In Tönnis et al. (2018), we illustrate our method by means of the following example. Consider a sentence of the form S-V-O and a sentence of the form S-V-O-O. For the first sentence, the probability that the subject is clefted is 0.5, as only S and O, not V, can be clefted. For the second sentence, this probability for the subject to be clefted drops to 0.33, as there are now three possible grammatical arguments to be clefted. Using this procedure, we calculated the probability of being clefted for each subject and object in each sentence from the comparison corpus and calculated their average p_{cleft} over all sentences.

This approach assumes that each sentence is equally likely to become a cleft. It ignores the fact that probably a sentence containing just a subject and an intransitive verb is less likely to become a cleft than a sentence containing more arguments. This intuition is supported by a corpus study conducted by Prince (1978) on English *it*-clefts and *wh*-clefts.

clefts. For *it*-clefts, she found that on average the cleft relative clause is nearly twice as long as the pivot, and the content of the relative clause made up 66% of the corresponding canonical sentence (Prince, 1978:886). Still, each of the approaches by Tönnis et al. (2018) can be seen as a useful simplification because the aspects they ignore are independent of each other.

The relative frequencies, f_{cleft} and f_{comp} , and the average probability p_{comp} in the comparison corpus are displayed in Table 9.2, taken from Tönnis et al. (2018:236).

	f_{cleft}	f_{comp}	p_{comp}
Subjects	0.91	0.67	0.76
Objects	0.09	0.33	0.24

Table 9.2: Relative frequencies of subjects and objects in the cleft sample (f_{cleft}) and the comparison sample (f_{comp}), and the average probability (p_{comp}) of subjects and objects in the comparison sample.

For the first approach, f_{cleft} and f_{comp} are compared. For the second approach, f_{cleft} and p_{comp} are compared. Tönnis et al. (2018) show that both analyses of the data yield that subject clefts occur significantly more often than object clefts, also when compared to the general frequency of subjects and objects. For the first approach, we tested f_{cleft} of subjects and objects and f_{comp} from the comparison corpus for significant deviation using a χ^2 -test. The test showed that subject clefts were significantly more frequent in the cleft sample ($p < 0.01$). For the second approach, we tested f_{cleft} and p_{comp} of subjects and objects for significant deviation. A t-test showed that subject clefts are significantly more frequent in the cleft sample than predicted by p_{comp} ($p < 0.01$).

Such an argument asymmetry has been attested for clefts in many other languages (see Carter-Thomas 2009, Reichle 2014 for French; Collins 1991, Roland et al. 2007 for English; an Skopeteas and Fanselow 2010 comparing English, French, Georgian, and Hungarian). Tönnis et al. (2018), Hartmann and Zimmermann (2007), and Lambrecht (2001), among others, argue that this asymmetry is a consequence of an asymmetry in focus marking

depending on the grammatical function. De Veaugh-Geiss et al. (2015) and Tönnis et al. (2018) hypothesize that the cleft structure is a good device to unambiguously mark focus in written German. Tönnis et al. point out that this is particularly necessary for subjects because, in contrast to objects, subjects do not receive prominence by default. A subject cleft, in contrast, moves the subject in a position where it receives the main accent per default. A more elaborate version of this argument can be found in section 8.5. This hypothesis still needs to be tested empirically, which I will leave for future research.

I will now get back to the other questions that Tönnis et al. (2018) aimed to investigate, namely whether clefts express exhaustivity or contrast. In the annotation process, it turned out that there was a very low inter-annotator agreement with respect to the decision whether the respective cleft expresses exhaustivity. One such case is presented in (9).

- (9) Obgleich *7xjung* mit dem Zusatz „Künstlerische Ausstellung über den Holocaust“ daherkommt, ist hier keine Ausstellung im herkömmlichen Sinn zu sehen. Klassische museale Exponate aus der NS-Zeit finden sich kaum in den lichten, reduziert gestalteten Räumen unter der Stadtbahntrasse unweit des S-Bahnhofs Bellevue. Stattdessen bilden sieben Themenräume, die Namen tragen wie „Mein Sport“, „Mein Laden“ oder „Meine Stadt“, die heutige Alltagswelt von Jugendlichen nach.

Kissenberge und NS-Zeit

Es sind harte wie subtile Brüche, mit denen Terror und Bedrohung in die realistisch gestaltete Aktualität der Räume einfallen: In „Mein Zimmer“ etwa stapeln sich Kissenberge auf dem Teppichboden, an der Wand hängt ein Bushido-Poster. Und doch geht es auch hier um Geschichte, um die NS-Zeit, um Antisemitismus und Rassismus

(T12/APR.00569 die tageszeitung, 05.04.2012, S. 23; Weil Anderssein kein Spiel ist)

‘Although 7xjung promotes its exhibition with the slogan “Art exhibition on the Holocaust”, it is not an exhibition as such. Conventional museum exhibits from

the NS-times are barely found in the bright, minimally decorated rooms below the tram trail close to the tram station Bellevue. Instead, the everyday life of teenagers is recreated in seven thematic rooms, which are titled, e.g., “My Sport”, “My Shop”, or “My City”.

Mountains of pillows and NS-times

It is hard as well as subtle breaks with which terror and threat invade the authentically decorated actuality of the rooms: In “My Room”, for instance, mountains of pillows are piling up on the carpet, a Bushido poster is hanging on the wall. And still, we are dealing with history here, with the NS-times, with antisemitism and racism’

The *es*-cleft in (9), repeated in (10), is predicted to have the exhaustivity inference in (11).

- (10) Es sind harte wie subtile Brüche, mit denen Terror und Bedrohung in
It is hard as.well.as subtle breaks with which terror and threat in
die realistisch gestaltete Aktualität der Räume einfallen
the authentically decorated actuality of.the rooms fall.upon
‘It is hard as well as subtle breaks with which terror and threat invade the authen-
tically decorated actuality of the rooms’
- (11) Terror and threat invade the authentically decorated actuality of the rooms with
nothing else than hard and subtle breaks.

It is not an obvious decision whether (10) has the exhaustivity inference in (11) or not. In the annotation process, this was the case for many occurrences of clefts, which leads to the conclusion that this category is not suitable for annotation. Instead, it should rather be tested in an experiment. Such experiments will be presented in the next section.

Regarding the category *contrast*, we defined some categories for the annotation that we expected to be related to contrast as well as clearly annotatable. Those included the existence of explicit alternatives in the context, negation in the context, adversative or concessive adverbials and adversative or concessive prepositional phrases. The adjectives

and prepositions are summarized in (12), all meaning something like *nevertheless*, *in contrast to* or *however*, etc.

(12) a. *Adversative adverbials:*

hingegen, dagegen, jedoch, doch, dennoch, indes(-sen), allerdings, nur, vielmehr, demgegenüber, stattdessen, aber

b. *Concessive adverbials:*

trotzdem, trotz allem, dennoch, dessen ungeachtet, gleichwohl, immerhin, allerdings, sowieso, nichtsdestotrotz

c. *Prepositions:*

trotz, entgegen

None of these categories, however, turned out to be even remotely significant, compared to other sentences. Determining whether a cleft is contrastive or not without those clear categories was equally difficult as for exhaustivity and, I propose, is also better approached with experimental methods. Those will be discussed in section 9.3.

9.2 Experiments on Exhaustivity

Most experiments on clefts are concerned with testing the exhaustivity inference of clefts in different languages. This section will give a broad overview over the most important experiments and findings on exhaustivity in cleft sentences. Just as a reminder, example (13) illustrates the exhaustivity inference that the cleft is assumed to have in some way or the other (for a theoretical discussion see chapter 7).

(13) It is Arne who is smiling.

~> Nobody other than Arne is smiling.

As mentioned in chapter 7, the status of this inference is debated, as well as its existence in general. Accordingly, the following questions have been of interest for empirical research on the exhaustivity inference of clefts.

1. Is the inference semantic or pragmatic?
2. Is the inference at-issue or not-at-issue?
3. When does the exhaustivity inference arise, and when does it vanish?

For this purpose, the cleft is compared to other structures for which those questions are settled, and which can, thus serve as baselines, such as those in (14).

- | | | |
|------------|------------------------------------|----------------------|
| (14) a. | ARNE is smiling. | NARROW FOCUS |
| | b. Only ARNE is smiling. | EXCLUSIVE |
| | c. The one who is smiling is Arne. | DEFINITE PSEUDOCLEFT |

Exclusives are semantically, hence strongly, exhaustive. Focus, in contrast, shows a much weaker exhaustivity since it only has an exhaustivity implicature, as shown in section 3.5. Most studies found that clefts are less exhaustive than exclusives but more exhaustive than narrow focus.

De Veaugh-Geiss et al. (2018b) present an experiment in which they compare all of these constructions for German in one experiment. Their study consists of a mouse-driven picture verification task. For their experiment, they describe a scenario with four male flatmates. Hence, the domain is restricted to those four individuals. Each trial of their experiment starts with a screen showing four cards, as in Figure 9.1.

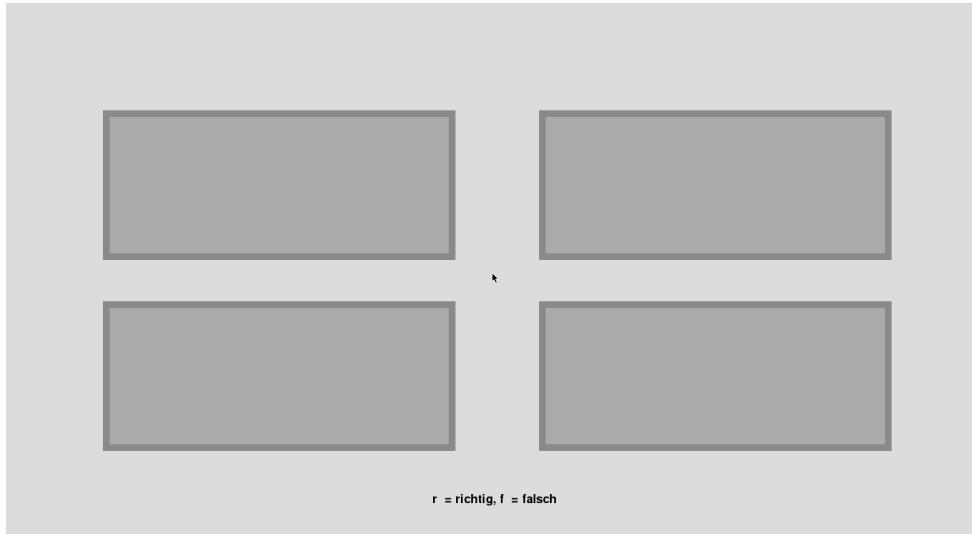


Figure 9.1: Screen on the start of each trial, with no card uncovered. Below the cards, it says ‘r = correct, f = false’.

Then, one of the test sentences, as exemplified in (15), or a filler is presented auditorily.

- (15) a. Es ist Ben, der einen Pullover angezogen hat.
 it is Ben who a sweater put.on has
 'It is Ben who put on a sweater.' es-CLEFT

b. BEN hat einen Pullover angezogen.
 BEN has a sweater put.on
 'BEN put on a sweater.' NARROW FOCUS

c. Nur BEN hat einen Pullover angezogen.
 only BEN has a sweater put.on
 'Only BEN put on a sweater.' EXCLUSIVE

d. Derjenige, der einen Pullover angezogen hat, ist Ben.
 DEF.MASC.SG who a sweater put.on has is Ben
 'The one who put on a sweater is Ben.' DEFINITE PSEUDOCLEFT

The participants' task is to uncover as many cards as necessary to judge the truth of the test sentence, and then hit *R* for *richtig* ('true') or *F* for *falsch* ('false'). Each card contains information about one of the flatmates in random order. Hence, in this set-up, participants incrementally receive more information, and decide themselves when they have enough information for making a judgment.

Two situations are of particular interest for the exhaustivity inference of clefts: (i) The prejacent of the cleft is verified on card 2, as in (16-a), or (ii) exhaustivity is violated on card 2 before the prejacent is verified, as in (16-b).

- (16) Audio: Es ist Ben, der einen Pullover angezogen hat.

'It is Ben who put on a sweater.'

- a. Card 1: Max put on a hat.
Card 2: **Ben** put on a sweater.
- b. Card 1: Max put on a hat.
Card 2: **Jens** put on a sweater.

For the situation in (16-a), De Veaugh-Geiss et al. (2018b) found significantly more early judgments (choosing ‘true’ after uncovering card 2) for the focus condition than for any other condition. Those that did not say true immediately, continued uncovering in order to check for exhaustivity. For focus, it did not seem necessary to check for exhaustivity in contrast to clefts and exclusives. This shows that focus is less exhaustive than clefts. For clefts and pseudoclefts, about half of the decisions were early ‘true’ judgments, the other half continued. There was not significant difference between clefts and pseudoclefts. For exclusives, almost all participants continued in order to check exhaustivity, as expected.

In the second situation, as in (16-b), De Veaugh-Geiss et al. (2018b) found that exclusives significantly more often than the other constructions resulted in early ‘false’ judgments (after uncovering card 2). Moreover, the focus sentence significantly more often resulted in the decision to continue than the other three constructions. Clefts and pseudoclefts were again not significantly different and showed a 50/50 ratio between ‘false’ judgments and continuations.

De Veaugh-Geiss et al. (2018b) show in a post-hoc analysis that the 50/50 ratio of judgments/continuations in those situations resulted from two different groups of participants. Group 1 always judged clefts and pseudoclefts like exclusives, hence exhaustive, while group 2 always judged clefts and pseudoclefts like focus, hence non-exhaustive.

Following Pollard and Yasavul (2014), De Veaugh-Geiss et al. (2018b) explain these results by assuming that participants follow different strategies to retrieve the antecedent of the existence presupposition of the cleft, which must be accommodated because the context did not provide it. The different strategies assume either an existential antecedent or a specific antecedent. The former would lead to a non-exhaustive interpretation, and the latter to an exhaustive interpretation (see De Veaugh-Geiss et al. (2018b) for a more detailed analysis).

This experiment brings several insights. It could not reproduce the finding that clefts are less exhaustive than exclusives and more exhaustive than narrow focus. In this experimental set-up, the participants either treated clefts and exclusives on a par or clefts and focus. This leads to the next insight. It seems that the context is quite important for the interpretation of clefts. As mentioned in section 9.1, a focus background cleft without any context is quite unnatural. It is not unlikely that speakers are never in a situation such as the one in the experiment, where they have to accommodate the referent without any cues. This could have made them come up with the strategies to group them with one of the other constructions.

Onea and Beaver (2009) introduced the *yes, but...-test*, which was used in many experiments to test whether the exhaustivity inference of clefts is at-issue or not at issue. The test shows that different constructions require different response strategies for pointing out an exhaustivity violation. It is exemplified for English data in (17).

- (17) It is Arne who is smiling.
- ?**No**, Nina is also smiling.
 - Yes, but** Nina is also smiling.
 - ?**Yes, and** Maya is also smiling.
- (18) Only Arne is smiling.
- No**, Nina is also smiling.
 - ?**Yes, but** Nina is also smiling.
 - ?**Yes, and** Maya is also smiling.

- (19) ARNE is smiling.
- a. **?No**, Nina is also smiling.
 - b. **?Yes, but** Nina is also smiling.
 - c. **Yes, and** Maya is also smiling.

Onea and Beaver (2009) showed in an empirical study that Hungarian preverbal focus, which is treated as an equivalent to the cleft, elicited mainly *yes, but*-responses, as in (17-b). Exclusives, however, led to mainly *no*-responses (18-a), and for canonical sentences with narrow focus the test subjects preferred the *yes, and*-response (19-c). Destruel (2013) and Destruel et al. (2015) show the same evidence for French. Destruel et al. (2015), in contrast, found no significant difference between the English *it*-cleft and focus, using the *yes, but...-test*.

Destruel et al. (2015), furthermore, show that the *yes, but...-test* indicates whether an inference is at-issue or not-at-issue, using examples that are uncontroversial with respect to at-issueness. They show that not-at-issue inferences have a tendency to trigger the *yes, but*-reply, just like the cleft. Therefore, they conclude that the exhaustivity inference of clefts is not-at-issue.

De Veaugh-Geiss et al. (2015) present a study that aims to determine whether the cleft exhaustivity inference is pragmatic or semantic, taking into account that it is a non-at-issue inference. As a baseline, they use exclusives with *nur* ('only'), for which the (not-)at-issue content is illustrated in (20).

- (20) Only Arne is smiling.
- a. *At-issue*: Nobody other than Arne is smiling.
 - b. *Not-at-issue*: Arne is smiling.

Targets like the following were presented to the participants in an acceptability judgment task. The sentence involved contradicting at-issue or not-at-issue content for both clefts and exclusives.³⁷

³⁷The original targets were in German.

(21) *Contradiction of not-at-issue content*

- a. It is Arne who is smiling and Nina is smiling too.
- b. Only Arne is smiling and he is not smiling.

(c.f. De Veaugh-Geiss et al., 2015)

(22) *Contradiction of at-issue content*

- a. It is Arne who is smiling and he is not smiling.
- b. Only Arne is smiling and Nina is smiling too.

(c.f. De Veaugh-Geiss et al., 2015)

De Veaugh-Geiss et al. (2015) found a significant interaction between sentence type (cleft vs. *only*) and contradiction type (at-issue vs. not-at-issue). They found that only targets of the kind in (21-a) were slightly acceptable. All the other conditions ranged equally low on the acceptability scale. This supports again that exhaustivity is weaker in clefts than in exclusives because clefts are less degraded when exhaustivity is violated, as in (21-a). De Veaugh-Geiss et al. (2015), however, argue that this is not due to at-issueness. If it was, the (not-)at-issue content would behave like the (not-)at-issue content of exclusives (their baseline) with respect to acceptability. This is not what they observe, given that acceptability for exclusives in the not-at-issue condition does not increase like for the cleft. They conclude that the difference arises because the inference is a pragmatic inference in clefts, as opposed to a semantic inference in exclusives. Byram Washburn et al.'s (2019) findings support this interpretation. They found that clefts do not always lose naturalness when exhaustivity is violated and concluded that it must be an exhaustivity implicature that can be canceled.

I argue, however, that the differences observed by De Veaugh-Geiss et al. (2015) do not arise due to at-issueness. The problem is that, for *only*, contradicting the not-at-issue content necessarily leads to also negating the at-issue content.³⁸ This explains the unacceptability of the exclusive irrespective of the contradiction type since both contradiction

³⁸Onea (2019a:13) has a similar point of critique.

types involve contradiction of the at-issue content for exclusives. For clefts, no such problem arises.

Drenhaus et al. (2011) ran one of the few online studies on cleft exhaustivity, an ERP study investigating the effect of exhaustivity violations in the brain. They compared German exclusives with *nur* ('only'), as in (23), to German *es*-clefts, as in (24), with and without exhaustivity violations.

- (23) a. *Exclusive, no exhaustivity violation:*

Nur Maria kann das Klavier spielen und außerdem noch die Geige, sagte...

'Only Mary can play the piano and, besides, the violin, said...'

- b. *Exclusive, exhaustivity violation:*

Nur Maria kann das Klavier spielen und außerdem noch Luise und Jana, sagte...

'Only Mary can play the piano and, besides, Luise and Jana, said...'

- (24) a. *Cleft, no exhaustivity violation:*

Es ist Maria, die das Klavier spielen kann und außerdem noch die Geige, sagte...

'It is Mary that plays the piano and, besides, the violin, said...'

- b. *Cleft, exhaustivity violation:*

Es ist Maria, die das Klavier spielen kann und außerdem noch Luise und Jana, sagte...

'It is Mary that plays the piano and, besides, Luise and Jana, said...'

The violation of exhaustivity in clefts, as in (24-b), led to an N400 effect, while the violation of exhaustivity in *nur*-exclusives, as in (23-b), led to a P600 effect. This shows that the two inferences are not only formally different, but they are also processed differently. Drenhaus et al. (2011:331) interpret the N400 for clefts as "an indicator of a semantic/pragmatic integration process which is typically triggered by the processing of non-stereotypical events". They assume this event to be the exhaustivity violation itself.

They point out that a semantic violation can be ruled out because semantic violations do not cause N400 effects. This speaks in favor of exhaustivity being a pragmatic inference.

