Eberhard Karls Universität Tübingen

Seminar für Sprachwissenschaft

**Acceptability Judgments On**
**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 23rd, 2023

Table of contents

Table of figures

Abstract

This paper focuses on the factors that influence how native speaker perceive fragments in dialogues involving contrastive focus in German. A total of 100 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…

1. Introduction

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

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).

(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)

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 the 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 [MS4]: 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 the 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 one 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 function 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.

The present study examines several 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 mostly 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

contrastive fragments. By determining the best medium for reliably obtaining these judgments, the 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 the 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 the 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 (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

In this chapter, the fragment theory is explained, encompassing its linguistic foundations. Moreover, the methodological approach of acceptability judgment tasks is discussed. Furthermore, 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.

**Kommentiert [MS6]:** Nochmal überprüfen, ob Reihenfolge stimmt

### 2.1 Fragment theory and its linguistic foundations

In a general sense, is used to refer to an interface phenomenon that occurs among syntax, semantics, and information structure, where linguistic material is omitted (cf. Lobeck, 1995; Winkler, 2019). In other words, "there is meaning without form" (Merchant, 2019, p. 19). While there are numerous types of ellipsis, an in-depth analysis of each of those types would go beyond the scope of this paper. Instead, the present paper focuses on a specific type of ellipsis, i.e., fragments. For a first understanding of fragments, consider (5).

(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, 2005, p. 661)

Ben's answer *Alex* in (5) only consists of one word and yet, the reader can easily derive that Ben's utterance is intended to convey that Beth is bringing Alex. Such short answers are called fragments (Merchant, 2005).

In the following, Merchant's (Merchant, 2005) theory is adopted, according to which fragments move to the clause-peripheral position and that

the concomitant ellipsis operation deletes the remaining constituents. Therefore, the remnant of ellipsis is the part of the underlying clause that survives ellipsis (cf. Griffiths et al., 2023). Hence, Ben's answer *Alex* is the only pronounced constituent of its underlying clause *Beth is bringing Alex*. For an in-depth explanation of how fragmentary answers are comprehended, see chapter 2.3.

Similar to other elliptical structures, fragments must meet specific licensing conditions. For instance, omission of material is only permissible if the material is given within the discourse (cf. Winkler, 2019, Merchant, 2001). That is, if a part of a sentence is anaphoric to linguistic material in the previous discourse or is inferred from it, that part is either deaccented or not phonetically realized (cf. Winkler, 2019). In the example (5), the given material *Beth is bringing* from the clause *Beth is bringing Alex* is eligible for omission since it can be deducted from the prior conversation. However, who Beth is bringing is not clear from the context. Therefore, the information *Alex* is new to the context and consequently, cannot be diminished.

## 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.

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. Krifka, 2008; Wagner, 2012). An example of contrastive focus is illustrated in (7), where the brackets subscripted with $F$ indicate the constituent in focus.

(7)    A:    Mary stole the cookie.
           B:    No, [Peter]$_F$ stole the cookie!
                           (adapted from Krifka, 2008, p. 252)

The sentence in (7A) 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 (6). As can be seen in (7B), 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). Placing focus on *Peter* serves to convey implicitly that the assertion made by speaker A is erroneous, indicating that Mary did not stole the cookie, but rather, it was Peter who did. Therefore, speaker B's utterance functions as a correction to speaker A's statement.

After examining (7B), which represents a complete sentence as correction, now consider the fragmentary equivalent (8B).

(8)    A:    Mary stole the cookie.
           B:    No, [Peter]$_F$!
                           (adapted from Krifka, 2008, p. 252)

The full correction in (7B) and the elliptical correction in (8B) have the same semantic meaning and pragmatic function, despite of the fact that (8B) 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.

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 (8), 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. Griffiths & Lipták, 2014; Krifka, 2008).

