

Eberhard Karls Universität Tübingen

Seminar für Sprachwissenschaft

**Kommentiert [MS1]:** Am Ende darauf achten: 1. Einzug nach jedem Absatz gleich breit? 2. Nummerierung der Beispiele im Text korrekt? 3. Jahreszahlen im Text mit Jahreszahlen in Zotero (und entsprechend den References) abgleichen 4. entweder überall we oder I/us oder me 5. verweise auf andere subchapters korrekt? 6. [check if I included all papers that James sent to me in May(?)] 7. acceptable/natural in text (perceived as more natural instead of "acceptable" because I changed it in the study)

**Acceptability Judgments About  
Contrastive Dialogues Involving Ellipsis:  
A Pilot Study**

Thesis submitted for the degree of Master of Arts

Supervised by

Prof. Dr. Michael Franke and

Jun.-Prof. Dr. James Griffiths

Summer Term 2023

Miriam Schiele

miriam.schiele@student.uni-tuebingen.de

Matriculation Number: 6098290

Käthe-Kollwitz-Straße 15

72074 Tübingen

Date of Submission: September 23<sup>rd</sup>, 2023

**Kommentiert [MS2]:** Mündliche Prüfung (Präsentation der MA) muss vor dem 30.09. stattfinden. Vermutlich geht es auch, wenn die Abgabe am 29. ist und die Präsentation am 30., aber sicherheitshalber (Krankheit, etc) wäre der 23.09. als Abgabetermin sinnvoller

## Table of contents

1. Introduction .....	6
1.1 Background and motivation .....	6
1.2 Research questions and objectives .....	7
1.3 Significance of the study .....	8
1.4 Scope and limitations .....	10
2. Literary review .....	11
2.1 Fragment theory and its linguistic foundations .....	11
2.2 Contrastive focus .....	12
2.3 The comprehension of fragmentary answers .....	14
2.4 Acceptability judgment tasks (AJTs) in linguistics .....	17
3. Data and method .....	20
3.1 Study design .....	20
3.2 Selection of critical and filler items .....	22
3.3 Recording of stimuli .....	26
3.4 Data collection .....	28
3.5 Participant recruitment and characteristics .....	28
4. Results and analysis .....	29
5. Discussion .....	37
5.1 Interpretation of the findings .....	37
5.2 Comparison with previous studies and theoretical predictions .....	39
6. Conclusions .....	39
6.1 Summary of findings .....	40
6.2 Contributions to the field .....	40
6.3 Limitations of the study .....	40
6.4 Suggestions for future research .....	41
7. References .....	43
8. Appendix .....	44
8.1 Abbreviations, symbols and other notational conventions .....	44
8.2 List of critical items .....	45
8.3 List of filler items .....	46

## Table of figures

**Kommentiert [MS3]:** Aktualisieren, Blocksatz, Zeilenabstand 1,5, hängender Einzug mit 1cm Abstand

Figure 1: Pitch contour of stimuli with emphasis .....	26
Figure 2: Pitch contour of stimuli without emphasis.....	27
Figure 3: Scatter plot of participants' ratings of all critical items .....	30
Figure 4: Mosaicplots of participants' ratings of each factor .....	31
Figure 5: Boxplot of participants' ratings of all critical items .....	32
Figure 6: Comparison of participants' ratings of stimuli with and without emphasis.....	34
Figure 7: Comparison of participants' ratings of auditory and written stimuli.....	35
Figure 8: Comparison of participants' ratings of stimuli with functional and lexical fragments. ....	36

## Abstract

This paper focuses on the factors that influence how native speaker perceive fragments in dialogues involving contrastive focus in German. A total of [number] participants were asked to rate fragmentary answers on a 7-point Likert scale.

A total of three variable were investigated: modality, emphasis, and fragment type. That it, it was investigated whether contrastive fragmentary answers are perceived more natural if they are presented as written or auditory stimuli, if the contrastive words were emphasized, and if the contrastive word in the fragmentary answer had lexical or functional meaning.

Insert hypotheses 1-3

The results suggest that...

These findings give insights into...

Kommentiert [MS4]: Insert number

## 1. Introduction

In this chapter, the background and motivation for the present study is provided, outlining the research questions and objectives that guide our investigation. I emphasize the significance of this study, while acknowledging the scope and limitations inherent in our research design.

### 1.1 Background and motivation

The pronunciation of a sentence holds significance, encompassing not only the stressed words but also the contents emphasized in the conversation. However, it is reasonable to assume that the perception of naturalness by native speakers may not solely depend on the stressed word but also on its associated meaning as well as the structure of the sentence. This becomes particularly intriguing in dialogues that involve contrastive focus and incomplete sentences as answers, as illustrated in the following example in German (1).

- (1) A: Peter hat AB 18 Uhr im Kino gearbeitet.  
'Peter worked at the cinema FROM 6pm.'  
B: Nein, BIS 18 Uhr.  
'No, UNTIL 6pm.'  
(own example)
- (2) A: Peter hat ab 18 Uhr im Kino gearbeitet.  
'Peter worked at the cinema from 6pm.'  
B: Nein, bis 18 Uhr.  
'No, until 6pm.'  
(own example)

Speaker B's response in (1) lacks a complete sentence structure. Therefore, to understand the intended meaning of speaker B's response, the reader must construct a complete sentence using both speaker A's preceding utterance and speaker B's response. Hence, the reader can derive the following sentence: *Peter worked at the cinema until 6pm.*

However, for the reader to grasp its intended message, they need to first identify the word *bis* 'until' as contrasting with something else and then discern the intended meaning behind speaker B's response. If the reader is not primed for the contrast through orthographic marking, understanding speaker B's response could become even more challenging. This becomes

particularly evident in dialogues that do not include any orthographic marking, as demonstrated in example (2).

In the dialogue (2), the contrasting words are not emphasized and therefore, the reader is confronted with the contrast in speaker B's answer unexpectedly. It is yet to determine to what extent native speakers encounter difficulties in interpreting dialogues lacking emphasis on the contrastive words such as (2) in comparison to dialogues such as (1) and what other factors are at play. Therefore, the modality and fragment type that might influence the acceptability ratings of dialogue involving contrastive focus and fragmentary answers are analyzed.

The present paper is subdivided as follows. Chapter 2 delves into the theoretical background of fragments, focus, and ellipsis comprehension, providing a more detailed explanation of our hypotheses. Chapter 3 centers on the study design and participant information. Chapter 4 presents the findings of the study, while chapter 5 explores and addresses any confounding factors related to the findings. Lastly, chapter 6 concludes with a summary of the study and offers insights into potential avenues for future research.

Kommentiert [MS5]: Update once MA is done

## 1.2 Research questions and objectives

The present paper aims to determine the most effective medium for reliably obtaining judgments about such dialogues involving contrastive focus and fragmentary answers, pathing the way for future research on fragments using acceptability judgement tasks. By examining the effects of different stimulus characteristics, we seek to deepen our understanding of how modality, emphasis, and fragment type contribute to the perceived acceptability of fragments. This section will present the factors and hypotheses investigated in the present paper.

First, as has been shown by the examples (1) and (2), dialogues that incorporate orthographic marked contrasting words are contrasted with dialogues that lack any emphasis. The present paper aims to investigate which dialogues are perceived as more natural by native speakers. It is hypothesized that dialogues emphasizing the contrasting words are regarded as more natural by native speakers. This prediction is grounded in the assumption that emphasizing the contrasting elements enhances their salience and

facilitates comprehension, leading to increased acceptability. For an overview of the role of emphasis in sentence comprehension, see chapter 2.3.

Furthermore, a comparison is made between (1) and (2) on the one hand and their verbal counterparts on the other hand. That is, the sentences (1) and (2) are recorded by native speakers and differ in whether they display default intonation or whether they prosodically stress the contrasting words. The presentation of auditory stimuli is expected to establish an authentic and natural context for fragmentary answers, leading to heightened acceptability ratings compared to written stimuli. Hence, it is hypothesized that generally, auditory stimuli are more likely to be accepted by native speaker than written stimuli.

Last, dialogues such as (1) and (2) are compared to dialogues, in which the contrastive words do not have functional meaning but lexical meaning. That is, instead of prepositions such as *bis* 'until' and *ab* 'from', the contrastive focus is placed on nouns such as *Bruder* 'brother' and *Vater* 'father' as in (3). See also (4) as the equivalent of (3) without the inclusion of emphasis.

(3) A: Peter hat seinem BRUDER ein Buch geschenkt.  
'Pete gave a book to his BROTHER.'

B: Nein, seinem VATER.  
'No, his FATHER.'

(own example)

(4) A: Peter hat seinem Bruder ein buch geschenkt.  
'Pete gave a book to his brother.'

B: Nein, seinem Vater.  
'No, his father.'

(own example)

Thus, it is hypothesized that stimuli with lexical words in contrastive focus such as *Bruder* and *Vater* receive higher acceptability ratings than stimuli with functional words in contrastive focus such as *ab* and *bis*.

### 1.3 Significance of the study

In the following section, the significance of the present study is demonstrated by exploring the implications of the findings for theoretical frameworks and practical applications.

**Kommentiert [MS6]:** Alles noch überprüfen, ob wirklich keine Studien dazu gemacht wurden und nochmal umformulieren

The present study examines a number of factors that have not been investigated in depth or at all in previous studies. First, while previous research has studied the importance of orthographic marking in other fields of linguistics, the role of orthography in the comprehension of contrastive focus has not been studied so far. The studies conducted in this field are reviewed in chapter 2.3.

Second, there exists a research gap concerning the significance of emphasis in comprehending fragmentary answers. While numerous studies have explored prosody and ellipsis comprehension, limited information is available regarding the impact of pitch accent placement on the perceived naturalness of fragmentary responses. That is, it is yet to be determined whether the pitch accent must be placed on the correlate of the fragmentary answer or whether a natural, default intonation of the preceding utterance is sufficient to comprehend the fragmentary answer. Moreover, the studies that have been conducted have focused on English, neglecting the possibility of crosslinguistic differences regarding where the default sentence accent falls. Recent findings regarding these research questions are discussed in chapter 2.4.