The P600, observed for exclusives, is interpreted as indicating a reanalysis process, or even a semantic violation by Drenhaus et al. (2011). They conclude that *es*-clefts and exclusives in German “involve qualitatively different processing mechanisms and can hence be seen as involving different generators” (Drenhaus et al., 2011:334). They suggest, like De Veaugha-Geiss et al. (2015), that the difference lies in the semantics/pragmatics distinction. Onea (2019a), in contrast, argues that these results are equally compatible with an analysis that distinguishes clefts and exclusives via at-issueness.

9.3 Experiments on Contrastivity

Another important line of empirical research on clefts is concerned with contrast (c.f. Destruel and Velleman, 2014; Destruel et al., 2018, 2019). In section 8.4, clefts were claimed to be more felicitous when they conflict with the expectations about the world or about the development of the discourse (see also Zimmermann, 2011b). This section will present the empirical findings of the two main studies on contrast in clefts, Destruel and Velleman (2014) and Destruel et al. (2019).

Destruel and Velleman (2014) conducted a production study on English *it*-clefts, in which they compared answers to *wh*-questions and corrections of presupposed content. Example (25-a) and (25-a) present a stimulus for an answer and a correction, respectively.

- (25) a. Answer Condition

Your friend says: This bean dip is fantastic. I really want to get the recipe.

Who made it?

Answer: Tim.

You say: _____

b. Correction Condition

Your friend says: This bean dip is fantastic. I really want to get the recipe.

I can't believe that **Shannon** made it – she's normally not a very good cook.

Correction: Tim.

You say: _____

(Destruel and Velleman, 2014:204)

The participants were asked to produce the part labeled *You say* using a full sentence. Destruel and Velleman (2014) expected a tendency towards the canonical sentence for (25-a), and a tendency towards the cleft for the sentence in (25-b). The respective answers are exemplified in (26).

- (26) a. *Answer:* Tim made it.
 b. *Correction:* But it was Tim who made it.

Indeed, they found that in the correction condition, such as for (25-b), participants produced a cleft significantly more often than in the answer condition, such as (25-a). However, even in the correction condition, clefts were only produced in about half of the trials. Hence, there does not seem to be a one-to-one-mapping between marking correction and the cleft structure. However, Destruel and Velleman (2014) did show that clefts do not constitute good answers to explicit *wh*-questions since the participants mainly produced canonical sentences in the answer condition.

Furthermore, Destruel and Velleman (2014) and Destruel et al. (2019) found in several experiments using rating tasks that English *it*-clefts and French *c'est*-clefts were rated higher the more the cleft conflicted with the expectations of a given speaker. In their experiment, they manipulated the Speaker A's beliefs in the context, which have an impact on the expectations about the content, as well as the at-issueness, which influences the expectations about development of the discourse (see section 8.4 for a more detailed explanation of these different kinds of expectations). The different conditions are illustrated in (27).

- (27) Speaker A: This bean dip is fantastic. I really want to get the recipe ...
- a. *Informational*
... Who made it?
 - b. *At-issue weak belief*
... I think that maybe Shannon brought it.
 - c. *At-issue strong belief*
... I'm sure that Shannon brought it.
 - d. *Counter-presuppositional*
... I can't believe that Shannon brought it – she's normally not a very good cook.
- (Destruel and Velleman, 2014:207)

The conditions represent the contexts for the test sentences. The strength of beliefs increases from (27-a) to (27-d). Condition (27-d) is special since it also presents the conflicting information as not-at-issue, while the others are at-issue. In each of these contexts, Destruel and Velleman (2014) compared the judgments of the participants for the cleft (*It was Tim who made it*) and the canonical sentence (*Tim made it*). For condition (27-a), the test sentence simply provided an answer. For the other three conditions, it constituted a correction. Destruel and Velleman assumed that the stronger the speaker's beliefs are, the less expected it is that there will be a contradicting reply.

Destruel and Velleman (2014) found that clefts were significantly worse in the informational condition, such as (27-a), compared to the other three. Furthermore, they found that canonical sentences are significantly worse in the counter-presuppositional condition, as in (27-d), than in the other three. As already mentioned in 8.4, I argue that the test item (27-d) does not only differ with respect to at-issueness, but also with respect to the distance of the cleft and the sentence it corrects. It is the only item that contains an intervening sentence (*She's normally not a very good cook*). If such a sentence is added to (27-c), as in (28), I suppose that the sentence is equally acceptable as (27-d).

- (28) a. A: This bean dip is fantastic. I really want to get the recipe. I am sure that Shannon brought it. **She is an amazing cook.**
 b. B: It is Tim who brought it.

Hence, it is not clear whether it is at-issueness that is measured here or the distance of the antecedent of the cleft. Destruel and Velleman (2014) did not find a significant effect of the factor strength of beliefs, but the factor at-issueness plays a role for the acceptability of the canonical sentence. They conclude that canonical sentences are required to address at-issue content while clefts may be used to address not-at-issue content. Furthermore, clefts are rather used to mark unexpected discourse developments than unexpected content per se. The effects are not as strong and clear-cut since canonical sentences and clefts are always competing structures. Therefore, both are acceptable in a lot of contexts. I will address this issue in chapter 10, where I distinguish contexts that clearly favor the cleft sentence from contexts that clearly favor the canonical sentence.

In a follow-up experiment, that has a similar set-up as the previously described rating study, Destruel et al. (2019) found quite the contrary. Similar to Destruel and Velleman (2014), they manipulated, among other factors, at-issueness and contrariness. Contrariness was composed of the factor contradiction and the factor strength of beliefs. Contradiction refers to the difference between test items in which the target sentence realized a correction (contradictive), and test items in which the target sentence just provided information (non-contradictive).³⁹ Example (29)–(31) illustrate three of the conditions used in Destruel et al. (2019).⁴⁰

- (29) *Non-contradictory (no fixed belief), At-issue*

Speaker A: This bean dip is fantastic. I will need to get the recipe. **I guess someone from the marketing department brought it.**

Speaker B: Yeah, it's Lyle who brought it. / Yeah, Lyle brought it.

³⁹The latter was realized by an overt question in Destruel and Velleman (2014).

⁴⁰The example stimuli are taken from their supplementary material (Appendix A), accessible under <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01400/full#supplementary-material>.

- (30) *Strong belief, At-issue*

Speaker A: This bean dip is fantastic. I really want to get the recipe. **It turns out that Sharon brought it.**

Speaker B: Actually, it's Lyle who brought it. / Actually, Lyle brought it.

- (31) *Strong belief, Not-at-issue*

Speaker A: This bean dip is fantastic. I really want to get the recipe. **But Sharon, who brought it, already left the party.**

Speaker B: Actually, it's Lyle who brought it. / Actually, Lyle brought it.

Destruel et al. (2019)

Destruel et al. (2019) had a total of six conditions, but I will just discuss those three because they illustrate the significant difference they observed. The non-contradictive condition in (29) is the only condition in which speaker B's reply is not a correction but just provides information. In this condition, Destruel et al. (2019) found that the cleft was significantly less acceptable than in all the other conditions, including (30) and (31). In contrast to Destruel and Velleman (2014), there was no difference in the judgments of the cleft with respect to at-issueness. The judgments did not differ significantly for the cleft in (30) and (31). Destruel and Velleman (2014), in contrast, had found an effect of at-issueness. Furthermore, the degraded acceptability of the cleft in the non-contradictory condition showed that the cleft prefers a salient contradictory alternative, the existential in (29) is not sufficient. Destruel et al. (2019) conclude from their results that “metalinguistic expectations about how a contrary point of view is changing in the discourse are less relevant to the acceptability of clefts than are salient beliefs about the world” (Destruel et al., 2019:10).

9.4 Summary

The empirical investigations had two main topics, exhaustivity and contrast, which were tested with both offline and online methods. It has been shown that the exhaustivity

inference of clefts is weaker than exhaustivity in exclusives and stronger than exhaustivity in focus. However, it is still not settled whether these differences arise due to differences in the (not-)at-issueness status of the inference or due to differences with respect to the inference being semantic or pragmatic. There is a lot of evidence, though, that the exhaustivity inference of clefts is a pragmatic inference. This is in conflict with many of the theoretical analyses, presented in chapter 7, which assume cleft exhaustivity to be semantically encoded. So far, there are very few approaches to bridging this gap between the empirical evidence and the theoretical analysis (they are discussed in chapter 7).

Furthermore, the relevance of contrast for clefts has been attested. The issue whether clefts involve a contrast to information provided in the context (Destruel et al., 2019) or a contrast to expectations about the development of the conversation (Destruel and Velleman, 2014) remains unclear.

Another result that is not as clear but still worth mentioning is the observation made by Byram Washburn et al. (2019), which is supported by corpus studies, that there are many clefts for which exhaustivity vanishes. Corpus studies even show examples of clefts that contain additive focus particles, which explicitly contradict exhaustivity. This does not seem to be problematic for the acceptability of the cleft. This observation is in line with exhaustivity being a pragmatic inference in clefts, but also with clefts having no exhaustivity inference at all. The observed effects might all result from the requirement of contrast instead of exhaustivity, a claim that needs to be tested in the future.

Finally, I want to point out some aspects of clefts that were ignored in the empirical studies, but that are of great (maybe crucial) relevance. First of all, the context seems to play an important role. The only studies that took this into account were the corpus studies presented in section 9.1. Those studies generated hypotheses that also involved the cleft as part of a discourse. However, these hypotheses have not really been tested empirically. The experiments on clefts, in contrast, only embedded the cleft stimuli in rather short contexts, if any at all. I argue that the discourse effect is lost in most of the experiments, which would be fatal if the purpose of a cleft was actually one that was crucially related to discourse. I will propose something like that in chapter 10.

Another aspect which was not covered in most of the empirical studies was the competition between the cleft and canonical. Destruel et al. (2019) pointed out that the cleft is hardly ever the only acceptable option in a context. Frequently, the cleft and the unclefted canonical sentence may be used interchangeably with no (big) difference in text coherence or meaning. This makes it quite difficult to approach the difference between the two constructions in naturally occurring examples. I want to emphasize the importance of finding and testing contexts in which the cleft is significantly better than the canonical sentence and vice versa. Such data is discussed in section 10.1.

Chapter 10

Expectedness-based Analysis of German Clefts

This chapter is the centerpiece of this thesis. It presents the final analysis of German *es*-clefts in discourse, in contrast to unclefted German sentences. In section 10.1, I present and discuss additional examples, that other approaches struggle to account for. Furthermore, I develop a hypothesis about the function of *es*-clefts in German. In section 10.2, I first informally introduce the concept of expectedness with respect to questions arising in discourse. Then, I provide a formal definition of expectedness in form of an expectedness function. Based on this function, section 10.3 is concerned with how to incorporate the function into the discourse model. Accordingly, the discourse model, presented in chapter 4, is adjusted. In section 10.4, I illustrate the application of the new discourse model, and I show that it makes the correct predictions for the cleft and the canonical sentence. Finally, section 10.5 discusses the new approach and gives an outlook.

10.1 More Data and Hypothesis

In my approach, I focus on how an *es*-cleft in German contributes to the discourse structure and, in particular, how it does this differently than its canonical counterpart. Exam-

ples (1) and (2) represent the two extreme cases which clearly distinguish the cleft from the canonical sentence.

- (1) Speaker A: Who did Lena talk to?

Speaker B:

- a. ?It was Peter she talked to.
- b. She talked to PETER.

- (2) Lena hat gestern auf der Party mit einem Typen₁ gesprochen. Die beiden haben viel gelacht und sich direkt für den nächsten Abend verabredet. Dann ist Lena glücklich nach Hause gefahren.

'Yesterday at the party, Lena talked to some guy₁. The two of them laughed a lot and they agreed to meet again the next evening. Then, Lena went home happily.'

- a. Es war Peter₁, mit dem sie gesprochen hat.
it was Peter₁ with whom she talked has
'It was Peter₁ she talked to.'
- b. ?Sie hat mit PETER₁ gesprochen.
She has with Peter₁ talked
'She talked to Peter₁'

In (1), the canonical sentence (1-b) is clearly more acceptable than the cleft (1-a). In (2), in contrast, the cleft (2-a) is clearly more acceptable than the canonical sentence (2-b). These examples show that the context plays a crucial role for the acceptability of the cleft. I will argue that the difference between the cleft and the canonical sentence in the two examples is rooted in discourse coherence. The unacceptable sentences in (1) and (2) each interrupt the flow of reading. I will explicate this intuition below.

Before, I will summarize again why other approaches struggle to provide a convincing explanation for the difference between the cleft and the canonical sentence in discourse. First of all, the unacceptability of the cleft in (1) does not arise from an exhaustivity violation (see chapter 7 for a summary of exhaustivity in clefts). Quite the contrary, the context of a question supports an exhaustive interpretation, even for the canonical sentence in (1-b). However, there are many other more plausible explanations for the

difference in (1). Related to exhaustivity, one could argue that the cleft in (1-a) is redundant in its function of marking exhaustivity, since a context including an explicit question is frequently interpreted to require an exhaustive answer. Hence, a canonical plain focus sentence, as in (1-b), suffices and is a less complex structure, which makes it a better option than the cleft. With respect to complexity, one could argue that the cleft is dispreferred because it competes with the less complex canonical sentence and the even more less complex term answer *Peter*. As long as it does not serve an additional function besides answering the question, the other two options are preferred again for reasons of reduced complexity. However, this raises the question if the cleft has an additional function as opposed to canonical sentences at all. Example (2) suggests that it does, given that the cleft is preferred this time. Thus, the explanations based on complexity need to account for why the more complex structure is not a problem in (2) anymore, and is even preferred.

Velleman et al. (2012) argue that the cleft functions as an inquiry-terminating construction. However, the marking of the termination of an inquiry is only necessary when the inquiry is extended to a certain degree. They argue that the inquiry in (1) is too short and the marking of its end would be redundant. As mentioned in section 8.6, this argument can easily be reversed claiming that the inquiry in (1) is too short and should, thus, be marked as terminating unusually fast. Then, also based on inquiry-termination, the cleft would not be predicted to be less acceptable.

De Veaugh-Geiss et al. (2015) and Tönnis et al. (2018), among others, argue that the cleft is a focus-marking device in order to disambiguate focus. This provides an explanation for example (1). Given that focus is not ambiguous since the current question (CQ) is expressed as an explicit question, there is no need for focus disambiguation. Hence, the cleft would be redundant. In example (2), however, the canonical sentence is ambiguously focus marked. It could be narrow focus on *Peter* or a wider focus, e.g. on *talked to Peter*. Hence, the cleft serves its purpose of disambiguating focus in this example. I argue, however, that this approach, too, needs to explain why focus can be disambiguated easily in some contexts, as in (1) and many other examples, but not in other contexts in

which the cleft is more acceptable than the canonical sentence. Furthermore, even in a situation in which focus is neither ambiguous for the canonical nor for the cleft, we can still sometimes observe a preference of the cleft. This is demonstrated in example (3), repeated from section 4.7. Note that this example only shows the desired effect in spoken German, when focus is marked by prosody.

- (3) Lena wurde gestern auf der Party von einem Typen₁ angesprochen. Die beiden haben viel gelacht und sich direkt für den nächsten Abend verabredet. Dann ist Lena glücklich nach Hause gefahren.

‘Yesterday at the party, Lena was approached by some guy₁. The two of them laughed a lot and they agreed to meet again the next evening. Then, Lena went home happily.’

- a. Es war PETER, der Lena angesprochen hat.
it was PETER who Lena approached has
‘It was PETER who approached Lena.’
- b. ?PETER hat Lena angesprochen.
PETER has Lena approached
‘PETER approached Lena.’

A main accent on the subject unambiguously marks narrow focus on the subject in this example. Hence, the cleft would be redundant in its function of disambiguating focus, and should be dispreferred, contrary to what we observe for this example.

I conclude that the cleft does have a function different from the canonical which just does not apply in all contexts, e.g. it does not apply in the context of an explicit question, as in (1). I will propose that a cleft can only occur in specific discourse contexts, which I will specify by analyzing the kinds of questions that are present in the discourse, and by investigating the relation of clefts and canonical sentence to these questions.

I assume that in the course of a story or a conversation, there will be questions that arise to the addressee, and there will be expectations on the side of the addressee whether a question is likely to be addressed or not. Furthermore, I assume that the speaker or writer wants to address questions her-/himself and decides based on the addressee’s expectations

how to address them without losing text coherence. I will show in this chapter that clefts address different questions than canonical sentences. In a nutshell, I propose the following for the discourse behavior of clefts in contrast to canonical sentences.

- (4) a. An *es*-cleft addresses a question that came up in the preceding context, but that the addressee does not expect to be answered at that point in the discourse compared to other questions.
- b. Those questions that are more expected are preferably addressed with a canonical sentence instead of a cleft.
- c. Those questions that are neither particularly expected nor particularly unexpected can equally well be addressed by a cleft or a canonical sentence.

The idea crucially relies on the concept of expectedness of questions to be addressed in discourse, which will be discussed in section 10.2. I will now present the rationale behind this idea using some more examples, some of them mentioned above and some new ones. In section 10.3, the idea will be implemented in an adapted discourse model in order to account for the discourse behavior of clefts in German.

First of all, I claim that expectedness can explain the differences between the cleft and the canonical sentence in (1) and (2), above. Based on the assumptions made in section 4.4, I assume that the cleft in example (2-a) as well as the canonical sentence in (2-b) address the question *Which guy did Lena talk to?*. This question is not a question that would naturally come to the addressee's mind at that point in the discourse. In other words, the addressee would not very strongly expect that question. The hypotheses in (4) predict the cleft to rather address unexpected questions and, therefore, the acceptability of the cleft in (2) is predicted. The canonical sentence, in contrast, is predicted to be unacceptable because the question it addresses is too unexpected. It cannot address the question *Which guy did Lena talk to?* without creating an incoherent discourse. It is assumed to only address questions that are expected to a certain degree, and this question does not reach this degree. What this degree might be will be discussed in subsection 10.2.2.

In example (1), the cleft and the canonical sentence again address the same question, namely *Who did Lena talk to?*, which is explicitly posed in the context. This time, however, it is very expected that Speaker B would address this question. Hence, the hypotheses correctly predict the canonical sentence to be acceptable, and the cleft to be unacceptable. Intuitively, the difference between the cleft and the canonical sentence is modeled via expectedness of the discourse move they are realizing. This also explains the feeling of interruption of the reading flow. Unexpected moves make the reader stumble, unless she/he is somehow warned, e.g. by a marked structure such as the cleft.

A cleft can, however, not address a question that is extremely unexpected. This is illustrated by example (5).

- (5) **Lena ist gestern Abend bei der Party angekommen und hat erstmal einen leckeren Cocktail getrunken.** Danach hat sie mit ihrer Freundin Andrea getanzt und die beiden hatten sehr viel Spaß. Dann ist Lena glücklich nach Hause gefahren.

'Lena arrived at the party yesterday and first of all she had a tasty cocktail. Thereafter, she danced with her friend Andrea and the two of them had a lot of fun. Then, Lena went home happily.'

- a. ?Es war ein Bloody Mary, den sie getrunken hat.
it was a Bloody Mary that she drunk has
'It was a Bloody Mary she drank.'
- b. ?Sie hat einen Bloody Mary getrunken.
She has a Bloody Mary drunk
'She drank a Bloody Mary.'

Here, the sentence in bold raises the question *q*: *Which cocktail did Lena drink?*. However, the addressee would not strongly expect the speaker/writer to address this question. Moreover, the subsequent context is not concerned with the cocktail any longer, which makes it even less expected that *q* is addressed. As example (5) shows, both the *es*-cleft and the canonical sentence are degraded in such a context. For the canonical sentence, the reason is obvious. The canonical sentence is not used to address unexpected questions. In

contrast to the examples above, the cleft is not an acceptable alternative that can be used to address the unexpected question in this case. I argue that this is the case because even a cleft needs at least a certain level of expectedness of the addressed question in order to be acceptable. It seems that the cleft cannot address a question that is too unexpected. Therefore, the claim (4-a) must include a greatest lower bound of required expectedness. Note that this issue is related to the very general assumptions about acceptable discourse moves, as described in terms of relevance in section 4.4. Addressing a very unexpected question would result in an incoherent discourse and, thus, such a discourse move would not be accepted by the interlocutors.