Contrastive focus has often been characterized phonologically with a L+H* pitch for German (Wagner, 1999). This is defined in the Contrastive Remnant Condition (henceforth, CRC), as formulated in (9).

(9) Given information licenses a contrastive focus interpretation of the remnant(s). The contrastive remnant(s) must be assigned a strong contrastive pitch accent.

(Winkler, 2019, p. 363)

Adhering to the terminology used in the ToBI-framework (cf. Beckman & Ayers, 1997), the bitonal L+H* pitch accent is a combination of a low tone (L) and a high pitch (H*). When combined as L+H* pitch accent, it indicates a pitch accent pattern where a low tone is followed by a high tone with a rising pitch. This pattern is often associated with a pitch accent that starts low and rises to a high pitch, creating a pattern of emphasis and prominence on the syllable. It's used to mark significant or accented syllables within an utterance (ibid.).

The prosodic marking on the remnant and correlate indicates their contrast with each other (cf. Winkler, 2019; Rasekhi & Harris, 2021). Adhering to the CMC formulated in (5), contrastive focus is placed on *Peter* in both (7B) and (8B). The importance of 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 involving contrastive focus. First, the process of the comprehension of elliptical utterances is explained. Then, I discuss to what extent emphasis and semantics 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.

(10)    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 (8B), 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*$_1$ *stole the cookie* t$_1$.

Even though the significance of information structure, specifically the differentiation between previously mentioned and newly introduced information, has been demonstrated in chapter 2.1, it remains essential to explore additional factors that influence the interpretation of clausal ellipsis. Hence, the subsequent section delves into examining both the importance of parallelism, the correlate's position, and the role of orthographic and prosodic emphasis.

Morphological Parallelism refers to the observation that "the processor favors correlate-remnant pairings for which the DPs are maximally similar along semantic and morphological dimensions" (Rasekhi & Harris, 2021, p. 7). As a method for resolving sentence ambiguity and determining the appropriate correlate for the remnant in elliptical structures, it stands as a valuable tool, especially in the context of German elliptical structures. This is due to the case marking present on all determiner phrases (DPs) in the German language. Consequently, the processor can readily pinpoint the accurate correlate-remnant pairing by identifying matching case markers (cf. Rasekhi & Harris, 2021).

However, parallelism between the correlate and remnant is not limited to their morphological properties. Frazier et al. (1984) found that if the ante-

cedent clause and the elliptical structure both had similar syntactic structures and thematic roles, elliptical structures can be processed faster than dissimilar sentences.

Another factor that influences how elliptical structures are interpretated is the locality of the correlate. That is, the processor tends to choose the nearest constituent in the preceding clause to be contrasted with the remnant (cf. Harris, 2015). For instance, consider (11).

(11)   John didn't take the poodle to the park, let alone…
       (a) the zoo. (local contrast with *the park*)
       (b) the pug. (Nonlocal contrast with *the poodle*)
                        (Harris & Carlson, 2016, p. 6, parentheses in original)

Therefore, a contrast between correlate and remnant such as in (11a) is preferred because the correlate and remnant are structurally closer together than in (11b). If this locality bias is violated, the processing of the sentences is slowed down. That is, the eye-tracking study by Frazier & Clifton (1998) indicated that ambiguous sentences with two potential correlates were read faster than unambiguous sentences and a forced-choice study showed that the more local DP of an ambiguous sentence was selected as appropriate correlate more often than the less local DP. Furthermore, Harris & Carlson (2016) conducted self-paced reading studies to explore how elliptical structures involving *let alone* phrases are processed. Their results indicate that sentences that do not adhere to the locality bias incur a processing cost (ibid.).

Moreover, Rasekhi & Harris (2021) investigate to what extent those factors facilitate the comprehension of elliptical clauses. Their study indicates that in Persian, whether the correlate and remnant have maximally similar semantic and morphological properties seems to play a greater role than the position of the remnant when finding the appropriate contrasting correlate for the remnant. In addition, a processing cost might also occur if the DPs differ in number (cf. Black et al. 1985).