Next, previous studies have focused solely on lexical contrastive answers, i.e., mostly proper names. Based on this data, theories on how such structures are processed were built. However, it very well might be that there are differences between processing proper names and processing other words that either do not denote human referent or do not have lexical meaning at all. The processing theories must account for such differences.

Last, the ongoing debate about formal and informal methods of conducting acceptability judgements has emphasized the need for more reliable data obtained by scientific standards. The present study will provide such formally obtained data.

Therefore, this study will be the first of its kind to systematically examine and compare different media for collecting acceptability judgments about fragments. By determining the best medium for reliably obtaining these judgments, our research will contribute to a deeper understanding of the nature of fragments and provide valuable insights for future studies in linguistics and psycholinguistics.



#### 1.4 Scope and limitations

In the following, the limitations inherent in our research design, acknowledging the potential constraints and scope of the investigation are addressed.

The investigation focuses exclusively on clausal ellipsis, disregarding semantic and pragmatic ellipsis. This decision was made to ensure that our analysis remains well-defined and manageable within the given scope of the study. Semantic and pragmatic ellipsis could be potential avenues for future research, but they fall outside the boundaries of the present investigation.

Furthermore, the study is limited to exploring syntactic ellipsis solely within the clausal context. It is acknowledged that ellipsis can occur at various linguistic levels and that analyses of these phenomena offer valuable insights, they lie beyond the current investigation's scope. Therefore, for the sake of depth and coherence, it was opted to concentrate on clausal ellipsis only. Future studies could explore other types of ellipsis to gain a more comprehensive understanding of ellipsis phenomena.

Additionally, the study does not extend its analysis to dialectal variations, differences between age groups, genders, or any other participant-specific features. Previous research indicated that except for age and geography, such sociolinguistic features do not significantly influence the acceptability ratings (cf. Delbar 2019). It is because of this that age and geography are controlled for in the present study.

Within its defined scope, the research design utilizes acceptability judgment tasks. However, this method inherently presents limitations. Participant responses may be influenced by individual linguistic competence, biases, and subjective interpretations, introducing potential sources of uncertainty. Although measures to mitigate these issues were taken, such as

ensuring a diverse participant pool, providing example dialogues in the introduction to the study, and utilizing statistical analysis, it is essential to recognize these inherent limitations.

In conclusion, the present study is bounded by specific limitations and a carefully defined scope. By recognizing these limitations, the present study ensures the reliability of its findings and identifies potential avenues for future research.

## 2. Literary review

Whether fragments are derived by ellipsis or not is still under debate, making it necessary to understand the theories of ellipsis before delving into the theory of fragments. Consequently, the chapter begins by introducing the two primary paradigms of ellipsis theories: the deletion and nondeletion accounts. Following that, the fragment theory is explained, encompassing its linguistic foundations. Moreover, I discuss the methodological approach of acceptability judgment tasks, which allow us to capture native speakers' intuitions about the grammaticality and acceptability of linguistic constructions. Furthermore, previous the role of emphasis in sentence comprehension is discussed and research on written and auditory stimuli as well as functional and lexical stimuli is reviewed to establish a strong foundation for our own empirical analysis and contribute to the broader understanding of these linguistic phenomena.

**Kommentiert [MS7]:** Nochmal überprüfen, ob Reihenfolge stimmt

### 2.1 Fragment theory and its linguistic foundations

The present paper focuses on a specific type of ellipsis, i.e., fragments. For a first understanding of fragments, consider (9).

- (5) Abby and Ben are at a party. Abby asks Ben about who their mutual friend Beth is bringing as a date by uttering: "Who is Beth bringing?"  
Ben answers:  
"Alex."

(Merchant 2004, p. 661)

Ben's answer *Alex* in (9) only consists of one word and yet, the reader can easily be derived that it is intended to convey that Beth is bringing Alex.

Such short answers are called fragments (cf. Merchant 2004). [include better definition of fragment]

According to the deletion approach as proposed by Merchant (2001), fragments are the only pronounced item in a full-fledged yet unpronounced clause. Merchant's theory is called move-and-delete approach (MDA) because it assumes that fragments move to the clause-peripheral position and that the concomitant ellipsis operation deletes the remaining constituents (cf. Merchant 2004). That is, Ben's answer *Alex* is the only pronounced constituent of its underlying clause *Beth is bringing Alex*.

**Kommentiert [MS8]:** Nochmal umformulieren mit Merchant 2001

That is, while clausal ellipsis suppresses the phonological realization of most parts of the sentence, one or more subconstituents of the clause survive ellipsis. These remnants of ellipsis are called fragments.

**Kommentiert [MS9]:** A better way to phrase it would be:  
Remnants of ellipsis (henceforth, fragments) ....

## 2.2 Contrastive focus

Before examining contrastive fragments, the notion of *contrastive focus* is explained and illustrated with examples. Subsequently, the chapter directs its attention specifically towards fragments that incorporate contrastive focus.

**Kommentiert [MS10]:** Probably write again after reading Merchant 2001 and making sure it's not a direct copy of Griffith et al 2023

The notion of *focus* is commonly associated with the element of an utterances that adds new information to the discourse (cf. Lambrecht 1994). A more elaborate definition would exceed the scope of this investigation and is not essential for comprehending the definition of contrastive focus. Contrastive focus is a specific type of focus and can be defined as follows.

- (6) Contrastive focus represents a subset of contextually or situationally "given" alternative elements for which the predicate phrase can potentially hold, and spells out this subset as the one for which the predicate actually hold.

(Griffiths & Lipták 2014, p. 200, quotation marks in original)

However, the alternative denotations must be of the same type and mutually exclusive (cf. Wagner 2012, Krifka 2008). An example of contrastive focus is illustrated in (12), where the brackets subscripted with *F* indicate the constituent in focus.

- (7) A: Mary stole the cookie.  
B: No, [Peter]<sub>F</sub> stole the cookie!  
(adapted from Krifka 2008, p. 252)

The sentence in (12A) is called antecedent clause, while speaker B's answer is named a contrastive utterance, since *Mary* and *Peter* are contrasting. Moreover, contrastive focus is placed on *Peter* in accordance with the definition in (11). As can be seen in (12B), the contrastive focus placed on *Peter* serves to emphasize *[[Peter]]* as an "alternative answer to an explicit or implicit statement provided by the previous discourse/situation" (Wagner 1999, p. 1529).

After examining (12B), which represents a full correction, now consider the fragmentary equivalent (13B).

- (8) A: Mary stole the cookie.  
B: No, [Peter]<sub>F</sub>!  
(adapted from Krifka 2008, p. 252)

The full correction in (12B) and the elliptical correction in (13B) have the same semantic meaning and pragmatic function, despite of the fact that (13B) only consists of a fragment. This is because fragmentary answers are assumed to be structurally identical to full sentences, as discussed in chapter 2.1, and contrastive focus is placed on *Peter* in both (12B) and (13B).

Contrastive fragments are a specific type of fragments. Contrary to other fragments, contrastive fragments include "an explicit relation of contrast between the elliptical remnant and its correlate in the antecedent clause" (Griffiths & Lipták 2014, p. 199, emphasis omitted). In the dialogue in (13), this is illustrated by the contrast between *Peter* and *Mary*, as these words denote contrasting referents. Therefore, contrastive fragments can be used for corrections and always include contrastive focus (cf. Krifka 2008, Griffiths & Lipták 2014).

Contrastive focus has often been characterized phonologically with a L+H\* pitch for English and German (cf. Wagner 1999)<sup>1</sup>. The importance of

---

<sup>1</sup> For the description of pitch accent and other intonational patterns, the ToBI labeling by Beckman & Ayers (1997) will be used.

not only pitch accents on elements in contrastive focus but the role emphasis in the comprehension of fragmentary answers in general is explained in the following subchapter.

### 2.3 The comprehension of fragmentary answers

This section explores the processes behind the comprehension of fragmentary answers. First, the process of the comprehension of elliptical utterances is explained. Then, I discuss to what extent emphasis and at-issueness play a role in said process.

Comprehending any utterance requires the evaluation and processing of information, i.e., organize lexical meaning and understanding syntactic structures (cf. Harris & Carlson 2018). This is particularly difficult for elliptical utterances such as dialogues involving fragmentary answers, since the syntactic structure is not present and has to be constructed by the processor, i.e., hearer or reader of the dialogue (cf. Phillips & Parker 2014). For comprehending elliptical utterances, the processor must finish the following three basic tasks.

(9) Basic tasks of the processor in ellipsis processing:

1. Parse the remnant by constructing the appropriate phrase structure for the remnant given the input.
2. Locate the correlate, if any, from the antecedent clause.
3. Construct the elided phrase by regenerating or copying a structure at Logical Form

(Harris & Carlson 2018, p. 485)

In the example (13B), that means that first, *Peter* is identified as remnant. Next, *Mary* is retrieved as correlate to *Peter*, as it appears to be a suitable contrasting denotation. Lastly, the elided phrase is contrasted, i.e., *Peter<sub>1</sub> stole the cookie t<sub>1</sub>* (cf. Harris & Carlson 2018).

While the tasks in (14) mainly focus on the syntactical structure, the meaning of the elliptical utterance can be derived from the background as formulated in (15) for clausal ellipsis, where *Question Under Discussion* (*QUD*) refers to a “partially-ordered set that specifies the currently-discussable questions at any point during a conversation” (Griffiths 2019, p. 8) and the *maximal QUD* (*MaxQUD*) refers to the “most conventionally-prominent

discussable questions (or unordered subset of question) in this set" (ibid, brackets in original).

- (10) Given a question  $q$  in the MaxQUD with background  $Q$  and clause  $\alpha$  with background  $A$ , clausal ellipsis is recoverable in  $\alpha$  iff  $A \sqsubseteq Q$ .

(Griffiths 2019, p. 10)

[maybe discuss the MaxQUD a bit here before moving on to prosody]

Kommentiert [MS11]: See note in text

So far, this chapter has primarily addressed the understanding of elliptical utterances, without considering the potential impact on comprehension caused by emphasizing specific words. Therefore, in the following, the role of orthographic and prosodic marking is discussed.