Another example, that has been discussed a lot, e.g. by Pollard and Yasavul (2014), is a cleft that is preceded by an indefinite antecedent, as in example (6).

- (6) Lena hat gestern auf der Party mit einem Typen gesprochen.

'Yesterday at the party Lena talked to some guy.'

- a. Es war Peter.
it was Peter
'It was Peter.'
- b. ?Es war Peter, mit dem sie gesprochen hat.
it was Peter with who she talked has
'It was Peter she talked to.'
- c. ?Sie hat mit PETER gesprochen.
she has with PETER talked
'She talked to PETER.'

Pollard and Yasavul (2014) claim that the pivot of a focus background *it*-cleft (here *Peter*) anaphorically refers back to a discourse referent in the context. This referent can be an indefinite description, as in (6) (*some guy*). The acceptability judgments for the example in (6), however, are not obviously predicted by the hypotheses stated above.

In (6), the truncated cleft in (6-a) is the best continuation, while the full *es*-cleft in (6-b) and the canonical sentence in (6-c) seem to be equally degraded. This example causes two problems for my approach. First, it is not obvious what my approach predicts for truncated clefts compared to full clefts or canonical sentences. Second, my approach would

not predict the cleft and the canonical sentence to be equally degraded unless they are both completely unexpected discourse moves, which is not the case here. The identification of the guy Lena talked to is relevant information in the context of the indefinite description *einem Typen* ('*some guy*') and, hence, constitutes an acceptable discourse move. This is also supported by the fact that the truncated cleft is acceptable. The elliptical structure describes the same informational content as the full cleft. Therefore, the cleft cannot, contentwise, be categorized as irrelevant information in this context.

I argue that the reason for the higher acceptability of the truncated cleft, in contrast to the other two sentences, is independent of my analysis. The truncated cleft seems to be the best candidate simply because it is less complex. Both the full cleft and the canonical sentence repeat a lot of material from the immediately preceding sentence, which is stylistically degraded and a redundant effort. Hence, the full cleft and the canonical sentence are degraded because the truncated cleft is the better alternative. Whatever my approach predicts for the truncated cleft, this is an independent explanation for why it is more acceptable than the full cleft and the canonical sentence in this context.

For the discussion of the second problem (why the cleft and the canonical sentence are equally degraded), it is necessary to know how expected the addressed question (*Which guy did Lena talk to?*) is in this context. As presented above, I hypothesize that the canonical sentence is more acceptable than the cleft when a rather expected question is addressed. The cleft, however, is more acceptable than the canonical sentence in case of a rather unexpected question. In example (6), it is not intuitively obvious how expected the addressed question is. It seems neither very expected nor very unexpected. However, if one assumed that the question was a rather expected question, my approach would incorrectly predict the canonical sentence in (6-c) to be more acceptable than the *es*-cleft in (6-b), which is not what we observe here. If one assumed the question to be less expected, the opposite prediction would be made.

The only way how my approach can predict equal acceptability of a cleft and a canonical sentence is when they both contain irrelevant information with respect to the context, or when the expectedness of the addressed question lies in the overlapping area of the

expectedness required for clefts and canonical sentences, i.e. those levels of expectedness in which both the cleft and the canonical sentence are acceptable. The first case is already ruled out above. In the second case, my approach would predict the cleft and the canonical sentence to both be rather acceptable than unacceptable since the overlapping area is one where both the cleft and the canonical sentence are possible to use. One way to explain why the cleft and the canonical sentence are equally unacceptable would be that due to the much better candidate in form of the truncated cleft the differences between the two comparably much worse candidates is eliminated (similar to a floor effect in rating studies). Hence, they are both perceived as just unacceptable.

In order to rule out such an effect, consider example (7), which eliminates the problem that (6) suffered from, namely repeating too much material and resulting in a too complex structure.

- (7) Gestern hatte Lena ein Meeting.

'Yesterday, Lena had a meeting.'

- a. #Es war Peter.
it was Peter
'It was Peter.'
- b. Es war Peter, mit dem sie sich getroffen hat.
it was Peter with who she herself met has
'It was Peter she met.'
- c. Sie hat sich mit PETER getroffen.
she has herself with PETER met
'She met PETER.'

In this example, the truncated cleft is unacceptable, for the simple reason that material is elided that cannot be recovered. The difference in acceptability of the full cleft in (7-b) and of the canonical sentence in (7-c) is not easily identifiable. Both sentences are acceptable, but it seems that, without any additional information, the canonical sentence is slightly better in this example.⁴¹ According to the hypothesis (4-c), this could be a case of addressing a question that is neither particularly expected nor particularly unexpected.

⁴¹So far, this judgment has only been confirmed by a few speakers.

Indeed, the question addressed by the cleft/canonical sentence (*Who did Lena meet?*) is expected but not extremely expected, maybe with a tendency towards more expected. Hence, the hypothesis would correctly predict either both sentences to be acceptable or the canonical sentence to be more acceptable.

Another observation that needs to be explained is illustrated in example (8).

- (8) Lena hat gestern auf der Party mit einem Typen₁ gesprochen. Die beiden haben viel gelacht und sich direkt für den nächsten Abend verabredet. **Dann hat Lena ihm sogar ein Geheimnis verraten.**

‘Yesterday at the party, Lena talked to some guy₁. The two of them laughed a lot and they agreed to meet again the next evening. Then, Lena even told him a secret.’

- a. ?Es war Peter₁, mit dem sie gesprochen hat.
it was Peter₁ with whom she talked has
‘It was Peter₁ she talked to.’
- b. ?Sie hat mit Peter₁ gesprochen.
She has with Peter₁ talked
‘She talked to Peter₁.’

The acceptability of the cleft in (8) decreases when a pressing question which is not the question that the cleft addresses is evoked right before the cleft. In this example, the sentence in bold evokes the question q_1 : *What was the secret?*, while the cleft and the canonical sentence address the question q_2 : *Which guy did Lena talk to?*. Earlier versions of this example had a less “interesting” sentence (*Then, Lena went home happily*), instead of the bold sentence. That sentence did not raise equally strong new questions. In (8), however, the evoked question q_1 seems very strong. Even though the addressee very urgently wants to know the answer to the question q_1 , it might not be expected to be addressed immediately. It is possible that the speaker is building up suspense (the difference between urgency and expectedness will be explained in subsection 10.2.1), which makes the addressee expect to get the information later in the discourse. Nevertheless, this question seems to affect the expectedness of other questions, such as q_2 , making them

less expected. It seems that the addressee is considerably less interested in information about the guy after having heard about a secret. A first brief explanation for this could be the following: Even though the addressee does not expect to be told what exactly the secret was, she/he still expects to be led towards it. Therefore, the addressee does not expect anymore that the speaker/writer would pick up an unrelated issue that was raised before, such as the one raised by q_2 . In this case, the question q_2 is probably too unexpected compared to the secret issue. Hence, the hypotheses would predict the cleft to be unacceptable, similarly to example (5).

If a sentence interferes between such a pressing question and the cleft/canonical sentence, such as the bold sentence in (9), the situation is similar to (2). The cleft is again better than the canonical sentence.

- (9) Lena hat gestern auf der Party mit einem Typen₁ gesprochen. Die beiden haben viel gelacht und sich direkt für den nächsten Abend verabredet. Lena hat ihm sogar ein Geheimnis verraten. **Dann ist Lena glücklich nach Hause gefahren.**
‘Yesterday at the party, Lena talked to some guy₁. The two of them laughed a lot and they agreed to meet again the next evening. Lena even told him a secret. Then, Lena went home happily.’
- a. Es war Peter₁, mit dem sie gesprochen hat.
 it was Peter₁ with whom she talked has
‘It was Peter₁ she talked to.’
 - b. ?Sie hat mit Peter₁ gesprochen.
 She has with Peter₁ talked
‘She talked to Peter₁.’

This example is more difficult to explain since it seems that the sentence in bold goes against the expectations of the addressee, and the speaker in a way refuses to tell the secret by this move. Nevertheless, the bold sentence does not seem to realize an unacceptable discourse move. This example will be discussed in more detail in section 10.5. More generally, this demonstrates again how questions that are raised by a new sentence affect the expectedness of previously raised questions that have not been addressed yet. It seems

that recently raised issues are more expected to be addressed than less recent ones. In section 10.3, I will define a formalism that automatically accounts for this effect. I will present a discourse model that can account for all these observations. I argue that existing approaches, which focus on the exhaustivity or contrastivity of clefts, cannot account for the effect of discourse (in-)coherence of clefts or canonical sentences, as was shown by the data in this section already.

10.2 Expectedness

This section represents the core of this thesis. It introduces the concept of EXPECTEDNESS in subsection 10.2.1, and it provides a formal definition of expectedness in subsection 10.2.2, which is the foundation of the adaption of the discourse model presented in section 10.3. I will show how expectedness crucially contributes to predicting the acceptability of German *es*-clefts in different contexts.

10.2.1 Concept

Expectedness has been mentioned in this thesis before. In section 8.4, two variants of expectedness were distinguished based on Zimmermann's (2011b) contrastive focus hypothesis. Based on this, Destruel and Velleman (2014) distinguish *expectations about the world* and *expectations about the discourse*, as described in (10).

- (10) a. *Expectations about the world*: These expectations may involve beliefs about the world, expressed as assertions or presuppositions.
 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.

(c.f. Destruel and Velleman, 2014:199)

I will make the same distinction, calling expectations about the world CONTENT EXPECTEDNESS and expectations about the discourse DISCOURSE EXPECTEDNESS. The former describes which propositions would be considered likely to hold in the context. The latter is concerned with which speech acts are likely to be performed by the interlocutors. In particular, I interpret discourse expectedness as describing which questions are likely to be addressed next in a speech act, given a context. A question could be addressed by asking it, if it has not been asked yet, or answering it. In order to understand the difference between the two notions of expectedness, consider the following example.

- (11) You know what I had for dinner yesterday? I had three fries, five peas and a spoonful of jam.

The second sentence in (11) is highly discourse-expected for the addressee after having heard the question *You know what I had for dinner yesterday?*. This first sentence triggers the expectation to soon receive the answer to that question, no matter what the answer is, and the second sentence provides this answer. Hence, discourse expectedness in this case only describes whether addressing the question *What did you have for dinner?* is expected for the addressee. However, the content of the answer provided by the second sentence is definitely not expected, given that it is certainly surprising to eat such a combination of food items. Accordingly, the second sentence is not at all content-expected.

In example (12), the opposite is the case: The sentence in bold is content-expected but not discourse-expected.

- (12) Lena arrived at the party yesterday and first of all she had a tasty cocktail. Thereafter, she danced with her friend Andrea and the two of them had a lot of fun. Then, Lena went home happily. **?She enjoyed the cocktail.**

Knowing that Lena had a tasty cocktail, it is expected that she enjoyed that cocktail. Hence, the sentence in bold is content-expected. However, it is not expected for the discourse to develop in the direction it does in (12). In other words, the addressee does not expect the speaker to address the question *What did Lena do with the cocktail?* or

Did Lena enjoy the cocktail? at this point in the discourse. Therefore, the bold sentence is not discourse-expected.

Example (11) and (12) illustrate an important difference between the two variants of expectedness: If a sentence is too discourse-unexpected but content-expected, as in (12), it leads to infelicity or incoherence. If it is content-unexpected but discourse-expected, as in (11), the addressee just receives surprising information, but the sentence is still felicitous. The reason for the infelicity in (12) is that it is difficult for the addressee to establish a discourse relation between the previous text and the bold sentence. Accordingly, she/he perceives the text as incoherent. In my analysis of German *es*-clefts, I will be mainly concerned with discourse expectedness, showing that clefts can rescue otherwise incoherent sentences (see more in section 10.4).

In order to understand how clefts do that, we have to look at discourse expectedness from the speaker's perspective. Discourse expectedness has an effect on the speaker or writer in the process of planning to address certain questions. In order to maintain text coherence, the speaker/writer decides how to address the intended question based on what she/he anticipates about the expectations of the addressee. The consequence of these considerations about the addressee is that the speaker/writer would pick different constructions or phrases depending on whether the addressee expects the next question to be addressed or not. Example (13) illustrates how the writer could address a very unexpected question.

- (13) Lena arrived at the party yesterday and first of all she had a tasty cocktail. Thereafter, she danced with her friend Andrea and the two of them had a lot of fun. Then, Lena went home happily. **By the way, she did not only drink a cocktail, but also a lot of shots.**

The last sentence in (13) can be assumed to address the question *What else did Lena drink at the party*, or something similar, which is not expected at that point of the discourse. The discourse move is realized by repeating a lot of material already explicitly mentioned before, and by using the phrase *by the way*. Both are means that the writer uses in order

to mark that she/he addresses an unexpected issue. As it was phrased above, the writer is “warning the reader not to stumble”. Without the marking the sentence would be less acceptable, as (14) indicates.

- (14) Lena arrived at the party yesterday and first of all she had a tasty cocktail. Thereafter, she danced with her friend Andrea and the two of them had a lot of fun. Then, Lena went home happily. **?She (also) had a lot of shots.**

In this example, the continuation without extra marking seems a bit incoherent. However, when the same issue is addressed earlier in the discourse, less marking is needed, as (15) indicates.

- (15) Lena arrived at the party yesterday and first of all she had a tasty cocktail.
- ?By the way, she did not only drink a cocktail, but also a lot of shots.
 - She also had a lot of shots.

The sentence in (15-a) is not completely unacceptable, but at least parts of it seem redundant. Given that it is now expected for the reader to hear more about Lena’s drinks, there is no need for the writer to mark that she/he addresses an unexpected issue. Hence, the repetition of the lexical material *drink a cocktail* and the phrase *by the way* seem unnecessary.

The examples (13) to (15) show that the speaker/writer needs to take the expectations of the addressee into consideration in order to chose an appropriate phrasing. An unexpected issue justifies, and maybe even requires, more marking than an expected issue. An expected issue sometimes even improves with less marking.

This observation is not new. Krifka and Musan (2012), for instance, state that unexpected moves should be marked explicitly. As an example, they present the discourse relation of contrast in comparison to all the non-contrastive discourse relations, suggesting that the discourse relation of contrast is less expected than other discourse relations. According to them, contrast needs to be marked, e.g. by *but* or *nevertheless*, while non-contrastive

relations do not have to be expressed explicitly. Onea and Zimmermann (2019) point out the need for additional effort in discourse management, e.g. by using a different prosody or a particle, when the speaker decides to depart from an alternative space that has just been built. An example for such a situation is provided in (16), repeated from chapter 2 (example (6)).

(16) Speaker A: Was hast du gerade in meinem Zimmer gesehen?

'What have you just seen in my room?'

Speaker B:

- a. ?Was versteckst du in deinem Zimmer?
what hide you in your room
'What are you hiding in your room?'
- b. Was versteckst du denn in deinem Zimmer?
what hide you Q.PRT in your room
'Why, what are you hiding in your room?'
lit. *'What are you Q.PRT hiding in your room?'*

In this example, Speaker B decides to not address the question raised by Speaker A, but poses a different question that is not a sub-question of A's question, which is a rather unexpected discourse move. By ignoring this question, Speaker B refrains from choosing one of the alternatives proposed by Speaker A, i.e. she/he departs from the alternative space. Speaker B implies with this move that Speaker A was asking because she/he was scared that Speaker B would see something in the room that she/he was not supposed to see. The acceptability of this move increases when the question receives extra marking by the question particle *denn*. Hence, (16-b) is more acceptable than (16-a). Csipak and Zobel (2014) claim that *denn* in *wh*-questions expresses that the speaker has a heightened interest in the answer to the question. Hence, it marks the relevance of the question for Speaker B in our case. This might also make Speaker A reconsider the relevance of Speaker B's question. I am not committing to this analysis of *denn*, though (see Theiler (2020) for a different analysis of *denn*). However, in that sense, *denn* could serve here as the required extra marking for a discourse-unexpected question. This additional marking helps the addressee to establish coherence even for unexpected discourse moves.

In my analysis, I analyze *es*-clefts in German as yet another linguistic means to mark a discourse-unexpected move.

Hedberg (1990), in contrast, relates the cleft to content expectedness. She states that it “is particularly common in mystery novels for clefts to be used to highlight the role of an unexpected participant in an expected event” (Hedberg, 1990:120). In her example, repeated in (17), the event of somebody answering the door is expected. It is, however, unexpected that it is Webber who answers the door instead of the apartment owner Ruskin.

(17) Beginning at the top of the list, I went along the landing and tapped at Ruskin's door. When it was opened, it was Webber who stood there. We stared at each other for a moment, both of us taken aback.

[Lucille Kallen, *The Piano Bird*, p. 95]

(Hedberg, 1990:120)

I argue that the cleft does not only introduce an unexpected participant, it also addresses an unexpected question at that point of the discourse. The question *Who stood there?* in the context of somebody knocking at Ruskin's door and the door being opened is not particularly expected to be addressed because the reader would probably already assume that Ruskin would stand there. Instead, other questions would be more expected, such as *What happened then?*. I will mainly be concerned with the discourse expectedness in this thesis. For reasons of brevity, I will, henceforth, use the term *expectedness* to refer to discourse expectedness if not explicitly stated otherwise. Focusing on discourse expectedness does not imply that content expectedness does not play a role for the analysis of clefts. Ideally, I aim for tracing this effect back to discourse expectedness, as I attempted for example (17).

So, what is it that makes a discourse move expected or unexpected? As it was already mentioned, this approach of expectedness is concerned with how expected is that a certain question is going to be addressed. Hence, an (un)expected discourse move addresses an (un)expected question. As I mentioned in chapter 4, I assume that in the course of a story or a conversation there will be questions that arise to the addressee, and there will

be expectations on the side of the addressee whether a question is likely to be addressed in the next discourse move. I assume that those questions are not all equally expected to be addressed from the perspective of the addressee. Consider example (18).

- (18) Lena told Andrea a secret.

The author of this example would anticipate the addressee to wonder what the secret was, but also why Lena told the secret to Andrea. However, intuitively, the first question is more expected than the second. Whether a question arises and how expected it is, depends on many factors such as the addressee, the situation, characteristic of the utterance, and the common ground. The latter is illustrated by (19-a) and (19-b).

- (19) a. Yesterday, I came to the meeting. My boss called me an idiot.
b. Yesterday, I forgot about the meeting. My boss called me an idiot.

In (19-a), the potential question *Why did she call you an idiot?* arises, and the addressee expects to some degree that this question will be addressed. In (19-b), in contrast, the answer to that question is already in the common ground and the question does not arise. Thus, the addressee does not expect that question to be addressed. Another factor that determines expectedness for a question is when in the discourse this question came up. Recall example (2), in which the question *Which guy did Lena talk to?* was not expected at the point when the cleft/canonical sentence was uttered. After the first sentence in the context of (2), however, it is actually rather expected to learn more about that guy. The discourse model, developed in section 10.3, will account for this effect.

The concept of expectedness will be incorporated into the discourse model. More specifically, it will play a crucial role in the process of accepting a discourse move, which can only be accepted if it addresses a sufficiently expected question. This is reminiscent of Roberts' (2012) concept of *Relevance*. According to her, a discourse move can only be accepted if it is relevant. However, she does not distinguish different degrees of relevance, which I will do for expectedness. And I assume that this is crucial for explaining the dif-

ference between clefts and canonical sentence. In section 10.3, we will see that relevance is covered by expectedness, but expectedness makes additional useful predictions, too.

In the following, I will discuss the difference between expectedness and urgency, and I will explain why it is expectedness and not urgency that makes the correct predictions for the acceptability of clefts. In many situations, if an addressee does not feel the urge to know the answer to a question, she/he also does not expect to be given that answer. This is the case in most of the examples discussed so far. In (19-a), for instance, the question *Why did she call you an idiot?* is both very expected to be addressed, and it is particularly urgent for the addressee to get an answer.

There are some cases, though, in which urgency and expectedness differ for the question that is addressed, such as example (20).

(20) Gestern hatte Lena ein heimliches Rendezvous.