An alternative reasoning for the locality bias is that the closest DP is selected as the correlate not due to its structural proximity, but because object DPs naturally carry the primary sentence accent in SVO sentences,

even during silent reading (cf. Rasekhi & Harris 2021). This holds true for German as well, as the same nuclear accent is also placed on the object DPs in present perfect sentences in the German language (cf. Féry, 2011; Féry & Herbst, 2004). Hence, the significance of orthographic and prosodic emphasis becomes another pivotal element influencing the perception of contrastive fragments.

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.

Moreover, 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. Carlson et al. (2009) investigated whether pitch accent affects how ambiguous replacive sentences such as (12) are interpreted.

(12)   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)

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 impact 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. Akinnaso, 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.

Furthermore, it can be postulated that not only emphasis and modality have the potential to affect the perception of contrasting, fragmented responses, but the semantics of the words used could also play a role in determining how effortlessly the listener or reader establishes a focus-oriented anaphoric connection.

Words in language can be divided by their meaning. While lexical words consist of content words such as nouns, verbs, and adjectives, function words are a closed group, have grammatical meaning, and include, for example, articles, prepositions, and pronouns (Biber et al. 1999). A number of studies have shown that lexical words often play a more crucial role in understanding messages. This can be seen in reading as well as in spoken language. In an eye-movement study, Roussel et al. (2018) found that function words are more likely to be eluded than lexical words and that function words are fixated less often and with shorter gaze durations than lexical words. In spoken language, stress is usually placed on lexical words (cf. Chomsky & Halle, 1991; Féry & Herbst, 2004).

In the context of this current investigation, participants will be exposed to conversations containing both lexical and functional fragments, with the aim of examining potential divergences. It is hypothesized that dialogues involving lexical fragments will receive higher acceptability rating than those incorporating functional fragments.

**Kommentiert [MS9]:** No PDF, Not in Zotero: https://books.google.de/books?hl=de&lr=&id=qSl-HEAAAQBAJ&oi=fnd&pg=PR1&ots=gnf3gRG_lw&sig=LpL847POM0JoRNOdwyGE7lS8kgk&redir_esc=y#v=onepage&q&f=false

**Kommentiert [MS10]:** Find other source?

## 2.4 Acceptability judgment tasks (AJTs) in linguistics

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, on can gain insights about the grammaticality of the respective syntactic patterns (cf. Schütze, 2016; Sprouse et al., 2013). Instead, in the present study, participants were asked to rate in their acceptability. That is, how natural the sentence sounds to them.

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 only considers informal methods on the one hand and Likert scales as formal method on the other hand, leaving out other formals 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 includes 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 unnatural* and 7 as *fully natural*. A 7-point scale was chosen because it allows participants to identify structures that are neither natural nor unnatural (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 give 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 (cf. Jasso, 2022; Kayali, 2023, 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.

**Kommentiert [MS15]:** Seitenzahlen fehlen .. Quelle ggfs. weglassen?

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 hypotheses to be investigated, 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.

The study design is based on two short pilot study, which was completed by a total of 18 participants. The aim of the pilot studies was to ascertain the clarity of instructions presented in the introductory pages of the experiment and to confirm of the study itself, including aspects such as the audio files and the randomized grouping process.

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 natural structures to 7, indicating full unnaturalness. Prior, fully natural, fully unnatural, and neither natural nor unnatural 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 natural nor unnatural, 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.

(13)  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)

(14)  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.)

(15)  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
          b*estimmt nie        werfen!*
          certainly never        throw
          'No, the towel, he would certainly never throw that in!'
                            (adapted from Wierzba et al., 2023, p. 16)

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 natural*, (16) is identified as *fully unnatural* and (17) is used as an example of an utterance that is *neither natural nor unnatural*.

   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.

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).