Rasekhi & Vahideh investigate to what extent information structure, semantic parallelism, and locality facilitate the comprehension of elliptical clauses.

→ Include Rasekhi & Harris 2021 in more detail

Previous research has studied the influence of how words are presented such as capitalization or color highlighting on comprehension. By employing different forms of orthographic marking, these studies have investigated, e.g., grammatical encoding of subject-verb agreement (cf. Franck et al. 2003), lexical access (cf. Opitz & Bordag 2022), and second language acquisition (cf. Meurers et al. 2010). However, the role of orthographic marking on the comprehension of ellipsis has not been studied so far.

Prosody is known for impacting language processing (cf. Warren 1999). However, the influence of intonation and emphasis is especially intriguing for the processing of ellipsis sentences and for structures involving

contrastive focus. As mentioned in the chapter 2.2, Contrastive focus has been characterized phonologically with a L+H\* pitch for English and German (cf. Wagner 1999). This leads to the hypothesis that prosodically marking contrasting words may enable the listener to access these words swiftly and effortlessly, ultimately facilitating a clearer comprehension of the contrast.

**Kommentiert [MS12]:** Include explanation for high and low pitch (see Harris & Carlson)

Carlson et al. (2009) investigated whether pitch accent affects how ambiguous replacive sentences such as (7) are interpreted.

- (11) a. ROGER insisted that Alice was reliable // not ANDREW[.]  
b. Roger insisted that ALICE was reliable // not ANDREW[.]  
c. ROGER insisted that ALICE was reliable // not ANDREW.  
(Carlson et al. 2009, p. 1077)

**Kommentiert [MS13]:** Nochmal ausführlicher beschreiben, dass es die erste Hypothese in der Studie ist und auch die zweite Hypothese (auditory stimuli > written stimuli) und die dritte (lexical > functional) erklären in anderen Subkapiteln

Indeed, they found that pitch accent significantly influenced participants' choice of the correlate of the replacive, i.e., whether the expression *Andrew* is used to replace *Roger* or *Alice*. There are two main differences between the sentences used in the present study by Carlson et al. (2009) and the sentences used in the present study. First, the former uses replacive sentences, while the latter uses fragmentary answers. Since both are forms of ellipsis, it is still reasonable to hypothesize significant differences comparing dialogues with and without emphasis on the contrasting words. Second, the former uses ambiguous sentences, while the latter uses sentences disambiguated using case marking. Although the prosodic marking is not used to find the correct correlate, one can hypothesize that emphasizing the contrasting words amplifies their prominence in the discourse, facilitating the comprehension of the contrast in the fragmentary answer.

However, it is not only the emphasis that impacts the understanding of elliptical sentences. It can be inferred that fragmentary answers are more common in spoken conversations compared to written dialogues due to the disparity between written and spoken language in terms of prioritizing complete expressions (cf. Akinaso 1982). Given that fragmentary answers are more frequent in spoken language than in written language, one can assume that these structures will be perceived as more acceptable, if they

occur in contexts that they are used in more often. Therefore, it is hypothesized that generally, auditory stimuli will receive higher acceptability ratings than written stimuli.

- Include research/1-2 paper on lexical and functional words (not necessarily fragments) and formulate and explain third hypothesis OR issueness??
- Include paragraph on at-issueness when talking about lexical and functional fragments?

#### 2.4 Acceptability judgment tasks (AJTs) in linguistics

This section provides an overview of acceptability judgment tasks, henceforth AJTs, and ... informal and formal methods and explains why acceptability judgments are a reliable source of?

More than 60 years ago, acceptability judgments were initially suggested as a substitute for assessing grammaticalness of syntactic theories.

Chomsky (1957) proposed that “[o]ne way to test the adequacy of a grammar proposed for [a language] is to determine whether or not the sequences that it generates are actually grammatical, i.e., acceptable to a native speaker” (p. 13).

Although AJTs are conducted to investigate the syntactic structures of a language, the term *grammaticality judgment* is misleading, as it assumes that participants have the ability to access their implicit knowledge about language. Instead, the term *AJT* is used to clarify that based on the acceptability of certain structures, one can gain insights about the grammaticality of the respective syntactic patterns (cf. Sprouse et al. 2013).

In the following, the term *informal method* refers to AJTs that were conducted with a low number of participants and that are associated with scale biases, judgment errors, etc. In contrast, the term *formal method* is used to describe AJTs, adhering to experimental standards, i.e., a common rating scale, a sufficient number of participants, etc. (cf. Juzek 2016). As the present study uses a 7-point Likert scale, the comparison of different AJTs

**Kommentiert [MS14]:** Include: (adopted from James) These are introspective judgments provided by native speakers about the acceptability of a particular linguistic expression in a prescribed context. In this project, we adopt the neutral position that introspective judgments reflect one's unconscious knowledge of how their native language can be used within and across sentences. Putting aside cases of language attrition,[63] we also assume that an adult's knowledge of their native language is complete.[64] To exhaustively test the predictions in (21) and (22) from §1.2, we need to know which fragments can and cannot be expressed in a particular language. Only judgment tasks can provide us with negative data, as other data sources, such as printed texts or electronic corpora of naturally-occurring speech, are composed solely of permissible linguistic expressions (i.e., positive data) or disfluencies. Also, the few fragments that appear in corpora show insufficient grammatical variation for our purposes. For instance, [65]'s study of 14,000 utterances from the British National Corpus of English returned only 1,300 fragments, 50% were plain acknowledgements, i.e., “yes” or “no”

**Kommentiert [MS15]:** No pdf found online, Cite in Zotero according to details on Uni Tü Bib katalog

**Kommentiert [MS16R15]:** Chomsky, Noam. 1957. Syntactic Structures. The Hague, The Netherlands: Mouton de Gruyter.

**Kommentiert [MS17R15]:** <https://rds-tue.ibs-bw.de/opac/RDSIndex/Se-arch?join=AND&bool0%5B%5D=AND&lookfor0%5B%5D=chomsky&lookfor0%5B%5D=Syntactic+Structures&type0%5B%5D=au&type0%5B%5D=ti&page=2>



only considers informal methods on the one hand and Likert scales as formal method on the other hand, leaving out other formal methods such as magnitude estimation, two-alternative forced-choice, etc. For a description of each method and an in-depth analysis of what methods are appropriate to conduct acceptability judgment, the reader is referred to Sprouse et al. (2013). The study opted for a multi-point scale, as it allows for statistical analysis of the judgments, including calculating sample means, standard deviations, and assessing the significance of the effects under investigation (cf. Featherston 2008).

The 7-point Likert scale used in the present study asked participants to rate each dialogue presented in the experiment on a scale from 1 to 7, representing varying degrees of acceptability. That is, in the present experiment, 1 was labeled as *fully unacceptable* and 7 as *fully acceptable*. A 7-point scale was chosen because it allows participants to identify structures that are neither acceptable nor unacceptable (cf. Sprouse et al. 2013).

The study was designed in a way that prevented participants from revisiting previous slides and changing their ratings or skipping dialogues without providing their acceptability ratings. Additionally, participants in the study involving auditory stimuli were required to listen to the entire audio files of the dialogues before they could provide their ratings.

Although AJTs have traditionally relied on written stimuli, auditory stimuli have been adopted by AJTs as a substitution for written stimuli in recent studies (Kayali 2023, Jasso 2022, Liu et al. 2022). This is particularly beneficial when examining structures that are uncommon in written language or necessitate prosodic cues for a comprehensive understanding of the syntactic structure. Therefore, auditory stimuli are deemed suitable in such cases (cf. Sedarous & Namboodiripad 2020). For an overview of the importance of emphasis in sentence comprehension, see chapter 2.5.

Juzek (2016) investigated whether the mode of stimuli influences participants' ratings in AJTs. That is, he stated as a null hypothesis that either the ratings are the same for both written and auditory stimuli. As an alternative hypothesis, he proposed that constructions that are more common in spoken language receive higher acceptability ratings as auditory stimuli and

constructions that are more common in written language receive higher acceptability ratings as written stimuli. In his experiment, the difference between commonly used written and spoken constructions as written and auditory stimuli, respectively, lacked significance. Hence, the null hypothesis could not be rejected. However, the experiment investigated resumptive pronouns. As the present study examines dialogues involving contrastive focus and fragmentary answers, the intonation of those sentences perhaps plays a larger role than for sentences with resumptive pronouns. Therefore, it might be that a significant difference in the mode of stimuli can be found in the present study. For a justification of the present methodology, see chapter 3.1.

As a final aspect, the limitations of AJTs are addressed to provide a comprehensive overview of this methodology. First, AJTs are inherently subjective and hence, leading to a high degree of variations. However, through a sufficient number of participants, reliable results are ensured in the present study. Second, although a 7-point Likert scale is more fine-grade than binary acceptability judgment, it still may be too general to capture the full range of variation and subtleties in native speakers' intuition. Furthermore, it is worth noting that in some cases, there might be a lack of contextual information, potentially hindering the resemblance to natural language use and compromising the reliability of judgments. However, in the current experiment, considerable efforts were made to select stimuli that minimize the risk of misunderstandings. Additionally, the inclusion of seven stimuli per condition serves to further reduce this risk and enhance the reliability of the results.

### 3. Data and method

This chapter presents an overview of the study design, stimuli selection, recording procedures, data collection, participant recruitment and characteristics, and data analysis methods. I discuss the rationale, procedures, and considerations involved in each aspect.

#### 3.1 Study design

The experiment was conducted using a 2 (modality: written or auditory) x 2 (emphasis: with or without emphasis) x 2 (fragment-type: functional or lexical word) study design. Therefore, eight conditions were tested by using three binary factors. A between-subject design was employed to examine the effects of modality, while a within-subject design was utilized to investigate the influence of emphasis and fragment-type. The chosen study design aims to mitigate participant perplexity or skepticism arising from varying modalities and to ensure that any observed differences in results for stimuli with different emphasis and fragment types are attributable to their influencing factors rather than participant variability. This design selection safeguards against potential confounding factors, reducing individual differences and increasing the sensitivity to detect effects.