‘*Yesterday Lena went on a secret rendezvous.*’

- a. Es war Peter, mit dem sie sich getroffen hat.
it was Peter with who she herself met has
‘*It was Peter she met.*’
- b. Sie hat sich mit PETER getroffen.
she has herself with PETER met
‘*She met PETER.*’

The observation, which is so far only confirmed by a couple of speakers, is that the cleft in (20-a) and the canonical sentence in (20-b) are either equally acceptable, or the cleft is even slightly better than the canonical sentence. The first sentence in this example induces the question *Who did Lena meet?* or *Whom did Lena have a rendezvous with?*, which can be considered very urgent questions in this context. The reason is the word *heimlich* (‘secretly’), which triggers the urge in the addressee to know the answer to that question. If the purpose of using a cleft was addressing a question that is not urgent, the cleft would be predicted to be dispreferred in this context. Accordingly, it seems that urgency does not play such a crucial role in predicting the acceptability of *es*-clefts.

Based on expectedness, however, we can make better predictions. The first sentence of example (20) does not only raise the urgency of the question *Who did Lena meet?*, but it also builds up suspense through the word *heimlich* ('secretly'). This can be argued to not only affect the urge to know more details, but it also affects the expectations of the addressee about the continuation of the storyline. When the author creates suspense by omitting details (e.g. by not mentioning the person Lena secretly went on a rendezvous with), the addressee would not expect to get those details easily and immediately, especially not right in the next sentence. We have learned from novels and stories that suspense is usually built up slowly and is not resolved too easily and soon. Otherwise, why would one even bother to build up suspense if it was resolved in the next sentence anyway. Based on these considerations, the addressed question in (20) might be very urgent but not very expected to be addressed just yet. If clefts address questions that are less expected and canonical sentences address more expected questions, the cleft in (20-a) is correctly predicted to be more acceptable than (20-b). Hence, the crucial factor should be expectedness instead of urgency, and the hypothesis is that a German *es*-cleft addresses a less expected (but not necessarily less urgent) question than the canonical sentence does. Only a tendency towards a difference between urgency and expectedness with respect to cleft acceptability can be observed. It must be tested empirically in future research. In general, however, the values of expectedness and urgency coincide in many cases.

10.2.2 Formalization

Having introduced the concept of EXPECTEDNESS above, I now formally define expectedness and will include it as a primitive category in the discourse model (see section 10.3). In my model, each question that can possibly be asked or has already been asked has an expectedness value between 0 and 1, depending on what has previously been uttered. Expectedness values can also be seen as probabilities for the addressee to expect a certain question to be addressed in a context. To give an example, a question that has already been answered in the previous discourse has an expectedness value of 0 (or close to 0) since

the addressee does not expect this question to be addressed again. A potential question (PQ) triggered by the last utterance, in contrast, might have a high expectedness value after it has been evoked by an utterance.

In order to formalize expectedness, I define an EXPECTEDNESS FUNCTION f_e , which takes a commitment space C (as introduced in chapter 4)⁴² from the set \mathcal{CS} of possible commitment spaces and a question q from the set of possible questions \mathcal{Q} as its input and yields an EXPECTEDNESS VALUE. The set \mathcal{CS} contains all commitment spaces that are consistent. The set \mathcal{Q} contains all questions that are well-formed semantic objects of interrogative sentence radicals. I assume questions to be represented as q-alternatives (as described in section 2.3)⁴³. However, the function does not hinge on this representation of questions. An expectedness value, henceforth EV, is a number between 0 and 1 that represents the level of expectedness of a question in a given context.

The expectedness function assigns an EV to every pair of a commitment space and a question. It determines how expected each question is, given the previous context, which is represented as the commitment space. In particular, it captures the idea that the expectedness of questions is likely to change with each update. For this purpose, f_e is defined as a recursive function in Definition 30, which I will explain step by step below.

⁴²A commitment space is a set of sets of propositions, which represents the common ground and the possible continuations of the discourse.

⁴³The q-alternatives are described as the set of the possible answers to q .

Definition 30 (Expectedness Function). *The expectedness function $f_e : \mathcal{CS} \times \mathcal{Q} \rightarrow \mathcal{E}$ is a recursive function, where*

- \mathcal{CS} is the set of all possible commitment spaces,
- \mathcal{Q} is the set of all possible questions,
- \mathcal{E} is the set of expectedness values from 0 to 1,

such that for any given $C \in \mathcal{CS}$

$$\sum_{x \in \mathcal{Q}} f_e(C)(x) = 1.$$

f_e is defined recursively as follows:

- i. For any $q \in \mathcal{Q}$, and for $C_0 \in \mathcal{CS}$, such that C_0 is the commitment space at the beginning of a conversation, $f_e(C_0)(q)$ assigns a prior EV to q .
- ii. For any $C \in \mathcal{CS}$, any $q \in \mathcal{Q}$, that is open in C , and any **assertive speech act** A_φ addressing the CQ cq_φ ,

$$f_e(C + A_\varphi)(q) \propto \begin{cases} f_e(C)(q) + \alpha & \text{iff } q \text{ is a PQ licensed by } \varphi. \\ \max((f_e(C)(q) - \epsilon), 0) & \text{iff } cq_\varphi \text{ q-entails } q. \epsilon \text{ varies} \\ & \text{with respect to completeness of } \varphi \text{ as answer to } q. \\ f_e(C)(q) & \text{otherwise.} \end{cases} \quad \begin{matrix} [a\text{-PQ}] \\ [a\text{-CQ}] \\ [OTHER] \end{matrix}$$

- iii. For any $C \in \mathcal{CS}$, any $q \in \mathcal{Q}$, that is open in C , and any **interrogative speech act** $A_{?\varphi}$ addressing the CQ $cq_{?\varphi}$,⁴⁴

$$f_e(C + A_{?\varphi})(q) \propto \begin{cases} f_e(C)(q) + \beta & \text{iff } q \text{ is a PQ licensed by } A_{?\varphi}. \\ f_e(C)(q) + \gamma & \text{iff } q \text{ is a subquestion of } cq_{?\varphi}. \\ f_e(C)(q) + \alpha + \beta + \gamma + \delta & \text{iff } q = cq_{?\varphi} = ?\varphi. \\ f_e(C)(q) & \text{otherwise.} \end{cases} \quad \begin{matrix} [i\text{-PQ}] \\ [i\text{-SUB.Q}] \\ [i\text{-CQ}] \\ [OTHER] \end{matrix}$$

(For questions that are not open, hence have been answered already, the EV stays 0.)

Normalization Consider first the condition that the EVs have to add up to 1 (for the range of f_e), repeated in (21).

⁴⁴As mentioned in section 4.4, addressing the CQ in an interrogative speech act means asking the CQ. Hence, the CQ $cq_{?\varphi}$ is identical to the uttered question $?_\varphi$ in this case.

(21) *Normalization Condition of EVs:*

For a given $C \in \mathcal{CS}$, $\sum_{x \in Q} f_e(C)(x) = 1$.

This requirement expresses the similarity of EVs and probabilities. It implies that once an EV is raised, others must decrease. As we will see below, this will make interesting predictions about the development of discourse. The requirement in (21) is called *Normalization Condition* because it has the effect that the values of all questions are normalized after each change due to an update. I will get back to this below in this section.

Recursion The function f_e is defined recursively, which means that the output of the function f_e again uses the same function f_e . The desired consequence is that each new update will incrementally effect the expectedness of each question because the output of f_e uses the previous commitment space, thereby excluding the most recent update.⁴⁵ The recursion starts with the initial step *i.*, repeated in (22).

(22) *Step i. of f_e :*

For any $q \in Q$, and for $C_0 \in \mathcal{CS}$, such that C_0 is the commitment space at the beginning of a conversation, $f_e(C_0)(q)$ assigns a prior expectedness value to q .

Step *i.* of the recursion provides the prior EVs at the beginning of a conversation, given that we never start a conversation without any shared information. Based on this common ground, some questions are a priori more expected than others. For example, it is more expected in an out-of-the-blue context to address the question *What happened today?* compared to the question *What would happen if aliens took over the world?*. Questions that are commonly assumed to be answered already, such as *Is the world round or flat?*, are assumed to have an EV of 0. I will not be concerned with those priors in more detail, and will just assume them to be well-defined.

Step *ii.* and step *iii.* are the recursive steps for an assertive and an interrogative speech act, respectively. I will first discuss the properties that both steps share, and will then

⁴⁵How an interrogative or assertive update change the commitment space is defined in section 4.2.

proceed to their specific requirements. Step *ii.* and *iii.* define the effect of an update on the expectedness of a question in relation to the previous commitment space. Put differently, with an update of a commitment space the EV of each question changes as well. In particular, these steps describe how the EV of a question and a commitment space is proportional to (marked by \propto) the EV of the same question in the commitment space without the update.

Different cases are distinguished, depending on features of the respective question. They present a first approach to provide an idea of what can affect expectedness. It is not meant as an exhaustive list of cases, which means that the case OTHER contains some cases that do affect expectedness but are ignored at this stage. Section 10.5 will discuss some further aspects that might affect expectedness of a question.

Formally, the EV after an update is either set to 0, it stays the same as in the previous commitment space, or a variable is added to or subtracted from the EV of the previous commitment space.⁴⁶ Strictly speaking, those variables should not be added/subtracted. The variables α , β , γ , δ , and ϵ should rather be increasing or decreasing functions, that take $f_e(C)(q)$ as their argument. Below, I will explain this and will discuss some characteristic of those variables in more detail. For reasons of simplicity, I will treat the variables as context-dependent values (hence no constants), that can be added to or subtracted from an EV. Below, we will see a case that illustrates their context-dependence.

Some questions in \mathcal{Q} might not be immediately affected by an update, but since other questions will be, their EV will nevertheless change because of all the values adding up to 1. This leads to a normalization of all the EVs. For instance, if an update very strongly raises the EV of one question, this will lower the EVs for all the other questions. This also has the implication that when a question decreases to an EV close to 0, the other questions will automatically gain in expectedness. This describes, for instance, the situation where one issue is settled entirely, and only then older issues will become more expected again. This does something very similar to Roberts' (2012) question hierarchy. When answering

⁴⁶Definition 30 uses ‘+’ for two different operations: a context update, as in $C+A_\varphi$, and for adding a variable to an EV.

the sub-issues on the lowest level, we only go back to the next sub-question on the next higher level, once the other issues are settled.

I will now explain each of the cases of step *ii.* of Definition 30 for an assertive speech act, repeated in (23), in more detail. The cases are labeled in square brackets, where *a* stands for assertive speech act.

(23) *Step ii. of f_e for an assertive speech act:*

For any $C \in \mathcal{CS}$, any $q \in \mathcal{Q}$, that is open in C , and any **assertive speech act** A_φ addressing the CQ cq_φ ,

$$f_e(C+A_\varphi)(q) \propto \begin{cases} f_e(C)(q) + \alpha & \text{iff } q \text{ is a PQ licensed by } \varphi. \\ \max((f_e(C)(q) - \epsilon), 0) & \text{iff } cq_\varphi \text{ q-entails } q. \epsilon \text{ varies} \\ & \text{with respect to completeness of } \varphi \text{ as answer to } q. \\ f_e(C)(q) & \text{otherwise.} \end{cases} \quad \begin{matrix} [\text{a-PQ}] \\ [\text{a-CQ}] \\ [\text{OTHER}] \end{matrix}$$

This step distinguishes three cases, in which the EV of a question either increases or decreases after an update resulting from an assertive speech act.

Case a-PQ (Potential question (PQ) of an assertion) If a commitment space C is updated with an assertive speech act, asserting the proposition φ , and q is a PQ of φ , then the expectedness is higher after the update (in the commitment space $C + A_\varphi$) than before in C . An example for this situation is (19-a), repeated in (24-a).

- (24) a. Yesterday, I came to the meeting. (C) My boss called me an idiot. ($C + A_\varphi$)
b. Yesterday, I forgot about the meeting. (C) My boss called me an idiot. ($C + A_\varphi$)

The commitment spaces are indicated in brackets after the last update they include. The second sentence in (24-a) evokes the potential question q : *Why did your boss call you an idiot?*, as discussed, e.g. in section 4.6. In C , which is just updated with the first sentence in (24-a), q is not very expected to be addressed, and can be assumed to also not have a particularly high prior EV. In $C + A_\varphi$, however the situation has changed a lot. The update with the speech act of asserting the second sentence promoted q , which is now

much more expected to be addressed. This is modeled by f_e adding α to the EV that q has in C for the updated $C+A_\varphi$.

In example (24-b), the same PQ, but in a different context, does not gain in expectedness, though. The reason is that this time the PQ is already answered by the first sentence. This explains why only questions are considered in step *ii.* that are open in C . If a question has received the EV 0, it should not be addressed or asked again and stay 0. This explains the extra requirement in the definition of f_e , repeated in (25).

- (25) For questions that are not open, hence have been answered already, the EV stays 0.

An exception would be the speech act of correction, where one of the discourse participants explicitly readdresses a question that has already been answered. For such a speech act, one would have to assume that questions are not immediately set to 0 when they are answered, but instead are put on hold for some time. Farkas and Bruce's (2010) *Table*, already mentioned in section 4.2, models such an approach, where questions and assertions are left on the *Table* for negotiation before it is finally settled. The definition of f_e excludes such a situation. However, it is possible to adjust it in order to make it compatible with the concept of a *Table*.

Case a-CQ (Current question (CQ) of an assertion) This case is concerned with answering a question. It is the only case in which the expectedness of the respective question is reduced after the update. Hence, the variable ϵ is subtracted. Additionally, we need to make sure that no EV is negative after subtraction, which is solved by using the function *max*. This function yields 0 in case $(f_e(C)(q) - \epsilon)$ would be negative. Otherwise, it yields $f_e(C)(q) - \epsilon$. I will first discuss the special case, in which the CQ of φ does not only q-entail the question q but is identical to q (q-entailment is defined in section 2.6, Definition 4). If a commitment space C is updated with the assertive speech act A_φ and q is the CQ of φ , the expectedness of q is lower in $C+A_\varphi$ than in C , as illustrated in the following example.

- (26) a. (C) Lena talked to PETER. ($C+A_\varphi$)
 b. CQ: Who did Lena talk to?

The assertion in (26-a), addresses the CQ in (26-b), which must be expected to a crucial degree in C because, otherwise, it would not have been addressed by the assertive speech act A_φ . In the updated $C+A_\varphi$, however, it is not as expected to be addressed anymore, and, therefore f_e subtracts the variable ϵ .

This example also shows how ϵ could vary depending on whether we take φ to be an exhaustive answer, a partial answer or a mention-some answer. If we had expected that Lena would talk to many people, the expectedness would be reduced less than if we had expected her to talk to just one person. This relates to what van Kuppevelt (1995) calls *satisfactoriness*. This concept describes to what extent an issue is settled. A partial answer, for instance, is not as satisfactory as a complete answer. Therefore, the more satisfactorily a question is addressed, the more its EV is reduced. If A_φ provides a mention-some answer, the EV of the CQ decreases more than for other partial answers since the addressee does not expect more information than the one provided.

If the question q is q-entailed by the CQ of φ but not identical, we are dealing with a sub-question of the CQ. As for the CQ itself, also a sub-question can be answered partially or completely, which leads to the subtraction of a smaller or larger ϵ -value, respectively. In contrast to the CQ itself, it could happen that a sub-question q of the CQ stays the same ($\epsilon = 0$) after the update. This is only possible when A_φ is a partial answer to CQ, because then q could represent that part of CQ that is not addressed by this partial answer. Imagine a context in which Lena talked to Peter, Nina, and Laura. In such a context, the assertive speech act of (27-a) partially answers the CQ *Who did Lena talk to?*. However, sub-question (27-b) loses expectedness, while sub-question (27-c) stays the same, because it was not covered by the partial answer.

- (27) a. Lena talked to PETER. (CQ: Who did Lena talk to?)
 b. Did Lena talk to Peter?
 c. Did Lena talk to Nina?

One could argue that in some situations a partial answer could even make the sub-questions that are not addressed more expected than before. This would be in line with van Kuppevelt (1995), who argues that discourse participants aim for settling an issue until it satisfies their need of detail. Hence, if Lena usually talks to many people, a partial answer would not be satisfying and the EVs of the unsettled sub-questions of the partially answered CQ should increase. I argue that we still do not need to actually add ϵ , instead of subtracting it, for unsettled sub-questions, but it suffices to leave their value as it is. Those sub-question will automatically increase their EVs because the EVs of other sub-questions and of the CQ itself decrease. The process of normalization will result in those unanswered sub-questions having relatively higher EVs again.

Case OTHER All the other cases of questions are subsumed in the category OTHER, which does not affect expectedness. As mentioned above, this category most likely contains sub-categories that do affect expectedness, such as dependent questions (mentioned in section 4.4). So far, those sub-categories are not accounted for by f_e . However, nothing speaks against extending the list of cases for f_e . For now, I focus on these few cases, and leave the other cases to future research.

We will now move on to the cases of step *iii.* of Definition 30 for an interrogative speech act, repeated in (28). The cases are again labeled in square brackets, where i stands for interrogative speech act.

(28) *Step iii. of f_e for an interrogative speech act:*

For any $C \in \mathcal{CS}$, any $q \in \mathcal{Q}$, that is open in C , and any **interrogative speech act** $A_{?\varphi}$ addressing the CQ $cq_{?\varphi}$,

$$f_e(C+A_{?\varphi})(q) \propto \begin{cases} f_e(C)(q) + \beta & \text{iff } q \text{ is a PQ licensed by } A_{?\varphi}. \\ f_e(C)(q) + \gamma & \text{iff } q \text{ is a subquestion of } cq_{?\varphi}. \\ f_e(C)(q) + \alpha + \beta + \gamma + \delta & \text{iff } q = cq_{?\varphi} = ?\varphi. \\ f_e(C)(q) & \text{otherwise.} \end{cases}$$

[i-PQ] [i-SUB.Q] [i-CQ] [OTHER]

It seems more obvious than for the assertions that raising a question changes the expectations of the addressee. When a question is explicitly asked, it requires quite some effort to not address it. It is possible, though, to not address it, which I showed in section 4.4, example (26). However, the expectations are high that a question will be addressed when explicitly posed. In the following, I explain the four cases of f_e for the interrogative speech act. I will not explain the OTHER-case because it is the same as for assertions.

Case i-PQ (Potential question (PQ) of an interrogative speech act) As shown in section 4.6, questions may, too, license PQs. The example is repeated in (29).

- (29) a. (C)When did you get home yesterday? ($C+A_{?\varphi}$)
- b. PQ: Why are you asking?

Example (29-a) evokes the PQ in (29-b), which is much more expected in $C+A_{?\varphi}$ than in C . It even refers to the speech act of asking the question in (29-a), which would not be possible before the speech act has taken place. Hence, the variable β is added to the EV of this PQ in C to model the increase of expectedness of a PQ after the licensing interrogative utterance.

Case i-SUB.Q (Sub-question of an interrogative speech act) In a commitment space that has been updated by an interrogative speech act $A_{?\varphi}$, it is expected to address sub-questions. This describes the strategy of asking a sub-question when one cannot answer $?\varphi$, as in example (30).

- (30) a. Peter celebrated his birthday.(C) Who had a present for him? ($C+A_{?\varphi}$)
- b. Sub-question: Did Nina have a present for him?

If we did not know anything specific about Nina's and Peter's relationship beforehand, we would more strongly expect the speaker to address whether Nina had a present for Peter in $C+A_{?\varphi}$, after having heard the question in (30-a) about presents than just after having heard that Peter celebrated his birthday. Hence, the presence of the super-questions raises

the expectedness of any sub-question to be addressed. Hence, the variable γ is added to the value of the sub-question in (30-b) in $C+A_{?\varphi}$, by Definition 30.

Case i-cQ (Current question (CQ) of an interrogative speech act) If q is the identical to the CQ of the question $?\varphi$, it means it is also identical to $?\varphi$ itself. Definitely, the expectedness of a question increases when it is explicitly asked. As already pointed out, it needs a lot of extra effort to not address a question since it is highly expected in discourse to address overt questions. By adding the sum of all the variables that could possibly be added to a question plus an additional constant δ , it is guaranteed that an explicit question will always have the highest increase of expectedness compared to all the other questions. After normalization, this means that an explicit question q will reduce the expectedness of all questions in $C+A_{?\varphi}$ compared to C , except for q itself of course.