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

(19)   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. Moreover, to allow for an unbiased research on contrastive focus, only the indirect object, that precedes the direct object, was chosen for the contrastive focus in the allow for an unbiased research on contrastive focus in the condition with lexical fragments, as outlined in chapter 2.3. Similarly, in the condition with functional fragments, the prepositional phrase precedes the object phrase to ensure a consistent pattern.

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

**Kommentiert [MS18]:** For discussion: all stimuli were adjusted to have definite remnant and disambiguated with dative marked etc. Therefore, parallelism hypothesis cannot be checked in the present stud

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.

(20)  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)

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

(22)  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)

(23)  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 natural and unnatural filler items, and the range of acceptability reflects the range assumed for the critical items.

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.
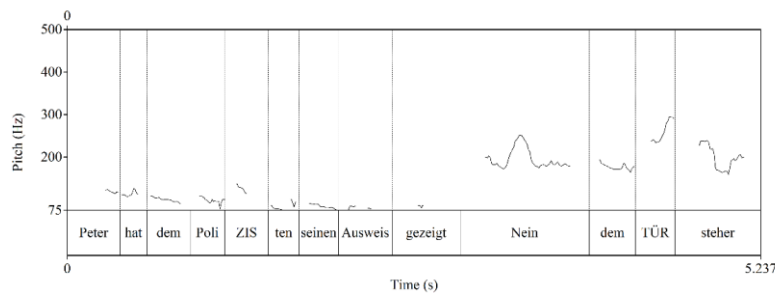


Figure 1: Pitch contour of stimuli with emphasis

Figure 1 shows the recoreded 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-accenated 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.
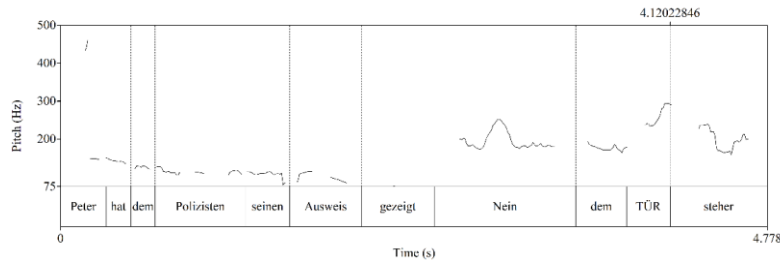


Figure 2: Pitch contour of stimuli without emphasis.

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

Acceptability judgements were crowdsourced from *Prolific* (cf. Prolific Academic 2019). Prolific serves as an online platform employed to recruit participants for research studies. It ensures a high level of transparency for both researchers and participants, features a user-friendly interface, and enables the pre-screening of potential participants (cf. Palan & Schitter, 2018).

In the experiment, stimuli were presented either as text or audio file. Chapter 3.2 gives an overview of how the stimuli were selected. For a description of how the dialogues were recorded, see chapter 3.3. In the auditory scenario, participants were able to provide their ratings only after they had finished listening to the audio file. In both conditions, participants were required to assign a rating to the current dialogue before proceeding to the subsequent one, and the presentation occurred one dialogue at a time. Participants were not permitted to revisit prior dialogues to modify their ratings. There were no time constraints in place. Participants were allowed to take as much time as necessary for reading the dialogues and potentially relistening to the audio files.

When presented with the dialogues, participants were asked the following question in the experiment.

(24)     Wie natürlich wirkt die Aussage der Sprecherin B auf Sie?
         'How natural does speaker B's response sound?'
                              (adapted from Featherston, 2008, p. 6)

The question of naturalness was taken from other studies incorporating AJTs (cf. Featherston, 2008). However, since all participants were native German speakers, the question was translated to German. Moreover, the question was adapted in such a way that it includes *auf Sie* 'to you', asking for participants' individual unconscious knowledge of the German language. Therefore, it is ensured that participants do not rate the naturalness of the

presented dialogues based on prescriptive grammar rules but their own intuition as a native speaker.