In the study, participants were asked to rate dialogues in the naturalness. The AJT was an ordinal response task on a 7-point Likert scale. The scale ranged from 1, representing fully acceptable structures to 7, indicating full unacceptability. Prior, fully acceptable, fully unacceptable, and neither acceptable nor unacceptable examples were given in the introductory part to the study. This methodological choice was motivated by several factors. First, the 7-point Likert scale offers an appropriate range of response options, allowing participants to express nuanced judgments effectively. It includes a balanced midpoint that signifies structures perceived as neither acceptable nor unacceptable, as discussed in chapter on 2.4. Second, the inclusion of auditory stimuli allows us to capture the full range of linguistic cues present in natural speech. Dialogues, particularly those involving contrastive focus and fragmentary answers, often rely on prosodic features, as discussed in more detail in 2.5. By presenting participants with auditory stimuli, we provide them with a more ecologically valid representation of

these linguistic cues compared to written stimuli alone. Third, including written stimuli alongside auditory ones allows us to explore potential differences or convergences in acceptability judgments between the auditory and written presentations, shedding light on the role of modality in the perception of naturalness, specifically for fragments and contrastive focus.

The experiment was designed as follows. After a welcoming page, participants were presented with three dialogues (15-17) that had a similar structure to the critical and filler items of the experiment and varying acceptability ratings. The dialogues did not include the variables that were investigated in the experiment. However, through the introduction of similar dialogues, participants became familiar with the rating scale. Moreover, it was ensured that all participants understood that the study aims to determine what sentences would be acceptable in daily speech contrary to written language (cf. Sedarous & Namboodiripad 2020). Note that the glossing and translation is given in (15-17), while participants were only presented with only the German sentences.

- (12) A: *Was mag Peter?*  
 what likes Peter  
 'What does Peter like?'  
 B: *Peter mag Ingwer.*  
 Peter likes ginger  
 'Peter likes ginger.'  
 (adapted from Sedarous & Namboodiripad 2020, p. 7)
- (13) A: *Was hat Peter gestern gemacht?*  
 what AUX Peter yesterday did  
 'What did Peter do yesterday?'  
 B: *\*Vater Fußball gestern.*  
 father football yesterday  
 'father football yesterday.'  
 (ibid.)
- (14) A: *Hat Peter inzwischen aufgegeben?*  
 AUX Peter by.now gave.up  
 'Has Peter given up by now?'  
 B: *?Nein, das Handtuch, das würde er*  
 no the towel that AUX he  
*bestimmt nie werfen!*  
 certainly never throw  
 'No, the towel, he would certainly never throw that in!'  
 (adapted from Wierzba et al. 2023)

Participants are instructed that their acceptability ratings should be based on only speaker B's response to speaker A's utterance. While (15) is described to the participants as *fully acceptable*, (16) is identified as *fully unacceptable* and (17) is used as an example of an utterance that is *neither acceptable nor unacceptable*.

After the introductory pages, participants were randomly assigned to either only written or auditory stimuli and then presented with seven items of each variable, i.e., with and without emphasis of contrasting words as well as functional and lexical fragments. In total, 56 critical items, including written and auditory items, were used for the study. However, since each participant was assigned to either written or auditory stimuli, each participants encountered 28 critical items and 28 filler items. The critical, written items were equally distributed across the four conditions, i.e., with and without orthographic marking as well as lexical or functional fragment types. Similarly, the critical, auditory items were equally distributed across the four conditions, i.e., with and without prosodic marking on the contrasting words as well as lexical or functional fragment types. Therefore, each condition is exemplified by 7 items in each run of the experiment. This balanced design allows for a systematic examination of the effects of emphasis and fragment types on the experimental variables.

At the end of the questionnaire, participants were asked to indicate their age, level of education, where they grew up, and native language/dialect. Completing the entire questionnaire took the participants about 10 minutes approximately.

**Kommentiert [MS18]:** Check if that is correct or what else they were asked to indicate

**Kommentiert [MS19]:** Did it?

### 3.2 Selection of critical and filler items

This section gives an overview of the selected critical and filler items for the study and the rationale behind their choice. The list of written critical and filler items can be found in the appendix, while their verbal equivalents can be found here: <https://shorturl.at/blwGM>.

An exemplary overview of how the written and auditory stimuli vary regarding emphasis and fragment type is shown in (1-4), repeated here as (19-22).

- (15) A: Peter hat AB 18 Uhr im Kino gearbeitet.  
'Peter worked at the cinema FROM 6pm.'  
B: Nein, BIS 18 Uhr.  
'No, UNTIL 6pm.'  
(own example)
- (16) A: Peter hat ab 18 Uhr im Kino gearbeitet.  
'Peter worked at the cinema from 6pm.'  
B: Nein, bis 18 Uhr.  
'No, until 6pm.'  
(own example)
- (17) A: Peter hat dem POLIZISTEN seinen Ausweis gezeigt.  
'Peter showed his identity card to the POLICE OFFICER.'  
B: Nein, dem TÜRSTEHER.  
'No, the BOUNCER.'  
(own example)
- (18) A: Peter hat dem Polizisten seinen Ausweis gezeigt.  
'Peter showed his identity card to the police officer.'  
B: Nein, dem Türsteher.  
'No, the bouncer.'  
(own example)

The stimuli in (19) and (20) include functional fragments, i.e., prepositions, while the stimuli in (21) and (22) incorporate lexical fragments, i.e., nouns that denote human referents. Moreover, the stimuli in (19) and (21) emphasize the contrasting words. In contrast, the stimuli in (20) and (22) do not incorporate any orthographic marking and their verbal equivalents display natural intonation of the sentence.

Several steps were taken to minimize the influence of extraneous factors. First, the contrasting words in the stimuli with functional fragment type incorporated the prepositions *bis* 'until' and *ab* 'from', *mit* 'with' and *ohne* 'without' as well as *nach* 'after' and *vor* 'before', as these demonstrate opposite meanings.

Second, for the lexical fragments and their correlates, the contrasting nouns all denoted human referents to. Moreover, only masculine nouns marked with dative case were chosen to stand in contrastive focus to ensure that the reader or hearer can unambiguously identify the correlate of the fragment.

Third, critical items were adjusted to be in past tense to ensure that the word in contrastive focus is not in final position, as this position is claimed to be a default location (cf. Harris & Carlson 2018/ Carlson et al.

**Kommentiert [MS20]:** Look up again and re-formulate. Literature: utterance-final effect (Griffiths & Lipták 2014, p. 202, footnote 10; Barros et al. 2014 )

**Kommentiert [MS21R20]:** James said that any fragment that correlates with final position is accepted

2009???) . Therefore, fragments that correlate(?) to a word or phrase in this position are more likely to be accepted than if the correlate is inside the clause (Quelle).

Kommentiert [MS22]: Look up again

Next, the sentences were created in such a way that stimuli with lexical fragments include ditransitive verbs, while stimuli with functional fragments, that must include a preposition phrase based on the study design, only include transitive verbs, as can be seen in the example (18-21). Therefore, a comparatively equal length of all stimuli is guaranteed.

Last, the contrasting words are either orthographically or prosodically marked in the condition with emphasis on the one hand, but in the condition without emphasis, on the other hand, the stimuli either do not contain any orthographic marking or the nuclear accent is not on the contrasting word but on the default position (cf. Féry 2011). The orthographic marking involved writing the respective words in uppercase letters. For an overview of stimuli with prosodical marking and stimuli with default intonation, see chapter 3.3.

A total of 56 critical items were selected for the study. However, due to the study's design, each participant only encountered 28 critical items. This was because participants were randomly assigned to either the written or auditory stimuli group. In addition, 56 filler items were included in the study, of which 28 were written and 28 were auditory items.

Filler items involved dialogues that incorporated either non-fragmental contrast such as in (22) or dialogues without any contrast such as (23). The order of critical and filler items was randomly arranged, with each item being presented on its own individual page.

- (19) A: *Peter hat die SÜDDEUTSCHE gelesen.*  
Peter AUX the Süddeutsche read  
'Peter read the Süddeutsche.'  
B: *Nein, er hat die FAZ gelesen.*  
no he AUX the FAZ read  
'No, he read the FAZ.'

(own stimuli)

- (20) A: *Peter hat in der Mensa zu Mittag gegessen.*  
Peter AUX in the canteen for lunch  
ate

Kommentiert [MS23]: Die Stimuli sind eigentlich von Featherston, aber er hat meines Wissens nirgends seine Materialien publiziert, deswegen kann ich es nicht zitieren. Fußnote?  
Update: siehe Mail von James

'Peter had lunch in the canteen.'  
 B: *Ja, zusammen mit Freunden.*  
 yes together with friends  
 'Yes, together with friends.'

(own stimuli)

The acceptability of the fillers varied. The filler in (22) and (23) represent structures associated with full acceptability. The fillers in (24) and (25) signify complete unacceptability.

(21) A: *Peter hat mit Freunden UNO gespielt.*  
 Peter AUX with friends UNO played  
 'Peter played UNO with friends.'  
 B: \**Nein, beim Stammtisch die Freunde*  
 no at.the regulars' table the friends  
*haben mit Vorliebe SKAT gespielt.*  
 AUX with preference Skat played  
 'No, at the regular's table the friends played skat with preference.'

(own stimuli)

(22) A: *Peter hat seinem Sohn ein Geschenk*  
 Peter AUX his son a gift  
*gemacht.*  
 made  
 'Peter gave a gift to his son.'  
 B: \**Ja, ein Fahrrad in die Schule zum Fahren.*  
 yes, a bike to the school for riding  
 'Yes, a bike to the school for riding.'