Characteristics of the variables The actual prior EVs, as well as the actual values for α , β , γ , δ , and ϵ , must be determined empirically. So far, there is not sufficient evidence to hypothesize about how much an EV increases or decreases, only *whether* it increases or decreases. Hence, it cannot yet be determined whether this in-/decrease is indeed realized by addition or subtraction of a variable, as it was stated in Definition 30. Instead, it is very well possible that the EV of a question in the previous context is multiplied either by a value greater than 1 or a value between 0 and 1, in order to model the increase or the decrease, respectively. Given that we know so little about the kind of in- or decrease, it seems reasonable to make the variables themselves functions, that take $f_e(C)(q)$ as their arguments. This would allow for a lot more options to determine how exactly the EVs increase or decrease.

One way to get more insights about the characteristics of these variables is by testing them in a model. The variables (or variable functions) can be set as model parameters in order to estimate them from a data set. Another preliminary approach to empirically test the expectedness of questions is presented in Westera and Rohde (2019). They presented snippets of texts to participants and asked them which questions are evoked. The EV

of a question could be calculated from the relative frequency of that question. With this method, one would have to take into account that the values will not add up to 1 because people on average do not treat probabilities as adding up to 1 when asked in an experimental setting.

There are, however, a few hypotheses that I want to make about how the variables relate to each other. I chose different variables for each case, because I assume that the strength of the change of expectedness differs between the different cases in f_e . I suggest an explicit question q , as in case i -CQ, to raise the strongest expectations for exactly that question to be addressed. This is already captured in the definition of f_e , by adding the sum of all variables to the EV of q in the previous commitment space. The variable α for the assertive speech act subsumes the rather huge category of potential questions, which is most likely oversimplified. For instance, the EVs of the primary potential questions (PPQs) of an assertion will probably increase more strongly than some of the other PQs. This is exemplified in the following example, repeated from section 4.6.

- (31) a. A: Peter was not alone.
PQ: Who was he with?
b. A: Peter was with someone.
PPQ: Who was he with? (Onea, 2016:15)

Both the PQ and the PPQ have a higher EV after the assertions than before the assertion. Intuitively, it seems that the PPQ in (31-b), however, is even more expected than the PQ in (31-a). Therefore, the variable α should be more fine-grained with respect to different kinds of PQs. Furthermore, there are questions that are technically PQs according to Onea (2016), which I will not consider in a -PQ, such as the following.

- (32) Lena talked to Peter.
PQ: How many words per second did she utter?

This PQ is not at all expected after the licensing utterance. I only consider PQs that can be assumed to actually arise to the addressee. Which questions these are, needs to be

determined empirically (see again Westera and Rohde, 2019). Comparing PQs licensed by assertions to PQs licensed by questions, I suggest that the EV of the former increases more than the EV of the latter. I hypothesize that it is more expected to address a PQ that is evoked by an assertion. This describes a very natural way of the development of a discourse or a coherent text, as discussed extensively in Roberts' work. For questions, asking a PQ is a rather unexpected move, because it questions the speech act and not the content. The EV for a PQ of a question still increases but not as strongly. Implementing this for f_e , α would have to be required to be greater than β by a certain factor.

Something similar, I suggest for the variables β and γ , PQs and sub-questions of questions respectively. I propose that it is usually more expected to follow a strategy addressing a sub-question (and this means answering it or asking it) in the light of an explicit question than asking a PQ of that question. Following Roberts (2012), most conversations follow strategies of addressing sub-questions, and, accordingly, a sub-question should be rather expected. Therefore, γ should be greater than β by a certain factor. There might be some exceptions though, such as (33).

- (33) a. Do you have any plans for today?
- b. PQ: Why are you asking?
- c. Sub-question: Do you have any plans in the afternoon?

In this example, it is rather expected that the PQ is addressed, since a question like (33) is often asked not only to obtain information about the plans of the addressee but also to announce the coming proposal of a great plan that involves the addressee. Hence, the PQ is aiming for information about this great plan, and is, thus, rather expected in the context in (33). The sub-question in (33-c) also receives a higher EV after the super-question is asked. It seems to be natural for the addressee to reply with *In the afternoon I want to go swimming* (addressing the sub-question). However, I want to point out that this question does not seem particularly more expected than the PQ in (33).

I do not want to make a claim about the difference between *a*-PQ and *i*-SUB.Q. It is unclear which question's EV increases more, that of a PQ licensed by an assertion or that

of a sub-question of an explicit question. I do not have evidence or specific intuitions about this difference, but I will assume them to not be identical. Hence, α and γ are not identical. Finally, the variable ϵ , which was already discussed above, is already claimed to vary with respect to the level of completeness of the answer provided by the assertive speech act in Definition 30.

10.3 Adapted Discourse Model

This section is based on the discourse model developed in chapter 4 and the expectedness function defined in the previous section. I argue that including expectedness into the discourse model solves many of the puzzles, mentioned in section 10.1 and elsewhere throughout this thesis, that other approaches did not cover.

In section 4.4, I pointed out that the QUD stack by Roberts (2012) loses importance, and could in principle be given up. In the discourse model in 4, however, this would put a lot of weight on the discourse question (DQ), which I pointed out was hard to define on an abstract level, or even to determine in a given discourse. Furthermore, it was problematic that constructions like the German *es*-cleft can address questions that are neither related to a salient DQ nor present on the QUD stack (or they are very low in the stack). In order to explain the behavior of clefts, and most likely also other non-canonical sentences, we need to keep track of such questions in the discourse model.

I argue that many effects modeled in Roberts' (2012) QUD stack can equally well be explained by expectedness. Above, it was already discussed that expectedness reproduces the effects of question hierarchies. Moreover, it can also be used to define the acceptability of clefts, as we will see below. In the first step, I define an adapted version of Roberts' (2012) QUD stack, presented in Definition 31.

Definition 31 (QUD stack adapted). *For a $C \in \mathcal{CS}$, the QUD stack is defined as the set $\mathcal{S}_{qud} = \{\langle q, e_i \rangle \mid q \in \mathcal{Q} \wedge e_i > n\}$, such that*

- i. $e_i = f_e(C, q)$
- ii. *n is an expectedness threshold for questions to enter the QUD stack, which depends on the context and on the attention span of the discourse participants.*

Roughly, my version of the QUD stack describes the expectedness of all questions at a certain point of the discourse, and it changes with each update. Formally, it is the set \mathcal{S}_{qud} of pairs of questions and their respective expectedness value (EV). The EV is determined by the expectedness function. Technically, \mathcal{S}_{qud} is not a stack in the proper meaning of the word. However, the questions can be ordered with respect to the EV determined by f_e . In contrast to Roberts (2012), however, it is possible that two questions are on the same level, namely when they receive the same EV.

The set \mathcal{S}_{qud} depends on the current commitment space, since each update changes the EVs of questions. It contains a pair for each question that is above a certain minimum level n of expectedness. This threshold must be determined empirically and it serves the purpose of not having to drag along too many questions, which would exceed the working memory of discourse participants, or which are extremely unlikely to be discussed. As mentioned above, two four-year-olds will probably not discuss the question of how to solve a certain differential equation. Hence, for modeling their discourse, that question would fall below the threshold n .

This definition of the QUD stack differs from earlier versions with respect to including all kinds of questions into the stack, e.g. also potential and dependent questions etc., and in principle all of these questions are accessible at any time. One consequence is that accepting a discourse move never implies adding a new question to the stack, not even in the case in which the addressee would need to accommodate the CQ, as explained in section 4.5. I will get back to accommodation of the CQ below.

Based on this definition of the QUD stack, I will now present the final revisions of the definitions of *Relevance* and *Accepting a Discourse Move*. For the latter, I will distinguish

ACCEPTING AN UNMARKED ASSERTION from ACCEPTING A CLEFT ASSERTION. An unmarked assertion is a canonical sentence with no additional marking, such as discourse markers. Hence, a cleft is a marked assertion since it has a complex syntactic structure. Nevertheless, the definition of unmarked assertions is relevant for my analysis of clefts because the discourse effect of clefts will be described with respect to how it differs from the discourse effect of unmarked assertions. Accepting an unmarked assertion is defined as follows.

Definition 32 (Accepting an Unmarked Assertion). *For a given current question cq_φ , and a given commitment space $C \in \mathcal{CS}$, an assertive speech act of the unmarked sentence φ is acceptable iff*

- i. φ is focus congruent to cq_φ , and
- ii. $f_e(C, cq_\varphi) > e_{can}$, where e_{can} is the EV that is necessary for an unmarked sentence to be acceptable.

C is then updated with the speech act A_φ .

As proposed in section 4.4, I assume that the CQ is determined by the speaker. Thus, it is given in the current commitment space C , which represents the context. Below, in Definition 35, I explain how the addressee can accommodate the CQ if it was not given to her/him. Definition 32 requires an acceptable assertion to be focus congruent to the given CQ, which is the same requirement that was covered in the previous principle of relevance of an assertion (Definition 22). Furthermore, an unmarked sentence can only be accepted if the CQ has a sufficiently high EV in C , which is the commitment space before the update with φ . This EV must exceed a contextually determined value e_{can} (*can* for canonical), which again must be determined empirically.

Certainly, e_{can} will be greater than the threshold n , defined for \mathcal{S}_{qud} , because sentences that are not even contained in the set of QUDs will definitely not be acceptable. Furthermore, I expect that those EVs of questions which f_e determines to increase due to an update will increase above e_{can} in many cases. Vice versa, the EVs of questions that

decrease (condition *a*-CQ) can safely be assumed to fall below e_{can} . Finally, if a sentence is acceptable, the commitment state is updated with the speech act of asserting the unmarked sentence.

I will not be concerned with updates with questions in this chapter. For completeness, I will nevertheless provide the Definition of ACCEPTING AN UNMARKED QUESTION. It is almost the same as *Accepting an Unmarked Assertion*, as shown in Definition 33.

Definition 33 (Accepting an Unmarked Question). *For a given current question cq_φ , and a given commitment space $C \in \mathcal{CS}$, an interrogative speech act of the unmarked question $?_\varphi$ is acceptable iff*

- i. $?_\varphi$ is identical to $cq_{?_\varphi}$, and
- ii. $f_e(C, cq_{?_\varphi}) > e_q$, where e_q is the EV that is necessary for an unmarked question to be acceptable.

C is then updated with the speech act $A_{?_\varphi}$.

The two differences are the condition *i.* and the threshold e_q instead of e_{can} . It is plausible that e_q and e_{can} are actually identical. However, this needs to be determined empirically, and I am not committing to how the two values are related. Generally, the same assumptions made for e_{can} should also hold for e_q . The condition *i.* in Definition 33 states identity instead of focus congruence, which was the same in the principle of relevance for questions (Definition 23, repeated in the footnote 47 on the following page).

Recalling the previous revised definitions of relevance and accepting assertions and questions⁴⁷, we see that now relevance is included in accepting assertions and questions, as well as in the definition of the expectedness function. As mentioned above, the requirement of focus congruence or identity between of the CQ and an explicit question is covered in condition *i*. Furthermore, relevance required a sub-question relation or identity of the CQ and a DQ. This DQ is now replaced by expectedness. The process of accepting an assertion or a question, as defined now, can be seen as identifying the CQ with a question–EV pair in \mathcal{S}_{qud} for which EV is sufficiently high, above e_{can} or e_q respectively. Whatever the DQ was, it would have an EV above e_{can} and e_q , since the DQ should always be expected to be addressed. If CQ is identical to DQ, it has, thus, an EV above e_{can} and e_q as well. Given that the expectedness function raises the EVs for sub-questions in case of explicit questions, it seems plausible that it assigns an EV above e_{can} and e_q to the CQ if it is a sub-question of DQ, which has a high EV. Hence, both relevance requirements are equally well predicted by Definition 32 and 33. The former definition of accepting an assertion only adds to this that the common ground is updated and that CQ is removed from the stack. The former is phrased similarly in Definition 32, by the update of the commitment space. The latter is modeled slightly differently. A question is never really taken out of \mathcal{S}_{qud} , but the EV of the CQ of an accepted assertion is reduced, as defined in Definition 30:*a*-CQ. This still has the same effect aimed for by taking it off the stack, namely that it will not be relevant anymore.

⁴⁷Those definitions were given in section 4.4, repeated in the following.

Definition 22 (Relevance of an Assertion – revised). *For a given current question Q_c and a given discourse question Q_d , an assertion φ is relevant to Q_d , iff*

- a) φ is focus congruent to Q_c , and*
- b) Q_c is either identical to Q_d or a sub-question of Q_d .*

Definition 23 (Relevance of a Question – revised). *For a given current question Q_c and a given discourse question Q_d , a question $? \varphi$ is relevant to Q_d , iff*

- a) $? \varphi$ is identical to Q_c , and*
- b) $? \varphi$ is a sub-question of Q_d .*

Definition 24 (Accepting an Assertion – revised). *For a given current question Q_c and a given discourse question Q_d , an assertion is accepted iff it is relevant to Q_d . Then, the common ground is updated and Q_c is removed from the QUD stack.*

Definition 25 (Accepting a Question – revised). *For a given current question Q_c and a given discourse question Q_d , a question $? \varphi$ is accepted iff it is relevant to Q_d . Then, $? \varphi$ becomes a new discourse question.*

Hence, this model can achieve what the previous model could. However, it can also account for cases that were not captured before. One such case is modeling the acceptability of a cleft sentence. Turning back to the difference between clefts and canonical sentences, we see now in Definition 34 that clefts have different conditions for being accepted as a valid discourse move than unmarked sentences do.

Definition 34 (Accepting a Cleft Assertion). *For a given current question cq_φ , and a given commitment space $C \in \mathcal{CS}$, an assertive speech act of the cleft sentence φ is acceptable iff*

- i. φ is focus congruent to cq_φ , and
- ii. $e_{can} > f_e(C, cq_{cl}) > e_{cl}$, where e_{cl} is the EV that is necessary for a cleft sentence to be acceptable.

C is then updated with the speech act A_φ .

Definition 34 states that the EV of the CQ of an acceptable cleft has to exceed the threshold e_{cl} . This means that even for the cleft assertion, there are questions that are too unexpected to be addressed by it. As for e_{can} , the value of e_{cl} must be greater than n . Furthermore, the EV of an acceptable cleft has to fall below the threshold e_{can} for the EV of an acceptable unmarked assertion. In other words, clefts address less expected questions than unmarked sentences do in most of the cases. In order to account for those cases in which both a cleft and a canonical counterpart are acceptable, one would have to introduce a variable m , that is added to e_{can} in Definition 34. This would make sure that there is an interval of EVs $(e_{can}, e_{can} + m)$ where both the cleft and the unmarked sentence are acceptable. For presentational purposes in section 10.4, I will use the simpler definition in this thesis without an overlapping area.

In section 4.5, it was discussed that the addressee must sometimes accommodate the CQ, since the speaker is the one who determines it. This CQ is then added to the QUD stack. In this new model, those questions that can be accommodated would need to have a high EV, and would, thus, be included in the stack anyway. Hence, the conditions for

accommodation of a CQ can also be captured by expectedness: A question that is too unexpected cannot be accommodated. If there is more than one option for the CQ of an assertion, for instance due to ambiguous focus marking, the addressee will choose the most expected question that is focus congruent to this assertion.

Taking a closer look at this accommodation process, it seems to be very similar to the process of accepting an assertion. I assume that what the addressee is doing when accommodating the CQ is stepping into the speaker's shoes at the point where she/he planned her/his discourse move. The addressee, then, considers different candidates for the CQ, given the speaker's choice of utterance, and evaluates the acceptability of those candidates in the commitment space before the speaker's utterance. Based on these considerations, I define accommodation of a CQ as follows.

Definition 35 (Accommodation of a CQ of an Assertion). *For a given assertive speech act of the proposition φ , and a given commitment space $C \in \mathcal{CS}$, q is accommodated as the current question of φ iff*

- i. φ is focus congruent to q ,
- ii. q is acceptable in C , and
- iii. for all $q' \in \mathcal{S}_{qud}$ such that $q' \neq q$, $f_e(C, q) > f_e(C, q')$

C is then updated with the speech act A_q .

This process of accommodation can be understood as considering potential updates with the speech act of asking the CQ. For the addressee, this serves the purpose of finding out which would make the best CQ given the utterance. The most expected acceptable question will be accommodated and C will be updated with this question. The decision about accepting the actual assertion φ is then made on the basis of $C + A_{cq}$, instead of C . This is a rather trivial decision now since the context has just been updated with the CQ of φ . By Definition 30:*i*-CQ, this CQ receives a particularly high EV in $C + A_{cq}$, which will most likely exceed e_{can} . However, once the assertion is accepted, this EV of the CQ will immediately decrease again because the assertion provides an answer to the CQ.

10.4 Applying the Model

I will now apply the proposed model to the examples mentioned before, which showed a difference in acceptability of the canonical sentence and the cleft. First of all, I will provide a different explanation for the example of Velleman et al. (2012), repeated in (34) with commitment spaces added in brackets.

- (34) A: (C)What did Mary eat? (C+? φ)
a. B: ?It was a PIZZA that Mary ate.
b. B: Mary ate a PIZZA.

As mentioned above, the canonical sentence is much more acceptable than the cleft as an answer to an explicit question. This is predicted by the proposed discourse model using expectedness. Given an explicit question, it is very expected that this question is going to be addressed. According to Definition 30: i -CQ, the EV of the CQ of the cleft in example (34) will be raised more than any other question after the update with the question, in the commitment space $C+?\varphi$. This value will most likely exceed the threshold e_{can} for canonical sentences and, therefore, the cleft is not acceptable in (34), Definition 34 requires the CQ to be below e_{can} for clefts. A canonical sentence, however, would be acceptable, by Definition 32. My approach can also explain why the acceptability of the cleft improves in (51), repeated in (35).

- (35) A: (C_0) What did Mary eat? (C_1)
B: I thought she said she was gonna get a pasta dish, but I might be wrong. (C_2)
A: And did she also order a salad? (C_3)
C: Guys, I was there and actually paid attention. It was a PIZZA that Mary ate.

Velleman et al. (2012) argued that the cleft in its function of marking the end of an inquiry is not necessary in (34) because the inquiry is not extended enough. The inquiry in (35) is longer and its termination should, according to them, be marked as such. I argued in

section 8.6 that this is not a very strong argument. One could equally well argue that a very abrupt ending of the inquiry after one answer, as in (34), should be marked as such.

Furthermore, the difference between (34) and (35) is not only the length of the inquiry but also the kind of questions that are evoked. In (35) in the commitment space C_1 , the EV of the question *What did Mary eat?* is very high, by 30:*i*-CQ. Then, B provides a partial answer, though a very weak one, to A's question. By Definition 30:*a*-CQ, this reduces the EV of A's question at C_2 , though probably not very much because of the partial answer. Hence, this probably does not change the overall distribution of EVs much. This is different for A's second statement of an explicit question *Did Mary also order a salad?*. By 30:*i*-CQ, the EV of this question is raised significantly in C_3 . The process of normalization has now the effect that the other questions including the previously high question *What did Mary eat?* decrease. But exactly this question is the CQ of the cleft sentence used in C's statement. It is plausible that its EV was pushed below or at least close to e_{can} . If it is pushed below e_{can} , the cleft is predicted to be acceptable in examples (35) (by Definition 34), and an unmarked sentence would be predicted to be unacceptable.

Actually, example (35) is one of the cases where a canonical sentence would be acceptable, too. This would speak in favor of EV of the cleft's CQ to be very close to but greater than e_{can} . In this approach, it is not the length of the inquiry that is crucial but the intervening question that raises expectations to address that question first. Intuitively, this seems to be the correct description. Indirectly, the length of the inquiry does still play a role since, of course, with a longer inquiry there are more chances for intervening questions that are expected to be addressed first.

Example (22), repeated in (36), can now be explained by analyzing the anticipated questions and their EVs. The commitment spaces $C_0 - C_3$ are indicated in brackets in the repeated example.