Acceptability ratings could be given on a 7-point Likert scale. For an in-depth overview of AJTs and Likert scales, see chapter 2.4.

Due to the setup of the study, it was ensured that participants were randomly assigned to different conditions and although participants were grouped into either the written or auditory condition, they were unaware that there were any other conditions, thereby effectively minimizing the potential for biases.

The study was structured in a manner that ensured participants' socio-demographic information couldn't be connected to individuals, thereby preserving their anonymity. With the exception of removing filler items, no data points were eliminated from the dataset. The analysis incorporated the complete dataset obtained from the experiment's critical items.

3.5 Participant recruitment and characteristics

The data for this study was obtained from *Prolific*, with a minimum approval rate requirement of 90% to ensure reliable work. Additionally, the study was carefully setup 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.

Before the commencement of the study, participants were acquainted with the voluntary nature of their involvement and the exclusive utilization of their data for scientific research purposes. Stringent measures were taken to uphold the privacy and confidentiality of participants' data. Owing to the study's design, the socio-demographic details provided cannot be linked to

individual identities. 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 extend 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 setup of the study, all participants encountered stimuli with varying emphasis and varying fragment types.

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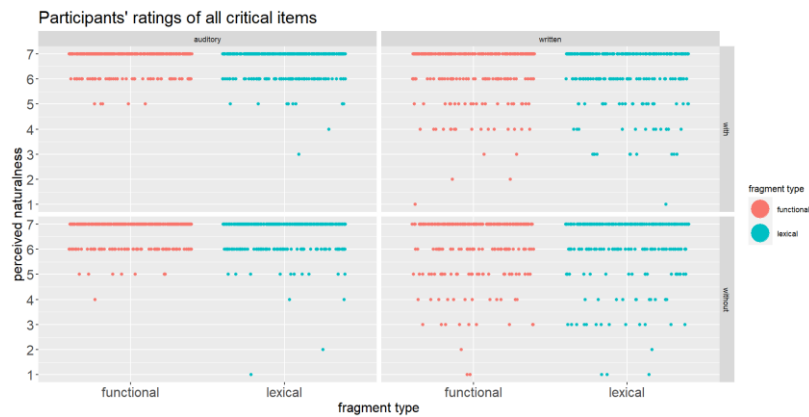
Figure 3: Scatter plot of participants' ratings of all critical items

As becomes apparent in Figure 3, by far the most responses cluster around 7 (fully natural). 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.
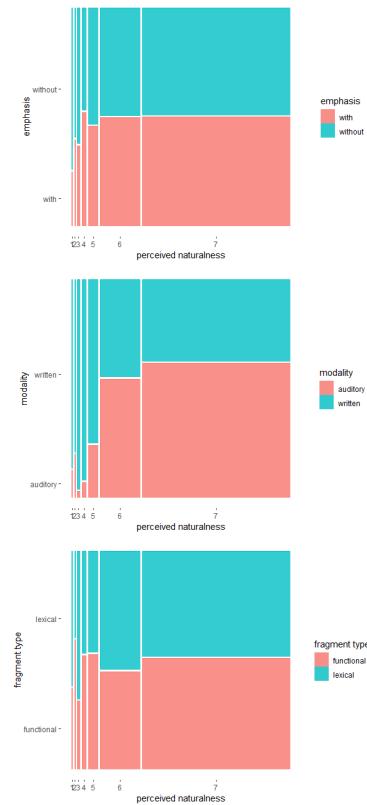
Figure 4: Mosaic plots of participants' ratings of each factor