(own stimuli)

To ensure consistency, several steps were taken. First, all stimuli, i.e., critical and filler items, were adjusted to be in past tense and start with *Peter*. Second, roughly half of the filler items incorporated orthographic or prosodic marking on the contrasting words, while the remaining half did not incorporate such marking, mirroring the variation of the critical items. Next, out of the 56 filler items, 10 items represented full acceptability, 12 items indicated some acceptability, 12 items denoted neutrality in terms of acceptability, 12 items implied partial unacceptability, and 10 items signified full unacceptability. This ensured that there are the same number of acceptable and unacceptable filler items, and the range of acceptability reflects the range assumed for the critical items.

**Kommentiert [MS24]:** Die Stimuli sind eigentlich von Featherston, aber er hat meines Wissens nirgends seine Materialien publiziert, deswegen kann ich es nicht zitieren. Fußnote?  
 Update: Siehe Mail von James

**Kommentiert [MS25]:** Die Stimuli sind eigentlich von Featherston, aber er hat meines Wissens nirgends seine Materialien publiziert, deswegen kann ich es nicht zitieren. Fußnote?  
 Update: Siehe Mail von James



### 3.3 Recording of stimuli

In this section, I delve into the critical process of recording stimuli, discussing the methodologies and considerations involved in capturing high-quality audio or visual materials for the present research study.

Stimuli were recorded in the open-source toolkit Praat in a soundproof room, using a Blue Snowball ICE microphone and saved to be in WAV-format. Silences before and after the sentences were cut out of the sound files. As each stimulus represents a dialogue, the two parts had to be recorded individually. Each part of every stimulus was recorded three times, of which the one with the highest clarity, intelligibility, and adherence was chosen for the experiment. The first part was recorded by the voice actor Roman Pertl, henceforth speaker A. The second part was recorded by the author, henceforth speaker B. Both speakers are native German speakers, were familiar with the sentence prior to the recording, and had the opportunity to re-record any sentence indefinite times. All sentences were recorded by condition (cf. Sederous & Namboodiripad 2020). To demonstrated this, the pitch contour of the stimuli (20) and (21) in chapter 3.2 are illustrated in Figure 1 and 2.

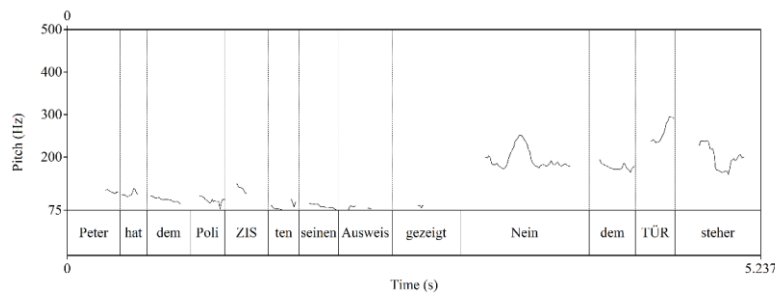


Figure 1: Pitch contour of stimuli with emphasis

Figure 1 shows the recorded intonation of the stimulus (20), that includes the emphasis of the contrasting words *Polizisten* ‘police officer’ and *Türsteher* ‘bouncer’. Both word are marked with L+H\* accent. The intonational contour of the preceding sentence in Figure 1 stands in stark contrast with its equivalent in Figure 2. Speaker A’s sentence in Figure 2 shows consistent, natural intonation. That is, the preverbal position, i.e., on the word *Ausweis* ‘identity card’, represents the default sentence accent (cf. Féry 2011). As becomes apparent in the pitch contours, the stimuli differ

in whether the word *Polizisten* 'police officer' is emphasized or not. While *Polizist* in the stimulus displayed in Figure 1 received L+H\* accent, it is de-accented in the stimulus in Figure 2. Note that the apparent distinction in pitch between the speakers is attributable to the gender contrast (cf. Simpson 2009), with the first speaker being male and the second speaker being female.

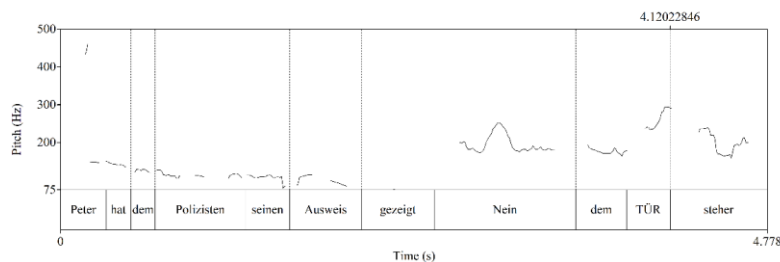


Figure 2: Pitch contour of stimuli without emphasis. The recording of speaker

B, that places L+H\* accent on the contrasting word *Türsteher* 'bouncer', were used for both conditions. That is, when combining the parts of speaker A and speaker B, the same recording of speaker B was used for the stimuli in the conditions with and without emphasis in the preceding sentence to ensure consistency and minimize confounding factors. Given that the contents of the sentences as well as the fragmentary answers were identical, the stimuli depicted in Figure 1 and Figure 2 solely vary in terms of the emphasis placed on *Polizisten* 'police officer'. This deliberate difference serves to eliminate alternative explanations for the observed outcomes, strengthening the validity of the results.

After the recording, the audio files of each speaker were concatenated in Praat. Therefore, each stimulus is composed of two merged audio files, seamlessly transitioning from the first to the second audio file without any audible disruptions or breaks.

Next, the audio files were controlled for loudness in Praat using the plugin (<https://www.praatvocaltoolkit.com/normalize.html>). The raw recordings as well as the combined, neutralized recordings can be found here: <https://shorturl.at/pCHM6>.

### 3.4 Data collection

In this chapter, the data collection process employed in our research study, which involved gathering acceptability judgments from participants, is discussed. I will outline the methods employed to obtain these judgments, including experimental design, and data collection procedures.

write a Paragraph on prolific and a Paragraph on collection procedure (see the respective chapter 3.1 on study design but go into more detail about the presentation format of stimuli, likert scale, control measures, etc.) and a Paragraph of managing and organizing of collected data for analyses, data cleaning, etc.

Participants are asked the following question about the dialogues in the experiment.

- (23) Wie natürlich klingt die Antwort der Sprecherin B?  
'How natural does speaker B's response sound?'  
(adapted from Featherston 2008, p. 6)

Kommentiert [MS26]: War das genau der Sprachlaut?

x

### 3.5 Participant recruitment and characteristics

Pilot study showed that sex and education play no significant role but age and geography must be controlled for (in CLEFS project summary gelesen, Quelle heraussuchen)

German native speakers

Age and geography must be controlled for!!

Sample size of 100

- Explain why
- Such sample sizes are much larger than necessary to obtain 80% statistical power for repeated-measures judgment tasks on 'standard' syntactic data with 2 to 4 factors because contrasts between syntactic stimuli generate large effect sizes.[68][69] [Quellen in project summary]
- Include paragraph of sample size in discussion

The data for this study was obtained from Prolific ([www.prolific.co](http://www.prolific.co)), with a minimum approval rate requirement of 90% to ensure reliable work. Additionally, the study was carefully set up to exclusively include participants who self-identified as native German speakers, as the present study focuses solely on the German language.

The study was completed by a total of 100 participants, comprising 69 males, 29 females, and 2 individuals who identified as diverse. The participants' ages varied between 19 and 73 years, with a mean age of 35.53 years. Among them, 9 participants did not hold a high school diploma, 29 participants completed high school as their highest level of education, 27 participants attained a bachelor's degree, and 35 participants had a higher degree beyond the bachelor's level.

Prior to the study, participants were informed about the Freiwilligkeit of the study and that the data is used for scientific research only. It was ensured that participants' privacy and data confidentiality were maintained. Due to the set-up of the study, the provided socio-demographic information cannot be traced back to individuals. Every participant was allowed to participate only once and was paid for their participation.

#### 4. Results and analysis

The following sections provides an overview of the study's findings. First, the raw data from the questionnaire, comprising participants' responses, is presented. Second, the methods used for data analysis are explained in detail. Next, the results of the data analysis are shown, followed by the addressing of the hypotheses. This comprehensive approach will help clarify the study's outcomes and shed light on the obtained results.

The study aimed to investigate to what extent the emphasis, modality, and fragment type in dialogues involving contrastive focus and fragmentary answers affects how native German speakers perceive the naturalness of such contrastive fragmentary answers. Below, the overall results of the study are displayed, followed by the presentation of the results for each investigated factor.

Out of the 100 participants who completed the study, 57 participants were presented with auditory stimuli, while 43 participants were faced with written stimuli. Due to the **set-up** of the study, all participants encountered stimuli with varying emphasis and varying fragment types.

**Kommentiert [MS27]:** Sometimes "set up", sometimes "set-up"

First, consider Figure 3, which shows participants' responses to all critical items. The graph displays participants' acceptability ratings based on the varying conditions of the three investigated factors. The ratings to stimuli with emphasis are illustrated in the graphs on the top, while the graphs on the bottom present participants' ratings to stimuli without emphasis, encompassing both conditions for the investigated factor emphasis. Similarly, ratings to auditory stimuli are illustrated on the left, while the rating for written stimuli is shown on the right. Moreover, functional fragments are shown in red, whereas lexical fragments are colored in blue.

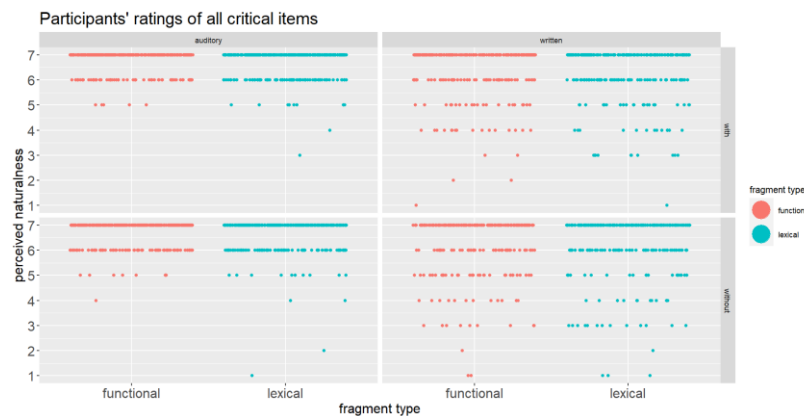


Figure 3: Scatter plot of participants' ratings of all critical items

As becomes apparent in Figure 3, by far the most responses cluster around 7 (fully acceptable). Furthermore, it can be said that overall, there is more variability in written stimuli, in stimuli with lexical fragments and in stimuli without emphasis. This will be analyzed in more detail later in the present chapter.