(36) (C₀) Lena hat gestern auf der Party mit einem Typen₁ gesprochen. (C₁) Die beiden haben viel gelacht und sich direkt für den nächsten Abend verabredet.

(C₂) Dann ist Lena glücklich nach Hause gefahren.(C₃)

‘(C₀) *Yesterday at the party, Lena talked to some guy₁.* (C₁) *The two of them laughed a lot and they agreed to meet again the next evening.* (C₂) *Then, Lena went home happily.*(C₃)’

- a. Es war Peter₁, mit dem sie gesprochen hat.
it was Peter₁ with whom she talked has
‘*It was Peter₁ she talked to.*’
- b. ?Sie hat mit Peter₁ gesprochen.
She has with Peter₁ talked
‘*She talked to Peter₁.*’

Recall that the other approaches could not explain why the cleft in (36-a) is more acceptable than the canonical sentence in (36-b). In contrast to Velleman et al. (2012), we are not dealing with an (extended) inquiry here that is terminated with the cleft. Other approaches that modeled discourse coherence by distinguishing CQ and DQ also struggled to make the correct predictions for this example. Both sentences have the same CQ *Which guy did Lena talk to?*, but it is unclear which DQ they address. If we assume the DQ to be identical to CQ, those theories still need to explain why a cleft can address it and a canonical sentence cannot.

I argue that my discourse model can solve these problems by taking into account the effect of all the updates between the first sentence and the last sentence, in particular how they affect the EV of the the CQ *Which guy did Lena talk to?*. I will discuss this example in a bit more detail, also in order to illustrate how the proposed model works. For simplicity, I assume a very reduced set of possible questions $\mathcal{Q} = \{q_1, q_2, q_3, q_4, q_5\}$. The QUD stack, or rather the QUD set, is explicated in (37) for the initial commitment space C_0 . It includes the pairs of all possible questions with their respective prior EV.

$$(37) \quad \mathcal{S}_{qud} = \{$$

$$\langle q_1: What did Lena do after the party?, 0 \rangle,$$

$$\langle q_2: What happened to Lena and the guy after the party?, 0 \rangle,$$

$$\langle q_3: Which guy did Lena talk to?, 0 \rangle,$$

$$\langle q_4: What happened to Lena at the party?, 1 \rangle,$$

$$\langle q_5: How was the conversation?, 0 \rangle\}$$

I assume (36) to take place in a context in which it is common knowledge that we are talking about Lena and yesterday's party. Based on this, I assume that only q_4 will be an a priori expected question in C_0 , all the other questions are assumed to be not at all expected. Strictly speaking, those unexpected questions should not have the value 0 but a very low EV. The value 0 would mean that those questions have been answered, as defined in Definition 30. Furthermore, the definition of \mathcal{S}_{qud} only allows questions above a certain threshold n , which would exclude all questions except q_4 from \mathcal{S}_{qud} in (37). However, I will ignore those restrictions for the sake of a less complex illustration of the model. Therefore, I assume that all questions stay in \mathcal{S}_{qud} , and that they can have an EV of 0 even though they have not been answered.

Table 10.1 presents how the result of applying f_e to each question and each commitment space changes, progressing from commitment space C_0 to C_3 . As mentioned before, the real values would have to be determined empirically. Since there are no overt questions in (36), only step *ii.* of f_e is relevant, which is repeated here.

$$(38) \quad Step \text{ } ii. \text{ of } f_e \text{ for an assertive speech act:}$$

For any $C \in \mathcal{CS}$, any $q \in \mathcal{Q}$, that is open in C , and any **assertive speech act** A_φ addressing the CQ cq_φ ,

$$f_e(C+A_\varphi)(q) \propto \begin{cases} f_e(C)(q) + \alpha & \text{iff } q \text{ is a PQ licensed by } \varphi. \\ max((f_e(C)(q) - \epsilon), 0) & \text{iff } cq_\varphi \text{ q-entails q. } \epsilon \text{ differs} \\ & \text{with respect to completeness of } \varphi \text{ as answer to q.} \\ f_e(C)(q) & \text{otherwise.} \end{cases}$$

[a-PQ] [a-CQ] [OTHER]

For the purpose of illustration, I stipulate the values for the variables α and ϵ , for PQs and CQs respectively, such that they make the correct predictions. I will distinguish two kinds of PQs and respective α 's. For a strongly expected PQ, I stipulate $\alpha_1 = 0.5$ and, for a weakly expected PQ, I set $\alpha_2 = 0.3$. Furthermore, I assume $\epsilon = 0.8$ for a rather satisfactory answer (in van Kuppevelt's (1995) sense of satisfactoriness). Those variables can, in principle, take any values, it does not have to be a number smaller than 1. However, the normalization will always yield final EVs between 0 and 1.

q	$f_e(C_0)(q)$	$f_e(C_1)(q)$	$f_e(C_2)(q)$	normalized
q_1	0	$f_e(C_0)(q_1) = 0$	$f_e(C_1)(q_1) + \alpha_2 = 0.3$	0.23
q_2	0	$f_e(C_0)(q_2) = 0$	$f_e(C_1)(q_2) + \alpha_2 = 0.3$	0.23
q_3	0	$f_e(C_0)(q_3) + \alpha_1 = 0.5$	$f_e(C_1)(q_3) = 0.5$	0.38
q_4	1	$\max(f_e(C_0)(q_4) - \epsilon) = 0.2$	$f_e(C_1)(q_4) = 0.2$	0.15
q_5	0	$f_e(C_0)(q_5) + \alpha_2 = 0.3$	0 (answered)	0

q	$f_e(C_3)(q)$	normalized
q_1	$f_e(C_2)(q_1) + \alpha_1 = 0.73$	0.41
q_2	$f_e(C_2)(q_2) + \alpha_2 = 0.53$	0.3
q_3	$f_e(C_2)(q_3) = 0.38$	0.21
q_4	$f_e(C_2)(q_4) = 0.15$	0.08
q_5	$f_e(C_2)(q_5) = 0$	0

Table 10.1: Application of the model for example (36) for commitment spaces C_0 – C_3 and $\alpha_1 = 0.5$, $\alpha_2 = 0.3$, and $\epsilon = 0.8$

I will now go through each of the updates, and explain its effect on expectedness on each of the questions in \mathcal{Q} . In C_1 , after the update with the first sentence (*Yesterday at the party, Lena talked to some guy*), q_1 (*What did Lena do after the party?*) and q_2 (*What happened to Lena and the guy after the party?*) do not change their EV, since they are not affected by this update. Accordingly, they fall under the OTHER-condition of (38) and stay 0. I interpret q_3 (*Which guy did Lena talk to?*) as a strong PQ of the first sentence and, thus, $\alpha_1=0.5$ is added to its EV by a.PQ. The question q_5 (*How was the conversation?*) is assumed to be a weak PQ of the first sentence and, again, according to a-PQ, $\alpha_2=0.3$ is added to its EV. The question q_4 (*What happened to Lena at the party?*) is the CQ of the

first sentence and it is answered rather satisfactorily, though not completely. Therefore, $\epsilon=0.8$ is subtracted from the EV of q_4 , by a-CQ in (38). The normalization in C_1 has no effect, since the values add up to 1 anyway. We see here how one assertion can at the same time answer a question and reduce its EV, as for q_4 , and evoke questions and raise their EVs, as for q_3 and q_5 .

In C_2 , after the update with the second sentence (*The two of them laughed a lot and they agreed to meet again the next evening.*), q_1 (*What did Lena do after the party?*) and q_2 (*What happened to Lena and the guy after the party?*) are interpreted as weak PQs of the second sentence. The expression *meet again the next evening* probably raises the expectations for those questions to be addressed. According to a-PQ, $\alpha_2=0.3$ is added to their EVs. The EV of q_3 (*Which guy did Lena talk to?*) falls under the OTHER-condition and does not change since its expectedness does not seem to be affected by the new update. Also the EV of q_4 (*What happened to Lena at the party?*) is not affected and stays the same. However, one could also assume that q_4 is further reduced since the CQ of the second sentence in a way contributes to the higher question. This would, however, require some adjustments in the definition of f_e . The value of q_5 (*How was the conversation?*) is set to 0 since the second sentence fully answers q_5 . After the application of f_e , the values are normalized for them to still add up to 1.

In C_3 , both EVs of q_1 (*What did Lena do after the party?*) and q_2 (*What happened to Lena and the guy after the party?*) increase. I treat those questions as PQs of the third sentence (*Then, Lena went home happily*), triggered by the discourse progressive element *then*. I take q_1 to be a stronger PQ than q_2 because, intuitively, it seems that the particle *then* seems to move the attention away from the conversation of Lena and the guy but not away from what happened to Lena. Therefore, $\alpha_1=0.5$ is added to q_1 , and $\alpha_2=0.3$ is added to q_2 . The correct prediction for the acceptability of the cleft in the next step, however, does not rely on one PQ being stronger than the other one, as long as they increase. It would not change the outcome with respect to the acceptability of the cleft if they were switched or if they both received +0.5. Question q_3 , q_4 and q_5 fall under the OTHER-condition and keep their EV. Again, the values are normalized.

The final \mathcal{S}_{qud} in C_3 , that serves as the basis for the cleft or the canonical continuation, is shown in (39).

$$(39) \quad \mathcal{S}_{qud} = \{$$

$$\langle q_1: \text{What did Lena do after the party?}, 0.41 \rangle,$$

$$\langle q_2: \text{What happened to Lena and the guy after the party?}, 0.3 \rangle,$$

$$\langle q_3: \text{Which guy did Lena talk to?}, 0.21 \rangle,$$

$$\langle q_4: \text{What happened to Lena at the party?}, 0.08 \rangle,$$

$$\langle q_5: \text{How was the conversation?}, 0 \rangle\}$$

The CQ of the cleft, cq_{cl} , and the CQ for the canonical sentence, cq_{can} , are identical to $q_3: \text{Which guy did Lena talk to?}$. Hence, they have the EV 0.21 in C_3 . Now, I follow Definition 34 and Definition 32 to determine the acceptability of the cleft and the canonical sentence in example (36). For this purpose, I stipulate the thresholds $e_{can} = 0.25$ and $e_{cl} = 0.1$. With these thresholds, the model would correctly predict the cleft in (36-a) to be acceptable (by Definition 34). The cleft is focus congruent to cq_{cl} and the EV of the current question cq_{cl} of the cleft falls below e_{can} and above e_{cl} , as indicated in (40).

$$(40) \quad e_{can} > f_e(C_3)(cq_{cl}) = f_e(C_3)(q_3) > e_{cl}$$

$$0.25 > 0.21 > 0.1 \quad \checkmark$$

The CQ, cq_{can} , of the canonical sentence in (36-b) also receives the EV 0.21, since it addresses the same CQ as the cleft. However, its EV still does not exceed the threshold e_{can} required for acceptable unmarked sentences, as indicated in (41).

$$(41) \quad f_e(C_3)(cq_{can}) = f_e(C_3)(q_3) \not> e_{can}$$

$$0.21 \not> 0.25 \quad \times$$

The canonical sentence is predicted to be unacceptable because the requirement in Definition 32 that the EV of the CQ of the canonical sentence must be greater than e_{can} is not met. Hence, the model successfully predicts the difference in acceptability of the cleft and the canonical sentence in example (36).

An important message to take away from this example is that the EV of q_3 , which is later identified with the CQ of the cleft, decreases from update to update just because the EVs of other questions (q_1 and q_2) increase. This is how the length of the inquiry is naturally incorporated as a predictor of acceptability of the cleft. It is, however, not the length per se, but also the relative expectedness of other questions after each update. If the new questions had a low EV, the cleft would be predicted to be less acceptable and the canonical sentence might be preferred.

I will discuss example (8) and (5) from section 10.1, repeated in (42) and (43), in less detail, but will provide the gist of it. Consider first example (42).

- (42) Lena hat gestern auf der Party mit einem Typen₁ gesprochen. Die beiden haben viel gelacht und sich direkt für den nächsten Abend verabredet. **Dann hat Lena ihm sogar ein Geheimnis verraten.**

'Yesterday at the party, Lena talked to some guy₁. The two of them laughed a lot and they agreed to meet again the next evening. Then, Lena even told him a secret.'

- a. ?Es war Peter₁, mit dem sie gesprochen hat.
it was Peter₁ with whom she talked has
'It was Peter₁ she talked to.'
- b. ?Sie hat mit Peter₁ gesprochen.
She has with Peter₁ talked
'She talked to Peter₁'

The problem with this example was that even the cleft is not acceptable anymore when the last sentence in the context is replaced with a sentence that evokes a very pressing question. My model accounts for this because it predicts the cleft to be unacceptable in this example, given that the update *Then, Lena even told him a secret* evokes the PQ *What was the secret?* with a very high EV. This would mean that the EV of the CQ of the cleft *Which guy did Lena talk to?* would be pushed below the cleft threshold e_{cl} and the cleft would be predicted to be an unacceptable discourse move, just like the canonical sentence of course.

Consider example (43), which caused the problem that clefts may address unexpected issues but not entirely unexpected ones.

- (43) **Lena ist gestern Abend bei der Party angekommen und hat erstmal einen leckeren Cocktail getrunken.** Danach hat sie mit ihrer Freundin Andrea getanzt und die beiden hatten sehr viel Spaß. Dann ist Lena glücklich nach Hause gefahren.

'Lena arrived at the party yesterday and first of all she had a tasty cocktail. Thereafter, she danced with her friend Andrea and the two of them had a lot of fun. Then, Lena went home happily.'

- a. ?Es war ein Bloody Mary, den sie getrunken hat.
it was a Bloody Mary that she drunk has
'It was a Bloody Mary she drank.'
- b. ?Sie hat einen Bloody Mary getrunken.
She has a Bloody Mary drunk
'She drank a Bloody Mary.'

The CQ of (43-a) and (43-b) (*Which cocktail did Lena drink?*) is evoked by the first sentence, and can be assumed to be a rather weak PQ. Hence, it receives a low EV. Given the intervening material, that again raises new questions, this already low value will be pushed down further, most likely below the cleft threshold e_{cl} , and thereby below the threshold e_{can} . Therefore, both the canonical sentence and the cleft are inappropriate discourse moves in this example.

10.5 Discussion and Outlook

In this chapter, I presented an approach to German *es*-clefts that analyzes them embedded in a broader discourse context, and discusses how they structure discourse. The analysis was based on several examples, many of which constituted rare, but insightful, cases in which the clefts were acceptable while the canonical sentences were not.

In order to describe the discourse function of the cleft, I made some adjustments on question-based discourse models such as the one presented in chapter 4. Building on Roberts (2012), I introduced an adapted QUD stack. It departed from the original by not only including current questions into the stack but also potential questions that were evoked by the preceding text. Furthermore, I added the concept of expectedness, which describes for each possible question in a given commitment space how strongly the addressee expects a certain question to be addressed. Expectedness is formalized by a function that involves parameters that can be estimated on a data set. I also formalized the process of accepting discourse moves based on expectedness, which makes the predictions of the model testable.

The differences between the acceptability of clefts in discourse, in contrast to canonical sentences, was modeled by different thresholds of expectedness that needed to be exceeded for the respective sentence type. More specifically, the cleft was hypothesized to address less expected questions than the canonical sentence. By requiring a lower expectedness threshold for clefts than for canonical sentences, the difference in acceptability could be correctly predicted by the model.

This model has features that are worth pointing out once again. By making expectedness a recursive function, it incorporates the progression of discourse updates, and how that affects the acceptability of clefts. It explains why clefts often address questions which are not raised right before, but which prefer some distance between the question evoking sentence and the cleft answer. To my knowledge, previous approaches had not properly captured that. Velleman et al. (2012) did consider the length of the inquiry. The model proposed here, however, provides a more convincing explanation for their examples. Furthermore, it can account for examples not captured by Velleman et al. (2012) (see section 10.4).

Another benefit that comes with the proposed model is that accommodation of the current question (CQ) and accepting a discourse move involve the same underlying mechanism, based on expectedness and a certain threshold. This is desirable because accommodation

can be understood as deciding to accept that a question *could* be addressed, which is quite similar to deciding to accept that a question *is* addressed.

Relation of the model to other approaches Finally, I will show why my model makes better predictions than many other approaches. Let us first look at exhaustivity. Approaches that make exhaustivity the main function of clefts struggle to explain the difference observed in (36), repeated again below.

- (44) Lena hat gestern auf der Party mit einem Typen₁ gesprochen. Die beiden haben viel gelacht und sich direkt für den nächsten Abend verabredet. Dann ist Lena glücklich nach Hause gefahren.

'Yesterday at the party, Lena talked to some guy₁. The two of them laughed a lot and they agreed to meet again the next evening. Then, Lena went home happily.'

- a. Es war Peter₁, mit dem sie gesprochen hat.
it was Peter₁ with whom she talked has
'It was Peter₁ she talked to.'
- b. ?Sie hat mit Peter₁ gesprochen.
She has with Peter₁ talked
'She talked to Peter₁.'

First of all, the reply in (44-a) does not require exhaustivity. The answer is compatible with Lena having talked to other people as well. This excludes semantic approaches to exhaustivity. Other exhaustivity approaches that are based on the anaphoric resolution to a discourse referent do not manage to explain this example either. Pollard and Yasavul (2014) would correctly predict (44-a) to be interpreted non-exhaustively. The anaphoric referent is resolved to the first sentence *Yesterday at the party Lena talked to some guy*. The pivot *Peter* in the cleft in (44-a) is identified with the referent which the indefinite description refers to. This does not imply there to be no other person or guy that Lena talked to. What Pollard and Yasavul (2014) do not aim to capture is the difference between the cleft and the canonical sentence in (44). Hence, this remains to be explained.

Onea's (2019a) principle does not make the correct predictions, either, since it predicts all clefts to provide a maximal answer. The acceptability of the cleft in (44-a) can only be

accounted for if the principle does not apply because it is a non-default cleft. I argue that this example, and other similar examples in this thesis, do not seem to be non-default cases, but rather standard cases of clefts in discourse.

My model does not require the cleft to provide a maximal answer to the question it addresses. Therefore, it can account for examples like (44), in which the answer does not mean that Lena did not talk to anybody else. Even though my approach does not focus on exhaustivity and maximality, it is compatible with the observation that clefts do quite frequently express exhaustivity. I analyze exhaustivity as a side effect of the discourse function of the cleft. My admittedly speculative explanation is that the reader/hearer might pragmatically infer the following: Given that the author/speaker bothered to pick up a question that was already settled or decreasing in expectedness, she probably has a complete/satisfactory answer to it, which would justify addressing it even though it is not expected. The fact that the exhaustivity inference was shown to be cancelable (e.g., by Horn, 1981) speaks in favor of it being a pragmatic inference anyway. This issue needs more profound investigations on the interaction of the expectedness of addressed questions and exhaustivity.

Let us now take a look at some of the various discourse functions that have been claimed for clefts (see section 8.2 and section 8.3). My approach can be interpreted as formalizing Hedberg's (1990) approach, which stated that the antecedent of the cleft relative clause requires different degrees of accessibility. The accessibility could be interpreted as the expectedness of this antecedent to be questioned. My approach adds to her approach, though, that the cleft is not only compared to other versions of the cleft but also to canonical sentences, which is crucial for explaining the difference in (44).

The various discourse relations realized by Hedberg (1990) and Delin and Oberlander (1995), among others, can be reinterpreted as being related to less expected questions, as will be explicated in chapter 11.

One might wonder whether this approach makes different predictions than Destruel and Velleman (2014) and Destruel et al. (2019), who also made their predictions based on

expectations of the addressee. They focus on expectedness in relation to at-issueness. I argue that the relation between expectedness and at-issueness still needs some further research. Destruel and Velleman (2014) claim that clefts address not-at-issue content rather than at-issue content, which would be in line with expectedness. It can be assumed that it is more expected to address a question that is at-issue instead of not-at-issue. Destruel et al. (2019), to the contrary, found that at-issueness did not play such an important role. I argue, however, that this result might not have been caused by at-issueness, but due to the distance between the cleft and their antecedent in their stimuli. The factor distance is something that my model can account for, as mentioned above. This needs further empirical research that systematically distinguishes at-issueness from distance.

More generally, my approach is in line with the claims made by Destruel and Velleman (2014), repeated in (45).