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 natural) 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 natural). 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.
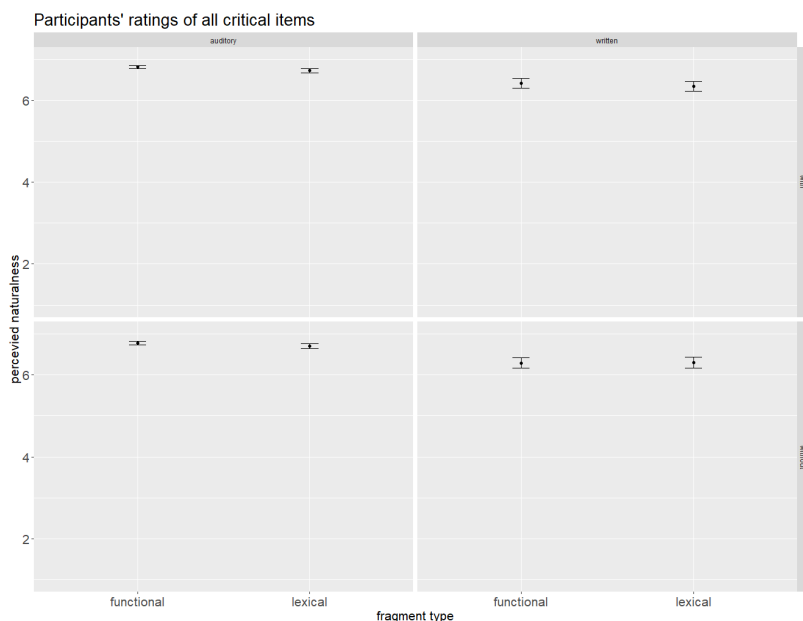
Figure 5: Box plot of participants' ratings of all critical items

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 version 4.2.1 (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. 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 (cf. Curtis et al. 2016).

The z-scored 7-point Likert scale ratings were analyzed using CLMM (cf. Christensen 2019). CLMM was chosen as the statistical approach due

**Kommentiert [MS23]:** See Harris & Carlson 2018 references for how to cite them

**Kommentiert [MS24]:** R. H. Christensen. 2019. ordinal – Regression models for ordinal data. R package version 2019.12-10

**Kommentiert [MS25R24]:** Christensen, R. H. B. (2020). ordinal: Regression Models for Ordinal Data. R package version 2019.12-10. https://cran.r-project.org/package=ordinal

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 (cf. Taylor et al. 2022). 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 (ibid.). Overall, the application of CLMM aligns with the nature of the data and research objectives, offering a robust and tailored framework for exploring the effects of different factors on participants' ratings in the 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.
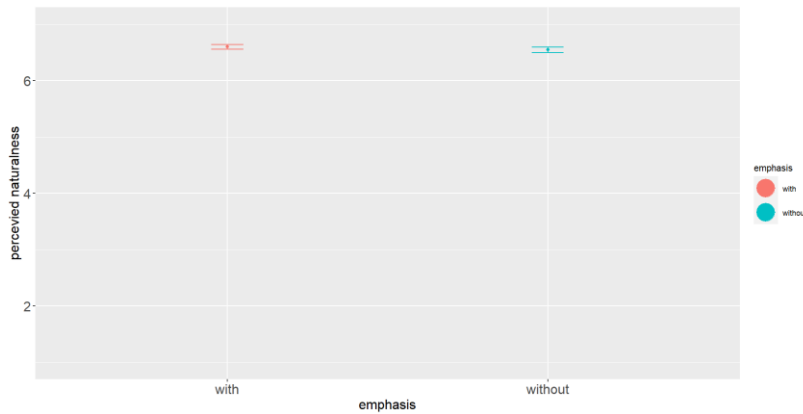
Figure 6: Comparison of participants' ratings of stimuli with and without emphasis