Now, consider Figures 4, displaying participants' ratings for each investigated factor in mosaic plots.

Participants' ratings shown for each factor

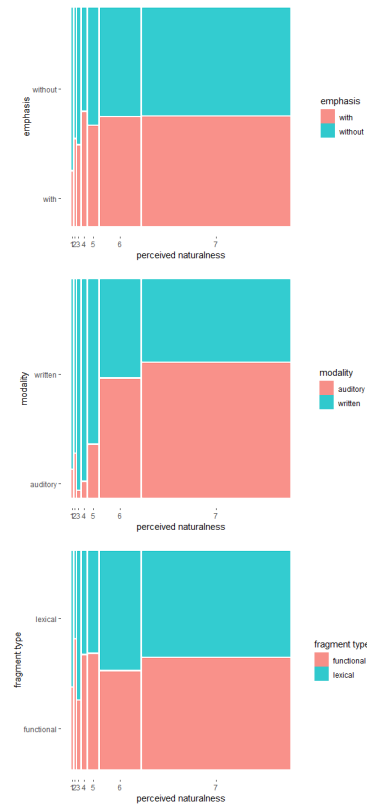


Figure 4: Mosaicplots of participants' ratings of each factor

As has already been shown in Figure 3, now becomes more apparent in Figure 4. That is, the majority of participants consistently assigned a rating of 7 (fully acceptable) on the 7-point Likert scale. Here, the investigated factors do not seem to influence participants' ratings. Nevertheless, when examining the impact of the investigated factor, we observe notable differentiation in participant responses, particularly for sentences that received ratings other than 7 (fully acceptable). This differentiation highlights the influence of the investigated factor on participants' acceptability judgments, revealing variations in how different factors affect participants' perceptions.

Next, consider Figure 5, illustrating the differences in participants' ratings between each condition, including the means and standard deviation.

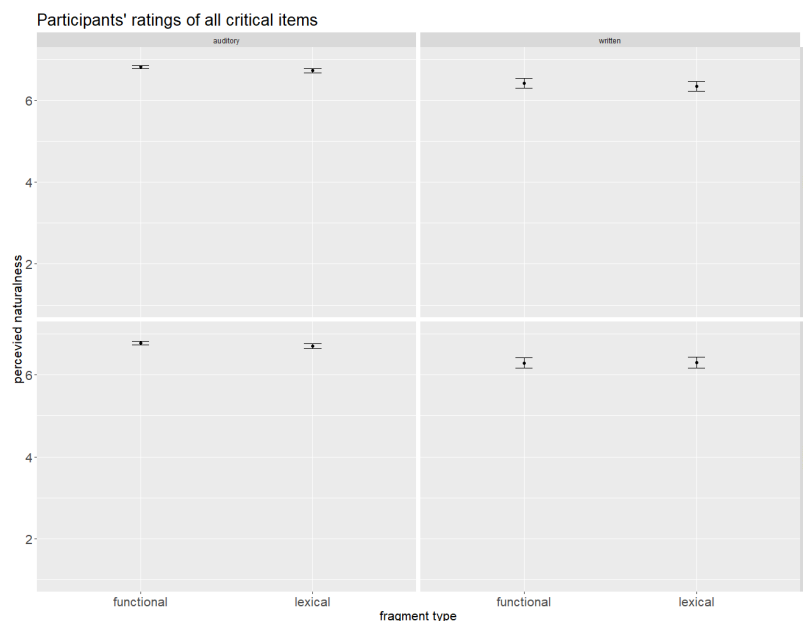


Figure 5: Boxplot of participants' ratings of all critical items

As can be seen in Figure 5, the means of participants' ratings are fairly similar in each condition. However, a more detailed presentation of the data and a statistical analysis is required to answer the study's research question. Therefore, participants' Likert scale responses were z-scored and analyzed using Cumulative Link Mixed Models (CLMM), using R 3.2.3 (cf. R Development Core Team 2015).

In order to facilitate meaningful comparisons and analyses, the 7-point Likert scale ratings provided by participants were standardized through a z-scoring procedure. Z-scoring involved subtracting the mean rating across all participants from each individual rating and then dividing by the standard deviation of the ratings. This transformation ensured that the ratings were placed on a common scale with a mean of 0 and a standard deviation of 1, allowing for relative comparisons and statistical analyses without the influence of varying response scales. Z-scoring enhances the interpretability and comparability of the ratings across different factors and conditions, enabling a more comprehensive exploration of the underlying patterns and effects.

**Kommentiert [MS28]:** Check if that is actually the version that was used

**Kommentiert [MS29]:** See Harris & Carlson 2018 references for how to cite them

**Kommentiert [MS30]:** Add quelle

The z-scored 7-point Likert scale ratings were analyzed using CLMM (cf. Christensen 2019). CLMM was chosen as the statistical approach due to its suitability for the nature of the data. The CLMM methodology effectively accommodates ordinal responses, making it a robust choice for analyzing the ordered Likert scale ratings. By accounting for the ordinal structure of the data, CLMM captures the inherent order and spacing between the response categories, providing a more accurate representation of participants' perceptions. CLMM models take into consideration both fixed and random effects, allowing to examine the impact of various predictor variables on the odds of participants choosing higher or lower response categories on the Likert scale. This approach is particularly advantageous when investigating factors that may influence participants' perceived naturalness in different conditions or contexts. The use of CLMM acknowledges the inherent correlations within the Likert scale ratings and provides a comprehensive understanding of the underlying relationships between the investigated factors and participants' responses. Overall, the application of CLMM aligns with the nature of our data and research objectives, offering a robust and tailored framework for exploring the effects of different factors on participants' ratings in our study.

In the following, the data undergoes analysis to address each of the three hypotheses that were the focus of the present study. Hence, consider Figure 6, which illustrates the means of participants' ratings of fragmentary answers, comparing stimuli with and without emphasis. Participants' ratings of stimuli with emphasis are shown in red, while those lacking emphasis are depicted in blue.

**Kommentiert [MS31]:** Add Quelle: zu Beschreibung über CLMM

Quelle für CLMM  
R. H. Christensen. 2019. ordinal – Regression models for ordinal data. R package version 2019.12-10



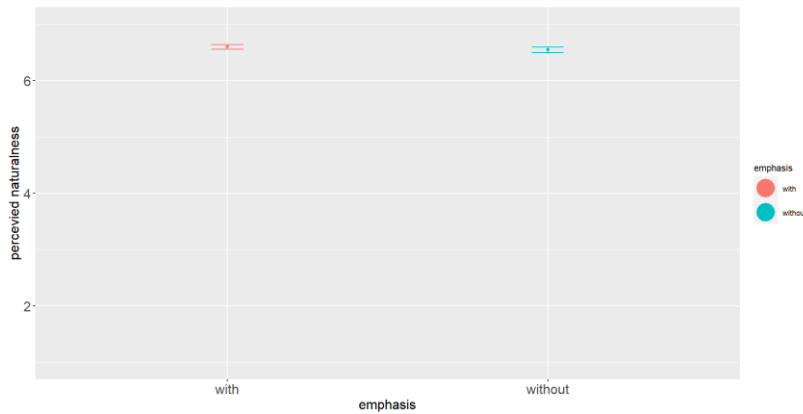


Figure 6: Comparison of participants' ratings of stimuli with and without emphasis

The first hypothesis examines whether the inclusion of emphasis on contrasting words impacts the perceived naturalness of contrastive, fragmentary answers. As depicted in Figure 6, a subtle distinction emerges in the ratings given to fragmentary responses. In fact, participants assigned a rating of 6.60 on the 7-point Likert scale to fragmentary answers with emphasis, whereas fragmentary answers lacking emphasis received a slightly lower rating of 6.55.

This difference aligns with the predictions of the first hypothesis, suggesting that acceptability ratings are higher for stimuli with emphasis compared to those without. Employing a CLMM, which accommodates potential variations among different participants and items within a given condition, the observed difference of emphasis is statistically significant with a p-value of 0.0268, adhering to the predetermined alpha level of 0.05. Therefore, the present data presents evidence in favor of the first hypothesis.

Next, consider Figure 7, which shows the mean ratings of auditory and written stimuli. Participants' ratings of auditory stimuli are illustrated in red, whereas those of written stimuli are colored in blue.

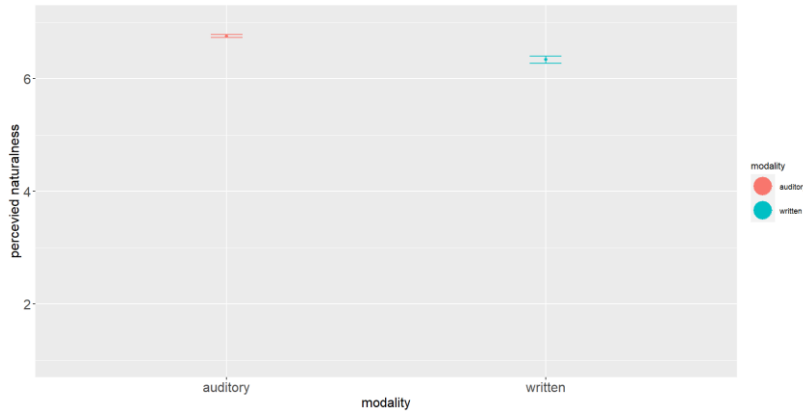


Figure 7: Comparison of participants' ratings of auditory and written stimuli

In the present experiment, contrasting words were emphasized either by orthographic marking in the written condition or by prosodic marking in the auditory condition. The present study does not only investigate the influence of emphasis on the perceived naturalness of fragmentary answers, but also delves into how the modality of presentation could influence how native speakers evaluate such fragmentary answers in terms of naturalness.