(45) *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 and Velleman, 2014:199)

Basically, my approach only refers to expectations about the discourse, modulo making a claim about at-issueness. Actually, I would argue that expectations about the world are not relevant for the acceptability of clefts. I propose that those acceptable clefts which involve conflicting expectations about the world will also involve conflicting expectations about the discourse. If a cleft only conflicts with expectations about the world, it is,

therefore, predicted to be unacceptable. This approach is supported by the following two examples.

- (46) Speaker A: What did Peter buy?

Speaker B: ?It was a kangaroo that Peter bought!

- (47) Beginning at the top of the list, I went along the landing and tapped at Ruskin's door. When it was opened, it was Webber who stood there. We stared at each other for a moment, both of us taken aback.

[Lucille Kallen, *The Piano Bird*, p. 95]

(Hedberg, 1990:120)

Example (46) only conflicts with expectations about the world, buying a kangaroo is unexpected, but not about the discourse. Addressing the explicit question is expected. In this case the cleft is marginally acceptable at best. In Hedberg' (1990) example, repeated in (47), there is a conflict between the expectations of who would answer the door, and who actually did (expectations about the world). However, there is also a conflict with respect to the expectations about the discourse. The question *Who stood there?* is not particularly expected to be addressed in the context. Given that somebody knocked at Ruskin's door, and the door was opened, the reader would probably already assume that Ruskin would stand there. Thus, other questions would be more expected, such as *What happened then?*. Comparing (46) and (47), I conclude that it is expectations about the discourse that crucially determine the acceptability of clefts.

Zimmermann (2011b) links expectedness to argument asymmetry in focus marking. He argues that subjects are less likely to be focused since the subject position is usually the position in which the topic is realized. Hence, he considers a subject focus to constitute an unexpected discourse development. According to his contrastive focus hypothesis, focus on subject must, therefore, be marked more than on objects, for instance. If we follow this reasoning, we can also explain the asymmetry between subject clefts and object clefts, by adding a requirement to the expectedness function that subject *wh*-questions are generally less expected than object *wh*-questions.

Open issues and possible extensions An important next step would be to determine the variables of the model by conducting a series of suitable experiments, which could be inspired by Westera and Rohde (2019). Alternatively, one could run a model on a data set in order to estimate the variables. In the course of that, it would make sense to investigate how the cases in the expectedness function could be refined. As was already mentioned in section 10.4, differently strong potential questions can be sub-categorized, e.g. likely potential questions might be more expected than other potential questions. Furthermore, one might want to include dependent questions, as in (48).

(48) A: Who is going to win the race?

Dependent question: Who is going to participate?

Furthermore, there are some examples that are not covered by the proposed model. For instance, example (9), repeated in (49), remains to be explained.

(49) Lena hat gestern auf der Party mit einem Typen₁ gesprochen. Die beiden haben viel gelacht und sich direkt für den nächsten Abend verabredet. Lena hat ihm sogar ein Geheimnis verraten. **Dann ist Lena glücklich nach Hause gefahren.**
'Yesterday at the party, Lena talked to some guy₁. The two of them laughed a lot and they agreed to meet again the next evening. Lena even told him a secret. Then, Lena went home happily.'

- a. Es war Peter₁, mit dem sie gesprochen hat.
 it was Peter₁ with whom she talked has
 'It was Peter₁ she talked to.'
- b. ?Sie hat mit Peter₁ gesprochen.
 She has with Peter₁ talked
 'She talked to Peter₁.'

In this example, there is a pressing question *q*: *What was the secret?* intervening between the cleft and the sentence that evoked the CQ of the cleft (*Which guy did Lena talk to?*). My model would predict *q* to have very high EV that pushes the value of the CQ of the cleft below the cleft threshold, and would make the cleft unacceptable. Indeed, the cleft

would be unacceptable if the bold sentence was left out. However, with the sentence in bold, the cleft seems to become more acceptable again.

So far, my approach cannot account for the effect of the last sentence, given that the sentence neither evokes a new very expected question, nor answers the question q (*What was the secret?*). Thus, nothing explains why the high EV of q would not still push down the EV of the cleft's CQ, and make the cleft unacceptable. In order to capture this example, one would probably have to incorporate the effect of a topic shift into the model. For instance, van Kuppevelt (1995) presents an informal account on discourse topics and related discourse questions, which could be a good starting point. Furthermore, it seems that the speaker in (49) uses the sentence in bold to mark that she/he refuses to tell what the secret was. This seems to have an effect on the expectedness of q in this case, namely reducing the EV of q . In that case, the EV of the CQ could be high enough again for the cleft to be acceptable. It needs to be investigated how exactly the EV is reduced here. Is it because of the marker *then*? Or just pragmatic reasoning?

Another problematic, but maybe not too problematic, case is correction. As many have noticed (e.g., Destruel and Velleman, 2014; Destruel et al., 2019), clefts are frequently used for correction. Correction is problematic for my account because it addresses a question that has already been answered and, thus, receives the EV 0. My model would, therefore, incorrectly predict that a cleft cannot address that question. In order to account for such cases, one would have to assume something like the denegation of a speech act, such as Krifka (2015) proposes. Alternatively, one could use a discourse restructuring mechanism (e.g., Farkas and Bruce, 2010). I leave the interaction between the cleft and this discourse restructuring for future research.

Chapter 11

Conclusion

This thesis investigated the function of German *es*-clefts in discourse. Part I of the thesis provided the theoretical background on which my analysis of clefts is based. It introduced different approaches to the linguistic representation of questions as speech acts, as they appear in a discourse update, and as semantic objects, as they are integrated into a QUD stack. My approach to clefts relies on an understanding of clefts as speech acts, because accepting a discourse move involves anticipating different interrogative speech acts. With respect to the semantic representation of questions, my approach to clefts is compatible with all the semantic approaches to questions presented in chapter 2.

Based on this, chapter 3 developed a representation of focus as indicating an implicit question, which is later identified as the current question (CQ). In chapter 4, I presented a discourse model that is set in Roberts' (2012) QUD framework. The discourse model in that chapter involved a new perspective on question-based discourse models, building on the notion of focus presented in section 3.2. Crucially, it disentangles the CQ from visible focus marking while still assuming that focus does indicate the CQ. Following Velleman et al. (2012), the CQ is not given by focus but by the speaker.

A consequence of this new perspective is that the accommodation of this CQ gains in importance, since the speaker may not always choose to address the most likely question, and the CQ is not always marked unambiguously by intonation. This seems to be a

plausible result, given the parallels between accommodating the CQ and accepting a discourse move. Deciding which question the speaker might have wanted to address (accommodation) seems to be similar to deciding whether it is acceptable that the speaker addressed a question (accepting a discourse move). As I pointed out in section 10.2, both processes involve expectedness.

Part II of the thesis was concerned with previous approaches to *es*-clefts in German but also to clefts and related structures in other languages. It presented an overview of the different syntactic approaches in chapter 5. The syntax of clefts is a topic of its own, which I was not aiming to cover exhaustively in this thesis. My overview mainly showed that there is a discrepancy between what the cleft structure resembles structurally, i.e. a copular construction, and what it represents semantically and pragmatically, i.e. something that is closer to a focus fronting construction. The different approaches each focused on a specific aspect of the cleft structure and built their syntactic analysis based on that aspect. Those aspects were, for instance, the (non-)referential status of the cleft pronoun, the integration of the relative clause into the structure, parallels to other structures, such as definite descriptions, *wh*-clefts, or Hungarian preverbal focus, etc.

Another line of research concerned exhaustivity of clefts, as discussed in chapter 7. The chapter discussed different categorizations of the exhaustivity inference of clefts, such as entailment, presuppositions, scalar implicatures, generalized implicatures, and other pragmatic inferences in the wider sense. Although many of the approaches treated exhaustivity as an inference which is lexically encoded in the cleft structure, the chapter concluded that the exhaustivity inference of clefts must be a pragmatic inference in the wider sense. It was left open how exactly this inference should be derived.

Chapter 8 discussed various functions that had been proposed in the literature on clefts in discourse, such as marking contrast or givenness, establishing text coherence, disambiguating focus, or requiring to answer a certain question. Those functions were shown to each describe only some functions of clefts, and some approaches covered more than others.

Before beginning my own analysis, chapter 9 presented the empirical findings concerned with exhaustivity and contrast in clefts. Previous experiments on exhaustivity showed that the exhaustivity inference is not a particularly robust inference. The *yes, but...* test revealed that clefts differ significantly from exclusives, suggesting that exhaustivity is a not-at-issue inference (c.f. Onea and Beaver, 2009; Destruel et al., 2015). A study by De Veaugh-Geiss et al. (2018b), involving a picture verification/falsification task, showed that participants were split into two groups either treating clefts exhaustively or non-exhaustively. This speaks in favor of a pragmatic inference. The experiments on contrast showed diverging results with respect to at-issueness and could not settle the relation of contrast and at-issueness. Moreover, the chapter presented some findings from corpus studies, which mainly pointed out the relevance of the wider discourse context for the analysis of clefts. The experiments frequently neglected the context.

The previous approaches are followed by my main contribution to the analysis of clefts, presented in section 10.2. I proposed the following hypotheses for the acceptability of German *es*-clefts in comparison to canonical sentences.

- (1) a. An *es*-cleft addresses a question that came up in the preceding context, but that the addressee does not expect to be answered at that point in the discourse compared to other questions.
- b. Those questions that are more expected are preferably addressed with a canonical sentence instead of a cleft.
- c. Those questions that are neither particularly expected nor particularly unexpected can equally well be addressed by a cleft or a canonical sentence.

I formalized expectedness as a function that assigns an expectedness value between 0 and 1 to each possible question that one could ask with respect to a possible context. In general, the process of accepting a discourse move was modeled as the process of considering whether the question addressed by that move is expected enough. If the question is too unexpected, the speaker should not address it in a discourse move, and the addressee would not accept the discourse move.

The conditions for accepting a cleft, in contrast to accepting a canonical sentence, were modeled via thresholds of expectedness values. A canonical sentence has to exceed a certain threshold of expectedness in order to be acceptable. A cleft, in contrast, was required to address a question with an expectedness value lower than the threshold for canonical sentences. Furthermore, a threshold for clefts was introduced to account for the fact that even clefts may not address questions that are too unexpected. Another useful feature of the model is that it incorporates the incremental change of the context by making the expectedness function recursive. This way, the model captures how the expectedness of questions is affected by every new discourse update.

This analysis of clefts is, to my knowledge, the only analysis which can account for the differences between clefts and canonical sentences in discourse as well as for the dependency of exhaustivity on the context. Moreover, it traces many of the previously observed discourse functions of clefts back to expectedness, as discussed in section 10.5.

Of course, this analysis does not manage to explain every example, but it leaves possibilities for extensions. As mentioned in section 10.5, correction cannot be explained by the model. However, that is a problem independent of the current analysis since correction in general requires discourse moves to not be finally accepted right away. Otherwise, an assertion made by an accepted discourse move cannot be picked up again for the sake of correction. Furthermore, there are many more factors which influence the expectedness of a question, which, as a consequence, influences the acceptability of the cleft and the canonical sentences, such as dependencies between questions that are not covered by q-entailment. Those factors are not covered by the expectedness function yet but could be incorporated by additional cases.

I propose that my model not only accounts for the difference between clefts and canonical sentences in German; it provides a useful discourse model in a more general sense, in that it makes predictions for the acceptability of discourse moves based on expectedness. Therefore, an interesting research question would be whether the model can be extended to also account for the acceptability of other constructions that might have to do with expectedness. Some of them are presented in (2).

- (2) Lena hat gestern auf der Party mit einem Typen₁ gesprochen. Die beiden haben viel gelacht und sich direkt für den nächsten Abend verabredet. Dann ist Lena glücklich nach Hause gefahren.

'Yesterday at the party, Lena talked to some guy₁. The two of them laughed a lot and they agreed to meet again the next evening. Then, Lena went home happily.'

- a. Sie hatte mit Peter₁ gesprochen. [PAST PERFECT]
'She had talked to Peter₁.'
- b. Mit Peter₁ hat sie gesprochen. [MOVEMENT]
'To Peter₁, she talked.'
- c. Der Typ, mit dem sie gesprochen hat, war Peter₁. [DEFINITE DESCRIPTION]
'The guy she talked to was Peter₁'
- d. Sie hat doch tatsächlich mit Peter₁ gesprochen. [PARTICLE]
'She PRT actually talked to Peter₁'

All of these constructions seem to be more acceptable than just the canonical sentence *She talked to Peter*. The model could be extended to allow different constructions, not only the cleft, to lower the threshold for acceptable discourse moves in order to predict the acceptability of the examples in (2).

The model might also make the correct predictions for the impact of clefts on discourse relations. It has been observed that clefts have a tendency to not realize the discourse relation *Narration* or *Elaboration*, as defined by Asher and Lascarides (2003). Example (3) shows that the cleft, as opposed to the canonical sentence, is not interpreted as realizing an elaboration.

- (3) Der Bus fuhr Schlangenlinien.

'The bus zigzagged along the road.'

- a. Am Steuer saß Peter. Sein Chef hatte ihm befohlen entlang der Schlangenlinien am Boden zu fahren.
'Behind the wheel sat Peter. His boss had told him to follow the zigzag lines on the road.'

- b. Es war Peter, der am Steuer saß. ?Sein Chef hatte ihm befohlen entlang der Schlangenlinien am Boden zu fahren.

'It was Peter who sat behind the wheel. His boss had told him to follow the zigzag lines on the road.'

The canonical sentence in (3-a) stands in the *Elaboration* relation to the preceding sentence. The second sentence further elaborates. It represents a well-formed discourse. The second sentence in example (3-b), in contrast, is only marginally acceptable because the addressee has to perform a reinterpretation of the discourse relation determined for the cleft once she/he has processed the last sentence. First, the cleft is interpreted as the *Explanation* for the first sentence, implying that Peter is a bad driver, or always drunk. However, in combination with the last sentence, this interpretation does not make sense anymore since that sentence provides a different explanation for the first sentence (Peter's boss told him). Hence, one would have to find a different discourse relation holding between the cleft and the first sentence.

If we assume that *Narration* and *Elaboration* are the most frequent discourse relations while the others are less frequent, the proposed model can predict the difference in (3) based on expectedness. The hypothesis would be that clefts only address questions that realize less frequent and, thus, less expected discourse relations. This would have to be tested with more data and different discourse relations.

It might also be worth investigating the relation of expectedness and suspense, an issue that was already raised in subsection 10.2.1. The two concepts seem to be related. It seems that once suspense begins to build, the addressee does not expect the speaker/writer to address certain issues. However, the more the suspense increases, the more the addressee again expects those questions to be addressed since suspense cannot go on forever. Hence, this is the end.

Bibliography

- Abrusán, M. and K. Szendrői (2013). Experimenting with the King of France: Topics, verifiability and definite descriptions. *Semantics and Pragmatics* 6(10), 1–43.
- Akmajian, A. (1970). On deriving cleft sentences from pseudo-cleft sentences. *Linguistic Inquiry* 1(2), 149–168.
- Asher, N. and A. Lascarides (2003). *Logics of Conversation*. Studies in Natural Language Processing. Cambridge: Cambridge University Press.
- Atlas, J. D. and S. C. Levinson (1981). It-clefts, informativeness and logical form: Radical pragmatics (revised standard version). In *Radical Pragmatics*, pp. 1–62. New York: Academic Press.
- Beaver, D. I. and B. Z. Clark (2008). *Sense and Sensitivity: How Focus Determines Meaning*, Volume 5 of *Explorations in Semantics*. Wiley-Blackwell.
- Benz, A. and K. Jasinskaja (2017). Questions Under Discussion: From Sentence to Discourse. *Discourse Processes* 54(3), 177–186.
- Biezma, M. and K. Rawlins (2017). Rhetorical questions: Severing asking from questioning. In D. Burgdorf, J. Collard, S. Maspong, and B. Stefánsdóttir (Eds.), *Proceedings of SALT*, Volume 27, pp. 302–322.
- Bolinger, D. L. (1958). A Theory of Pitch Accent in English. *Word* 14(2–3), 109–149.
- Bolinger, D. L. (1972). A look at equations and cleft sentences. In E. S. Firchow (Ed.), *Studies for Einar Haugen: Presented by Friends and Colleagues*, pp. 96–114. The Hague: De Gruyter.

Büring, D. (2003). On D-trees, beans, and B-accents. *Linguistics and Philosophy* 26(5), 511–545.

Büring, D. and K. Hartmann (2001). The Syntax and Semantics of Focus-Sensitive Particles in German. *Natural Language & Linguistic Theory* 19(2), 229–281.

Büring, D. and M. Kriz (2013). It's that, and that's it! Exhaustivity and homogeneity presuppositions in clefts (and definites). *Semantics and Pragmatics* 6(6), 1–29.

Byram Washburn, M., E. Kaiser, and M. L. Zubizarreta (2019). The English *It-Cleft*: No Need to Get Exhausted. In K. von Heusinger, E. Onea, and M. Zimmermann (Eds.), *Questions in Discourse*, pp. 198–236. Leiden, Boston: Brill.

Caponigro, I. and J. Sprouse (2007). Rhetorical questions as questions. In E. Puig-Waldmüller (Ed.), *Proceedings of Sinn und Bedeutung*, Volume 11, pp. 121–133.

Carlson, L. (1982). *Dialogue Games: An Approach to Discourse Analysis*. PhD thesis, MIT.

Carter-Thomas, S. (2009). The French c'est-cleft: Function and frequency. In D. Banks (Ed.), *La linguistique systémique fonctionnelle et la langue française*, pp. 127–157. Paris: L'Harmattan.

Cheng, L. L. (1991). *On the Typology of Wh-Questions*. PhD thesis, MIT.

Chomsky, N. (1982). *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, MA: MIT press.

Ciardelli, I. (2014). Interrogative Dependencies and the Constructive Content of Inquisitive Proofs. In U. Kohlenbach, P. Barceló, and R. de Queiroz (Eds.), *Logic, Language, Information, and Computation. WoLLIC 2014. Lecture Notes in Computer Science*, Volume 8652, pp. 109–123. Berlin, Heidelberg: Springer.

Ciardelli, I., J. A. G. Groenendijk, and F. Roelofsen (2013). Inquisitive Semantics: A New Notion of Meaning. *Language and Linguistics Compass* 7(9), 459–476.

- Cohen, A. and M. Krifka (2014). Superlative quantifiers and meta-speech acts. *Linguistics and Philosophy* 37(1), 41–90.
- Collins, P. C. (1991). Pseudocleft and cleft constructions: A thematic and informational interpretation. *Linguistics* 29(3), 481–520.
- Coppock, E. and D. I. Beaver (2011). Sole sisters. In *Proceedings of SALT*, Volume 21, pp. 197–217.
- Cremers, A. and E. Chemla (2016). A Psycholinguistic Study of the Exhaustive Readings of Embedded Questions. *Journal of Semantics* 33(1), 49–85.
- Csipak, E. and S. Zobel (2014). A condition on the distribution of discourse particles across types of questions. In *Proceedings of NELS*, Volume 44, pp. 83–94.
- De Veaugh-Geiss, J. P., S. Tönnis, E. Onea, and M. Zimmermann (2018a). An experimental investigation of (non-)exhaustivity in *es*-clefts. In *Proceedings of Sinn und Bedeutung*, Volume 21, pp. 347–366.
- De Veaugh-Geiss, J. P., S. Tönnis, E. Onea, and M. Zimmermann (2018b). That's not quite it: An experimental investigation of (non-)exhaustivity in clefts. *Semantics and Pragmatics* 11.
- De Veaugh-Geiss, J. P., M. Zimmermann, E. Onea, and A.-C. Boell (2015). Contradicting (not-)at-issueness in exclusives and clefts: An empirical study. *Semantics and Linguistic Theory* 25, 373–393.
- Delin, J. and J. Oberlander (1992). Aspect-Switching and Subordination: the Role of *It*-Clefts in Discourse. In *Proceedings of the Fourteenth International Conference on Computational Linguistics (COLING-92)*, pp. 281–287.
- Delin, J. and J. Oberlander (1995). Syntactic Constraints on Discourse Structure: the Case of *It*-clefts. *Linguistics* 33(3), 465–500.
- Delin, J. L. (1990). *Cleft Constructions in Discourse*. PhD thesis, University of Edinburgh.