Hence, the second hypothesis analyses whether modality has an impact on how natural fragmentary answers are perceived. As depicted in Figure 7, a notable contrast can be observed between the ratings in the auditory condition as opposed to those in the written condition. That is, fragmentary answers in the auditory condition received a rating of 6.76 on the 7-point Likert scale, whereas fragmentary answers in the written condition received a slightly lower rating of 6.34.

As predicted by the second hypothesis, acceptability ratings are higher for auditory stimuli compared to their written counterparts. In the CLMM, the observed difference of modality holds statistical significance with a p-value of 0.0158, adhering to the predetermined alpha level of 0.05. Therefore, the present data provides evidence to reject the null hypothesis and accept second hypothesis that modality influences the perceived naturalness of fragmentary answers.

Next, consider Figure 8, which shows the mean ratings of stimuli with lexical and functional fragments. Participants' ratings of stimuli with functional fragments are colored in red, while those of stimuli with lexical fragments are depicted in blue.

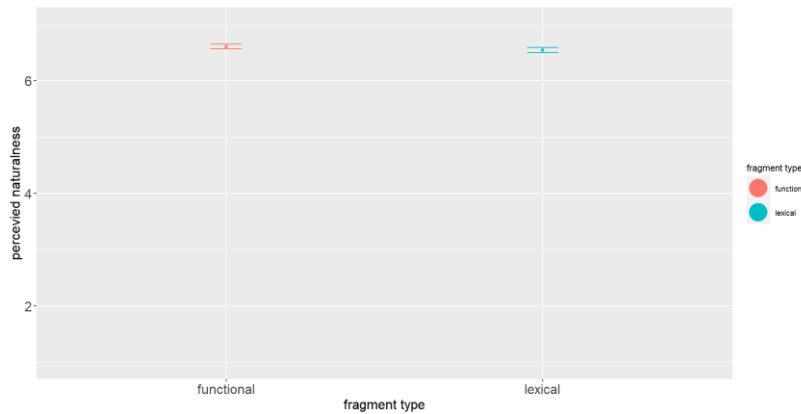


Figure 8: Comparison of participants' ratings of stimuli with functional and lexical fragments.

The third hypothesis explores whether fragment type influences the perception of naturalness in fragmentary answers. As illustrated in Figure 8, there is a subtle contrast in acceptability ratings between functional and lexical fragments. In fact, participants assigned a rating of 6.61 on the 7-point Likert scale to answers incorporating functional fragments, whereas those with lexical fragments received a slightly lower rating of 6.55.

Surprisingly, this difference in acceptability ratings does not coincide with the predictions of the third hypothesis, indicating that acceptability ratings tend to be higher for stimuli with lexical fragments than for stimuli with functional fragments. Instead, as shown in the present data, the ratings exhibit an inverse trend.

In the CLMM, the observed contrast in fragment type is statistically significant with a p-value of  $>0.01$ , adhering to the predetermined alpha level of 0.05. While the present data does not provide evidence in favor of the third hypothesis, it indicates that the difference in fragment type is statistically significant.

## 5. Discussion

In the following chapter, the key findings of the two studies are presented and the hypotheses are re-visited. In addition, unexpected results, confounding factors, and biases are discussed and comparisons to previous research in this field are drawn. Moreover, the implications for the future understanding of fragments in German are debated.

### 5.1 Interpretation of the findings

The results of the study reveal that overall, the vast majority of fragmentary answers were rated as 7 (fully natural). This was somewhat unexpected because as explained in chapter 2, contrastive, fragmentary answers naturally occur with pitch accent and in spoken language more than written language. Therefore, it was expected that contrastive, fragmentary answers in situations that do not resemble their natural environment would be rated far lower than what they have in the present study.

Nevertheless, the present research indicated that emphasis, modality, and fragment type significantly impact the perceived naturalness of fragmentary answers. Emphasis and modality influences participants' ratings in the expected way, even though with less of an impact as predicted. That is, ...

As for fragment type, while a statistically significant distinction between functional and lexical fragments was found, it was the other way around than expected. Lexical fragments were assumed to be receive higher acceptability ratings but in fact, functional fragments were got higher ratings. This could be explained by the choice of fragment types. In the present study, prepositions were used to represent functional fragments and nouns denoting human referents are used as an example for lexical fragments. Perhaps, one can assume that a different choice of words may result in different ratings. However, considering the limited number of functional words in German, in particular the ones that can be contrasted, prepositions seem to be the most appropriate ones to use. Similarly, nouns denoting human referents are probably the most used fragmentary answers in everyday speech and therefore, a suitable example to use in comparison to functional fragmentary answers.

The reliability of the data analysis was checked by assessing whether adding predictors to the model significantly improves the fit compared to a simpler CLMM model with only random intercepts for each participant. That is, the simpler model, so-called null model, was fit to the data and the Akaike Information Criterion (AIC) of this model was computed. AIC is a measure of the model's quality of fit while penalizing for the number of parameters. Lower AIC values suggest better model fit. Next, an analysis of variance (ANOVA) between the null model and the CLMM model used in the data analysis above was conducted to test whether adding predictors to the model significantly improves the fit. The result indicates that using emphasis, modality, and fragment type as predictors significantly improves the model fit. The difference between the two models holds statistical significance with a p-value of  $>0.01$ , adhering to the predetermined alpha level of 0.05

Kommentiert [MS32]: Quelle

The lower AIC and significant likelihood ratio test suggest that the more complex model that uses emphasis, modality, and fragment type as predictors, provides a better fit to the data and that the predictors have a significant impact on the model's performance.

-> Discuss how reliable the model is (see Santina's and Felix' notes)

- Confounding factors

- o Age
- o Geography
  - From a pilot study that assessed the influence of sociolinguistic factors on how fragmentary answers are perceived by native Dutch speakers, it was discovered that sex and educational background play no significant role, whereas age and geography must be controlled for. Therefore, the present study does not delve

deeper into the differences between varying age groups or dialects of participants. Instead, the

**Kommentiert [MS33]:** Alter und geographie konnten aber nicht eingeschränkt werden?

- - Not possible in Prolific
    - Not enough participants
    - Pool not big enough
- Biases
  - Participants knew that they were rating language structures according to their naturalness
  - However, participating in a research study is not a natural setting
  - So many ratings of 7 (fully natural): perhaps, because in everyday speech, even native speakers make mistakes, so therefore ungrammatical structures can be seen as natural too
  -

## 5.2 Comparison with previous studies and theoretical predictions

X

Highlight similarities and differences

- Read literature again
- Comparison of literature and present study's findings
  - Contrastive focus is associated with pitch accent
  - Therefore, in theory, contrastive focus in speech that lacks such prosodic marking should be rated as less natural
  - And in fact, it was rated less natural and the difference hold statistical significance
  - However, the ratings of the stimuli lacking emphasis were thought to be even lower
- Compare findings

## 6. Conclusions

The final chapter provides an overview of the findings from the present study, accompanied by an examination of its contributions to the realm of

fragments. In addition, the study's limitations are discussed and an outlook to future research is given.

#### 6.1 Summary of findings

X

Main findings of the research

#### 6.2 Contributions to the field

X

Highlight contributions of your research to the existing body of knowledge

Discuss the broader implications of your findings for the field of fragment acceptability and related research areas

Discuss the value of this pilot study for future research

#### 6.3 Limitations of the study

X

Discuss any limitations or potential biases that may have affected your research

What could not be answered?

- First, in the present study, recruitment is conducted via the Internet, which is associated with certain biases. For instance, only people who have access to the Internet, who have the required time to fill in the survey, and who are interested in participation are recruited. As a result, certain groups of people may be underrepresented, while others may be overrepresented in the sample. However, conducting the study online perhaps reduced the risk of participants changing their responses because they are being observed by the researchers. This is less likely to be the case in the present study since participants' responses were collected anonymously online.
- Only German

#### 6.4 Suggestions for future research

X

What aspects need further research?

Furthermore, it can be said that other factors may come at play when investigating the factors that influence how contrastive, fragmentary answers are perceived by native speakers. First, the difference in emphasis might be clearer if only the contrasting word instead of the complete phrase were given in the fragmentary answer. That is, for functional fragments, the fragmentary answer would only consist of the preposition instead of the prepositional phrase. Similarly, for lexical fragments, the answer would only consist of the contrastive noun instead of the noun phrase.

- (1) A: Peter hat AB 18 Uhr im Kino gearbeitet.  
'Peter worked at the cinema FROM 6pm.'  
B: Nein, BIS.  
'No, UNTIL.'

(own example)

- (2) A: Peter hat ab 18 Uhr im Kino gearbeitet.  
'Peter worked at the cinema from 6pm.'  
B: Nein, bis.  
'No, until.'

(own example)

As illustrated in (28-29), fragmentary answers can be reduced in such a way that they only incorporate the contrastive words. The dialogues in (28) and (29) differ in whether the contrasting words are emphasized or not. The present study has already shown that contrastive dialogues involving fragmentary answers receive lower ratings in terms of naturalness if the contrasting words are not either orthographically or prosodically emphasized. In such cases, the reader or hearer struggles more to understand the dialogue (29) because they are not primed for the contrast and it is difficult to understand what speaker B tries to convey with their message. Not only is the hearer required to figure out based on the preceding sentence what the preposition *until* refers to, e.g., until 6pm, but also, they must find the contrast and adjust the phrase accordingly, e.g., Peter worked at the cinema until 6pm instead of from 6pm. In cases such as (28), in which the contrastive words are emphasized, one can assume that is much easier for the hearer to understand



what the preposition refers to, what it contrasts with and what the propositional content of speaker B's utterance is. Therefore, it can be assumed that fragmentary answers with a complete prepositional or noun phrase and emphasized contrastive words receive the highest acceptability ratings. Acceptability ratings lower if the contrastive words lack emphasis. Fragmentary answers without complete phrases and without emphasis receive the lowest acceptability ratings.

**Kommentiert [MS34]:** Ggfs. weglassen / kürzen

Next, issuehood might also influence how contrastive, fragmentary answers are perceived. If not-at-issue content is contrasted, it may be more difficult for the hearer or reader to understand the meaning of speaker's B utterance. Since all dialogues in the present

**Kommentiert [MS35]:** Ggfs. weglassen

7. References

X

Insert references from Zotero

## 8. Appendix

### 8.1 Abbreviations, symbols and other notational conventions

?	questionable/marginal acceptability
#	infelicitous
*	ungrammatical
$1 \dots t_1$	syntactic movement
—	omitted linguistic material
$XP^i \dots YP^i$	coreference
//	intonational phrase boundary
UPPERCASE	emphasis (orthographical marking or pitch accent)
[...]F	focused position
[[...]]	denotation/semantic representation
L+H*	pitch accent
A, B, ...	speaker
ACC	accusative
AJT	acceptability judgment tasks
A-movement	argument movement
ASG	adposition stranding generalization
CLMM	Cumulative Link Mixed Models
DAT	dative
iff	if and only if
ISG	island sensitivity generalization
MaxQUD	maximal QUD
MDA	move-and-delete approach
NOM	nominative
p-omission	preposition-omission
p-stranding	preposition-stranding
QUD	Question Under Discussion
SQA	syntactic question approach

**Kommentiert [MS36]:** Check the entire list to see which abbreviations were even used in the thesis

## 8.2 List of critical items

The following list of critical items only includes written items. The auditory critical items can be found here: <https://shorturl.at/iwR78>.

- (1) A: Peter hat seinem BRUDER ein Buch geschenkt.  
B: Nein, seinem VATER.
- (2) A: Peter hat seinem Bruder ein Buch geschenkt.  
B: Nein, seinem Vater.
- (3) A: Peter hat dem POLIZISTEN seinen Ausweis gezeigt.  
B: Nein, dem TÜRSTEHER.
- (4) A: Peter hat dem Polizisten seinen Ausweis gezeigt.  
B: Nein, dem Türsteher.
- (5) A: Peter hat seinem CHEF den neuen Mitarbeiter vorgestellt.  
B: Nein, seinem KOLLEGEN.
- (6) A: Peter hat seinem Chef den neuen Mitarbeiter vorgestellt.  
B: Nein, seinem Kollegen.
- (7) A: Peter hat dem MALER ein Getränk angeboten.  
B: Nein, dem GÄRTNER.
- (8) A: Peter hat dem Maler ein Getränk angeboten.  
B: Nein, dem Gärtner.
- (9) A: Peter hat seinem KOLLEGEN Urlaubsbilder gezeigt.  
B: Nein, seinem NACHBARN.
- (10) A: Peter hat seinem Kollegen Urlaubsbilder gezeigt.  
B: Nein, seinem Nachbarn.
- (11) A: Peter hat seinem NEFFEN Werkzeug geschenkt.  
B: Nein, seinem NACHBARN.
- (12) A: Peter hat seinem Neffen Werkzeug geschenkt.  
B: Nein, seinem Nachbarn.
- (13) A: Peter hat seinem VORGESETZTEN einen Kaffee gebracht.  
B: Nein, seinem MITBEWOHNER.
- (14) A: Peter hat seinem Vorgesetzten einen Kaffee gebracht.  
B: Nein, seinem Mitbewohner.
- (15) A: Peter hat AB 18 Uhr im Kino gearbeitet.  
B: Nein, BIS 18 Uhr.
- (16) A: Peter hat ab 18 Uhr im Kino gearbeitet.  
B: Nein, bis 18 Uhr.
- (17) A: Peter hat BIS August Miete gezahlt.  
B: Nein, AB August.
- (18) A: Peter hat bis August Miete gezahlt.  
B: Nein, ab August.
- (19) A: Peter hat MIT seinem Bruder Unterschriften gesammelt.  
B: Nein, OHNE seinen Bruder.

**Kommentiert [MS37]:** Link to PRIVATE repository.  
Publish repository? Or publish recordings somewhere else?

- (20) A: Peter hat mit seinem Bruder Unterschriften gesammelt.  
B: Nein, ohne seinen Bruder.
- (21) A: Peter hat OHNE sein Team einen Vortrag gehalten.  
B: Nein, MIT seinem Team.
- (22) A: Peter hat ohne sein Team einen Vortrag gehalten.  
B: Nein, mit seinem Team.
- (23) A: Peter hat VOR seiner Mittagspause seine Chefin angerufen.  
B: Nein, NACH seiner Mittagspause.
- (24) A: Peter hat vor seiner Mittagspause seine Chefin angerufen.  
B: Nein, nach seiner Mittagspause.
- (25) A: Peter hat NACH seinem Urlaub den Handwerker gerufen.  
B: Nein, VOR seinem Urlaub.
- (26) A: Peter hat nach seinem Urlaub den Handwerker gerufen.  
B: Nein, vor seinem Urlaub.
- (27) A: Peter ist VOR seinem Einkauf noch zur Bank gegangen.  
B: Nein, NACH seinem Einkauf.
- (28) A: Peter ist vor seinem Einkauf noch zur Bank gegangen.  
B: Nein, nach seinem Einkauf.

### 8.3 List of filler items

The following list of filler items only includes written items. The auditory filler items can be found here: <https://shorturl.at/jsHV1>. The acceptability of the fillers varied, with A representing full acceptability, B indicating some acceptability, C denoting neutrality in terms of acceptability, D implying partial unacceptability, and E signifying complete unacceptability.

- A1 A: Peter hat in der Mensa zu Mittag gegessen.  
B: Ja, zusammen mit Freunden.
- A2 A: Peter hat den Gegenspieler vorsätzlich gefoult.  
B: Ja, den Stürmer.
- A3 A: Peter hat die SÜDDEUTSCHE gelesen.  
B: Nein, er hat die FAZ gelesen.
- A4 A: Peter hat einen ERDBEERKUCHEN gebacken.  
B: Nein, er hat einen SCHOKOKUCHEN gebacken.
- A5 A: Peter hat den KAFFEE gekocht.  
B: Nein, er hat den TEE gekocht.
- B1 A: Peter hat dem Fürsten jemanden empfohlen.  
B: Ja, dem Fürsten den Maler.
- B2 A: Peter hat dem Gast ein Getränk empfohlen.  
B: Ja, dem Gast den Wein.
- B3 A: Peter hat seinem Neffen ein Geschenk gegeben.  
B: Ja, seinem Neffen ein Fahrrad.

**Kommentiert [MS38]:** Link to github, however repository is private! Publish repository?

**Kommentiert [MS39]:** In fillers reden wir über acceptability, aber bei critical von naturalness, weil acceptability = was erwartet wird vs naturalness = was PP wirklich angeben

- B4 A: Peter hat geglaubt, dass sein CHEF Urlaub hat.  
B: Nein, er hat geglaubt, sein Chef gibt IHM Urlaub.
- B5 A: Peter hat sich GEWUNDERT, weil Maria zu Besuch kam.  
B: Nein, er hat sich GEFREUT, weil Maria hat Geschenke mitgebracht.
- B6 A: Peter hat angenommen, dass Franz ihm das Radio SCHENKT.  
B: Nein, er hat angenommen, er VERKAUFT ihm das Radio günstiger.
- C1 A: Peter hat dem Kunden etwas gezeigt.  
B: Ja, dem Kunden sich selbst im Spiegel.
- C2 A: Peter hat den Mann nach etwas gefragt.  
B: Ja, wen wer in dieser Affäre betrügt.
- C3 A: Peter hat seinen Nachbar zu dem Unfall befragt.  
B: Ja, wem wer aufgefahren ist.
- C4 A: Peter hat gedacht, dass der POLITIKER bestochen wurde.  
B: Nein, in ROTTENBURG hat Peter gedacht, hat der Händler den Politiker bestochen.
- C5 A: Peter hat erzählt, dass Franz einen UNFALL hatte.  
B: Nein, auf einer KREUZUNG hat Peter erzählt, hatte Franz einen Unfall.
- C6 A: Peter hat gehört, dass der Lehrer WÄHREND seinem Urlaub gekündigt hat.  
B: Nein, VOR dem Urlaub hat Peter gehört, hat der Lehrer gekündigt.
- D1 A: Peter hat ihn als kompetenten Begleiter empfohlen.  
B: Ja, sich selbst.
- D2 A: Peter hat Maria einen Brief geschrieben.  
B: Ja, einander.
- D3 A: PETER hat es dem neuen Tenor zugemutet.  
B: Nein, der KOMPONIST hat dem neuen Tenor es zugemutet.
- D4 A: Peter hat seinen Sohn eine GESCHICHTE vorgelesen.  
B: Nein, Peter hat ein GEDICHT ihm vorgelesen.
- D5 A: Peter hat Maria eine E-MAIL geschickt.  
B: Nein, er hat eine SMS ihr geschickt.
- D6 A: Peter hat am liebsten die FAZ gelesen.  
B: Nein, er liest am liebsten die SÜDDEUTSCHE, obwohl er lebt jetzt in Düsseldorf.
- E1 A: Peter hat den Rasen gemäht.  
B: Ja, obwohl der Hitze.
- E2 A: Peter hat den Fernseher eingeschaltet.  
B: Ja, um zu sehen eine Fernsehserie.
- E3 A: Peter hat seinem Sohn ein Geschenk gemacht  
B: Ja, ein Fahrrad in die Schule zum Fahren.

- E4    A: Peter hat mit Freunden UNO gespielt.  
      B: Nein, beim Stammtisch die Freunde haben mit Vorliebe  
          SKAT gespielt.
- E5    A: Peter hat Franz mit einem Geschenk überrascht.  
      B: Nein, da gerechnet mit hat der Franz natürlich nicht.

#### Declaration of Authorship

I hereby confirm that this paper and the work presented in it is entirely my own. Where I have consulted the work of others this is always clearly stated. All statements taken literally from other writings or referred to by analogy are marked and the source is always given. This paper has not yet been submitted to another examination office, either in the same or similar form.

Tübingen, September 23<sup>rd</sup>, 2023

Miriam Schiele

Miriam Schiele