- Destruel, E. (2013). *The French C'est-cleft: Empirical studies of its meaning and use*. PhD thesis, The University of Texas.
- Destruel, E., D. I. Beaver, and E. Coppock (2018). Clefts: Quite the contrary! In *Proceedings of Sinn und Bedeutung*, Volume 21, pp. 335–346.
- Destruel, E., D. I. Beaver, and E. Coppock (2019). It's Not What You Expected! The Surprising Nature of Cleft Alternatives in French and English. *Frontiers in Psychology* 10, 1–16.
- Destruel, E., D. Velleman, E. Onea, D. Bumford, J. Xue, and D. I. Beaver (2015). A Cross-Linguistic Study of the Non-at-issueness of Exhaustive Inferences. In F. Schwarz (Ed.), *Experimental Perspectives on Presuppositions*, pp. 135–156. Cham: Springer.
- Destruel, E. and L. Velleman (2014). Refining contrast: Empirical evidence from the English *it*-cleft. *Empirical Issues in Syntax and Semantics* 10, 197–214.
- Doetjes, J., G. Rebuschi, and A. Rialland (2004). Cleft sentences. In F. Corblin and H. de Swart (Eds.), *Handbook of French Semantics*, Chapter 23, pp. 529–552. Stanford, CA: CSLI Publications.
- Drenhaus, H., M. Zimmermann, and S. Vasishth (2011). Exhaustiveness effects in clefts are not truth-functional. *Journal of Neurolinguistics* 24(3), 320–337.
- Dufter, A. (2009). Clefting and discourse organization. Comparing Germanic and Romance. In A. Dufter and D. Jacob (Eds.), *Focus and Background in Romance Languages*, pp. 83–121. Amsterdam: John Benjamins.
- É. Kiss, K. (1998). Identificational focus versus information focus. *Language* 74(2), 245–273.
- Eckardt, R. (2014). *The Semantics of Free Indirect Discourse: How Texts Allow Us to Mind-read and Eavesdrop*, Volume 31 of *Current Research in the Semantics/Pragmatics Interface*. Leiden: Brill.

Faller, M. (2002). *Semantics and Pragmatics of Evidentials in Cuzco Quechua*. PhD thesis, Stanford University.

Farkas, D. F. and K. B. Bruce (2010). On Reacting to Assertions and Polar Questions. *Journal of Semantics* 27(1), 81–118.

Féry, C. (2001). Focus and phrasing in French. In C. Féry and W. Sternefeld (Eds.), *Audiatur Vox Sapientia. A Festschrift for Arnim von Stechow*, pp. 153–181. Berlin: Akademie Verlag.

von Fintel, K. (2004). Would You Believe It? The King of France is back! (Presuppositions and Truth-Value Intuitions). In A. Bezuidenhout and M. Reimer (Eds.), *Descriptions and Beyond: An Interdisciplinary Collection of Essays on Definite and Indefinite Descriptions and other Related Phenomena*, pp. 325–341. Oxford University Press.

Fischer, K. (2009). Cleft Sentences: Form, Function, and Translation. *Journal of Germanic Linguistics* 21(2), 167–191.

Fricke, L. M., D. Blok, M. Zimmermann, and E. Onea (to appear). The pragmatics of embedded questions. An experimental comparison of four verbs of embedding.

Garro, L. and F. Parker (1982). Some suprasegmental characteristics of relative clauses in English. *Journal of Phonetics* 10(2), 149–161.

Geurts, B. and R. van der Sandt (2004). Interpreting focus. *Theoretical Linguistics* 30, 1–44.

Ginzburg, J. (2012). *The interactive Stance. Meaning for Conversation*. Oxford: Oxford University Press.

Grice, H. P. (1975). *Logic and Conversation*, pp. 41–58. Leiden: Brill.

Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, London: Harvard University Press.

- Groenendijk, J. A. G. (2015). Then I don't know. *Manuscript submitted at ILLC*.
- Groenendijk, J. A. G. and F. Roelofsen (2009). Inquisitive Semantics and Pragmatics. University of Amsterdam. Available online on <https://eprints.illc.uva.nl/348/>. Accessed 30 November 2020.
- Groenendijk, J. A. G. and M. J. B. Stokhof (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers*. PhD thesis, University of Amsterdam.
- Groenendijk, J. A. G. and M. J. B. Stokhof (1991). Dynamic Predicate Logic. *Linguistics and Philosophy* 14, 39–100.
- Grubic, M. (2015). *Focus and Alternative Sensitivity in Ngamo (West-Chadic)*. PhD thesis, University of Potsdam.
- Gundel, J. K. (1977). Where do Cleft Sentences Come from? *Language* 53(3), 543–559.
- Gundel, J. K. (1985). ‘Shared knowledge’ and topicality. *Journal of Pragmatics* 9(1), 83–107.
- Gundel, J. K., N. Hedberg, and R. Zacharski (1993). Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69(2), 274–307.
- Halvorsen, P.-K. (1978). *The Syntax and Semantics of Cleft Constructions*. PhD thesis, University of Texas.
- Hamblin, C. L. (1973). Questions in Montague English. *Foundations of Language* 10(1), 41–53.
- Hartmann, K. (2008). Focus and Emphasis in Tone and Intonational Languages. In A. Steube (Ed.), *The Discourse Potential of Underspecified Structures*, pp. 389–411. Berlin: De Gruyter.
- Hartmann, K. and M. Zimmermann (2007). In place-out of place? Focus in Hausa. In K. Schwabe and S. Winkler (Eds.), *On Information Structure, Meaning and Form*:

- Generalizations across languages*, pp. 365–403. Amsterdam, Philadelphia: John Benjamins.
- Hedberg, N. (1990). *Discourse Pragmatics and Cleft Sentences in English*. PhD thesis, University of Minnesota.
- Hedberg, N. (2000). The Referential Status of Clefts. *Language* 76(4), 891–920.
- Heim, I. R. (1982). *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis, University of Massachusetts.
- Heim, I. R. (1994). Interrogative semantics and Karttunen's semantics for know. In *Proceedings of the Ninth Annual Conference and the Workshop on Discourse of the Israel Association for Theoretical Linguistics*, Volume 1, pp. 128–144.
- Horn, L. (1972). *On the semantic properties of logical operators in English*. PhD thesis, University of California.
- Horn, L. (1981). Exhaustiveness and the semantics of clefts. In *Proceedings of NELS*, Volume 11, pp. 125–142.
- Horn, L. (1985). Metalinguistic Negation and Pragmatic Ambiguity. *Language* 61(1), 121–174.
- Horn, L. (2016). Information Structure and the Landscape of (Non-)at-issue Meaning. In C. Féry and S. Ishihara (Eds.), *The Oxford Handbook of Information Structure*, pp. 108–127. Oxford University Press.
- Horvath, J. (2017). Pied-Piping. In M. Everaert and H. van Riemsdijk (Eds.), *The Wiley Blackwell Companion to Syntax* (Second ed.), pp. 1–64. Wiley-Blackwell.
- Jespersen, O. (1949). *A modern English grammar on historical principles. Part VII*. London: Allen and Unwin.
- Kamali, B. and M. Krifka (2020). Focus and Contrastive Topic in Questions and Answers, with particular reference to Turkish. *Theoretical Linguistics* 46(1-2), 1–71.

- Karttunen, L. (1977). Syntax and Semantics of Questions. *Linguistics and Philosophy* 1(1), 3–44.
- Klein, W. and A. von Stechow (1982). Intonation und Bedeutung von Fokus. Arbeitspapier des SFB 99. Universität Konstanz.
- Klinedinst, N. and D. Rothschild (2011). Exhaustivity in questions with non-factives. *Semantics and Pragmatics* 4(2), 1–23.
- Kratzer, A. (2020). Situations in Natural Language Semantics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020 ed.). Metaphysics Research Lab, Stanford University. Available on <https://plato.stanford.edu/archives/fall2020/entries/situations-semantics/>, accessed 01/14/2021.
- Krifka, M. (1992). *A Compositional Semantics for Multiple Focus Constructions*, pp. 17–53. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Krifka, M. (2001a). For a structured account of questions and answers. In C. Féry and W. Sternefeld (Eds.), *Audiatur Vox Sapientia. A Festschrift for Arnim von Stechow*, pp. 287–319. Berlin: Akademie Verlag.
- Krifka, M. (2001b). Quantifying into question acts. *Natural Language Semantics* 9(1), 1–40.
- Krifka, M. (2008). Basic Notions of Information Structure. *Acta Linguistica Hungarica* 55(3–4), 243–276.
- Krifka, M. (2011). Questions. In K. von Heusinger, C. Maienborn, and P. Portner (Eds.), *Semantics. An International Handbook of Natural Language Meaning*, Volume 2, Chapter 66, pp. 1742–1784. Berlin: De Gruyter.
- Krifka, M. (2015). Bias in Commitment Space Semantics: Declarative questions, negated questions, and question tags. In *Proceedings of SALT*, Volume 25, pp. 328–345.

- Krifka, M. and R. Musan (2012). Information structure: Overview and linguistic issues. In M. Krifka and R. Musan (Eds.), *The Expression of Information Structure*, Volume 5, pp. 1–43. Berlin: De Gruyter.
- van Kuppevelt, J. (1991). *Topic en Comment: Expliciete en impliciete vraagstelling in discourse*. PhD thesis, Radboud University Nijmegen.
- van Kuppevelt, J. (1995). Discourse Structure, Topicality and Questioning. *Journal of Linguistics* 31(1), 109–147.
- van Kuppevelt, J. (1996). Inferring from topics: Scalar implicatures as topic-dependent inferences. *Linguistics and philosophy* 19(4), 393–443.
- Lambrecht, K. (2001). A framework for the analysis of cleft constructions. *Linguistics* 39(3), 463–516.
- Lascarides, A. and N. Asher (1993). Temporal Interpretation, Discourse Relations and Commonsense Entailment. *Linguistics and Philosophy* 16(5), 437–493.
- Link, G. (1983). The Logical Analysis of Plurals and Mass Terms: A Lattice-Theoretic Approach. In P. Portner and B. H. Partee (Eds.), *Formal Semantics – the Essential Readings*, pp. 127–147. Blackwell.
- Mikkelsen, L. (2007). On so-called truncated clefts. In L. Geist and B. Rothstein (Eds.), *Kopulaverben und Kopulasätze: Intersprachliche und intrasprachliche Aspekte*, Volume 512 of *Linguistische Arbeiten*, pp. 47–68. De Gruyter.
- Mitterer, H. and T. Stivers (2007). *Annual Report 2006*. Nijmegen: Max-Planck-Institute for Psycholinguistics.
- Onea, E. (2013). *Potential Questions in Discourse and Grammar*. Habilitationsschrift, Georg-August-Universität Göttingen.
- Onea, E. (2016). *Potential Questions at the Semantics-Pragmatics Interface*, Volume 33 of *Current Research in the Semantics/Pragmatics Interface*. Leiden, Boston: Brill.

Onea, E. (2019a). Exhaustivity in *it*-clefts. In C. Cummins and N. Katsos (Eds.), *The Oxford Handbook of Experimental Semantics and Pragmatics*, pp. 401–417. Oxford: Oxford University Press.

Onea, E. (2019b). Underneath Rhetorical Relations. The Case of Result. In K. von Heusinger, E. Onea, and M. Zimmermann (Eds.), *Questions in Discourse*, Chapter 7, pp. 194–250. Leiden, Boston: Brill.

Onea, E. (2020). An Inquisitive Stroll in Commitment Spaces. *Theoretical Linguistics* 46(1–2), 103–111.

Onea, E. and D. I. Beaver (2009). Hungarian focus is not exhausted. In *Proceedings of SALT*, Volume 19, pp. 342–359.

Onea, E. and M. Zimmermann (2019). Questions in Discourse: an Overview. In K. von Heusinger, E. Onea, and M. Zimmermann (Eds.), *Questions in Discourse*, Chapter 2, pp. 5–117. Leiden, Boston: Brill.

Paul, H. (1880). *Prinzipien der Sprachgeschichte*. Leipzig: Niemeyer.

Pavlović, A.-C. (2019). *The Interpretation of it-Clefts*. PhD thesis, Georg-August-Universität Göttingen.

Percus, O. (1997). Prying Open the Cleft. In *Proceedings of NELS*, Volume 27, pp. 337–352. University of Massachusetts.

Polanyi, L. (1985). A Theory of Discourse Structure and Discourse Coherence. In P. D. Kroeber, W. H. Eilfort, and K. L. Peterson (Eds.), *Papers from the General Session at the 21st Regional Meeting of the Chicago Linguistics Society*, Volume 21, pp. 306–322.

Polanyi, L. (1988). A formal model of the structure of discourse. *Journal of Pragmatics* 12(5-6), 601–638.

Pollard, C. and M. Yasavul (2014). Anaphoric *it*-clefts: The myth of exhaustivity. In *Proceedings of the Chicago Linguistic Society (CLS)*, Volume 50, pp. 381–393.

- Prince, E. F. (1978). A comparison of *wh*-clefts and *it*-clefts in discourse. *Language* 54(4), 883–906.
- Reeve, M. (2011). The syntactic structure of English clefts. *Lingua* 121(2), 142–171.
- Reichle, R. V. (2014). Cleft type and focus structure processing in French. *Language, Cognition and Neuroscience* 29(1), 107–124.
- Repp, S. (2010). Defining ‘contrast’ as an information-structural notion in grammar. *Lingua* 120(6), 1333–1345.
- Roberts, C. (1996). Information Structure: Towards an integrated formal theory of pragmatics. In J. H. Yoon and A. Kathol (Eds.), *The Ohio State University Working Papers in Linguistics*, Volume 49, pp. 91–136.
- Roberts, C. (2011). Topics. In C. Maienborn, K. von Heusinger, and P. Portner (Eds.), *Semantics. An International Handbook of Natural Language Meaning*, Volume 33:2, Chapter 72, pp. 1908–1934. Berlin, Boston: De Gruyter.
- Roberts, C. (2012). Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics. *Semantics and Pragmatics* 5(6), 1–69.
- Roelofsen, F. and S. van Gool (2010). Disjunctive Questions, Intonation, and Highlighting. In M. Aloni, H. Bastiaanse, T. de Jager, and K. Schulz (Eds.), *Logic, Language and Meaning*, pp. 384–394. Berlin, Heidelberg: Springer.
- Roland, D., F. Dick, and J. L. Elman (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language* 57(3), 348–379.
- Rooth, M. (1985). *Association with Focus*. PhD thesis, GLSA, University of Massachusetts.
- Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics* 1, 75–116.

- Rooth, M. (2009). Second occurrence focus and Relativized Stress F. In M. Zimmermann and C. Féry (Eds.), *Information Structure: Theoretical, Typological, and Experimental Perspectives*. Oxford: Oxford University Press.
- Ross, J. R. (1967). *Constraints on variables in syntax*. Phd thesis, MIT.
- Russell, B. (1905). On Denoting. *Mind* 14(56), 479–493.
- Russell, B. (1957). Mr. Strawson on Referring. *Mind* 66(263), 385–389.
- Schwarzschild, R. (1999). Givenness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics* 7(2), 141–177.
- Selkirk, E. (1996). Sentence prosody: Intonation, stress, and phrasing. In J. A. Goldsmith (Ed.), *The Handbook of Phonological Theory*, Volume 1, pp. 550–569. Cambridge: Blackwell.
- Seuren, P. A. M. (1985). *Discourse Semantics*. Oxford, New York: Basil Blackwell.
- Sheil, C. (2016). *Scottish Gaelic Clefts: Syntax, Semantics and Pragmatics*. PhD thesis, UC Berkeley.
- Simons, M., D. I. Beaver, C. Roberts, and J. Tonhauser (2017). The Best Question: Explaining the Projection Behavior of Factives. *Discourse Processes* 54(3), 187–206.
- Skopeteas, S. and G. Fanselow (2010). Focus types and argument asymmetries: a cross-linguistic study in language production. In C. Breul and E. Göbbel (Eds.), *Contrastive information structure*, pp. 169–197. Amsterdam, Philadelphia: John Benjamins.
- Stalnaker, R. C. (1976). Propositions. In A. MacKay and D. D. Merrill (Eds.), *Issues in the Philosophy of Language*, pp. 79–91. New Haven: Yale University Press.
- Stalnaker, R. C. (1978). Assertion. In *Pragmatics*, pp. 315–332. Leiden: Brill.
- Stalnaker, R. C. (2002). Common Ground. *Linguistics and Philosophy* 25(5/6), 701–721.

- von Stechow, A. (1991). Focusing and Backgrounding Operators. *Discourse particles: descriptive and theoretical investigations on the logical, syntactic and pragmatic properties of discourse particles in German* 6, 37–84.
- Stenius, E. (1967). Mood and Language-Game. *Synthese* 17(3), 254–274.
- Strawson, P. F. (1950). On Referring. *Mind* 59(235), 320–344.
- Szabó, Z. G. (2003). Definite descriptions without uniqueness: A reply to Abbott. *Philosophical Studies* 114(3), 279–291.
- Szabolcsi, A. (1981). The Semantics of Topic-Focus Articulation. In J. A. G. Groenendijk, T. M. V. Janssen, and M. B. J. Stokhof (Eds.), *Formal methods in the study of language*, pp. 513–540. Amsterdam: Mathematisch Centrum.
- Szabolcsi, A. (1994). All quantifiers are not equal: The case of focus. *Acta Linguistica Hungarica* 42(3/4), 171–187.
- Theiler, N. (2020). *Denn* as a highlighting-sensitive particle. *Linguistics and Philosophy*, 1–40.
- Tonhauser, J. (2012). Diagnosing (not-)at-issue content. In *Proceedings of Semantics of Under-represented Languages of the Americas (SULA)*, Volume 6, pp. 239–254.
- Tönnis, S. (2018). Investigating the Distribution of Clefts in Written and Spoken German: The Role of Prosody. Talk presented at the 40th Annual DGfS Meeting. Stuttgart.
- Tönnis, S., L. M. Fricke, and A. Schreiber (2018). Methodological Considerations on Testing Argument Asymmetry in German Cleft Sentences. In E. Fuß, M. Konopka, B. Trawiński, and U. H. Waßner (Eds.), *Grammar and Corpora 2016*, pp. 231–240. Heidelberg University Publishing.
- Tuller, L. (1992). The syntax of postverbal focus constructions in Chadic. *Natural Language & Linguistic Theory* 10(2), 303–334.
- Uegaki, W. (2015). *Interpreting Questions under Attitudes*. PhD thesis, MIT.

van der Sandt, R. A. and B. Geurts (1991). Presupposition, Anaphora, and Lexical Content. In O. Herzog and C. Rollinger (Eds.), *Text understanding in LILOG. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence)*, Volume 546, pp. 259–296. Berlin, Heidelberg: Springer.

Velleman, D., D. I. Beaver, E. Destruel, D. Bumford, E. Onea, and L. Coppock (2012). *It-clefts are IT* (inquiry terminating) constructions. In *Proceedings of SALT*, Volume 22, pp. 441–460.

Wedgwood, D., G. Pethő, and R. Cann (2006). Hungarian ‘focus position’ and English *it*-clefts: the semantic underspecification of ‘focus’ readings.

Westera, M. and H. Rohde (2019). Asking between the lines: Elicitation of evoked questions in text. In *Proceedings of the 22nd Amsterdam Colloquium*, pp. 397–406.

Whorf, B. L. (1938). Some Verbal Categories of Hopi. *Language* 14(4), 275–286.

Zimmermann, M. (2007). Quantifying question particles in German: Syntactic effects on interpretation. In *Proceedings of Sinn und Bedeutung*, Volume 11, pp. 627–641.

Zimmermann, M. (2011a). Discourse Particles. In *Semantics. An International Handbook of Natural Language Meaning*, Volume 2, Chapter 76, pp. 2011–2038. Berlin: De Gruyter.

Zimmermann, M. (2011b). The grammatical expression of focus in West Chadic: Variation and uniformity in and across languages. *Linguistics* 49(5), 1163–1213.