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 auditive and written stimuli. Participants' ratings of auditory stimuli are illustrated in red, whereas those of written stimuli are colored in blue.
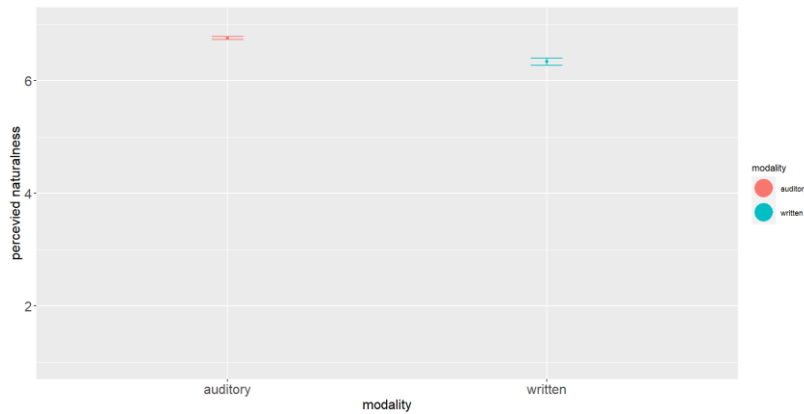
Figure 7: Comparison of participants' ratings of auditory and written stimuli

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.
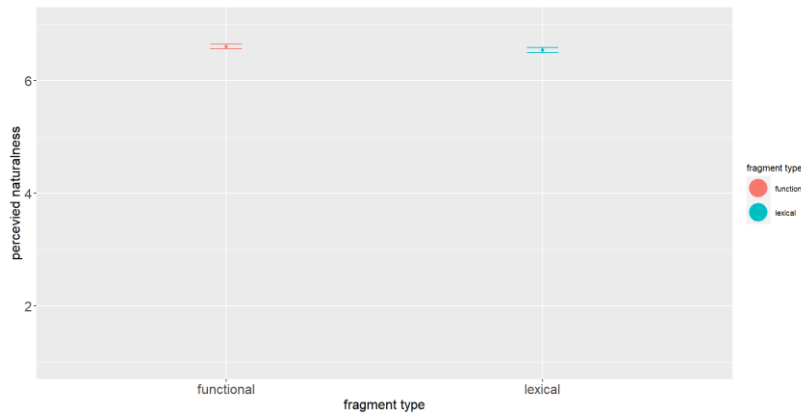


Figure 8: Comparison of participants' ratings of stimuli with functional and lexical fragments.

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, the hypotheses are re-visited, and the reliability of the data analysis is discussed. 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.

Regarding emphasis, the present study showed that dialogues in which the contrasting words are either orthographically or prosodically emphasized are perceived as more natural than if they are not. This allows the conclusion that emphasis positively influences recipient's ability to create a focus-based anaphoric relation.

Furthermore, modality is a significant factor for how fragmentary answers are perceived. That is, fragmentary answers receive higher acceptability ratings if presented in auditory form instead of written form. One can draw the conclusion that this is due to the fact that elliptical utterances occur more often in spoken language than written language and that language structures are rated more natural if they occur in their natural circumstances.

Regarding the 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 function 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.

To ensure a reliable data analysis, it was assessed 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. AIC takes into account both how well the model fits the data and the number of parameters in the model. The goal is to find a model that adequately explains the data while avoiding overfitting, which occurs when a model is too complex and captures noise in the data rather than genuine patterns (cf. Cavanaugh & Neath 2018).

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

The lower AIC and significant likelihood ratio test suggest that the more complex model incorporating predictors such as emphasis, modality, and fragment type, provides a better fit to the data. Furthermore, these predictors exert a notable impact on the model's performance. Despite the statistically significant distinction between the AIC of the null model and the

model encompassing the three predictors, this discrepancy remains relatively minor. This could potentially be attributed to an unaccounted-for effect within the model or the possibility that modality and emphasis, as predictors, possess relatively weak levels of significance.

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 (cf. Delbar 2019). However, as participants for the present study were crowdsourced from Prolific which has a limited number of potential participants, a pre-screening for age and geography was not possible, as the pool of participants would not been sufficient then. Nevertheless, the data shows that the differences in age are negligible.
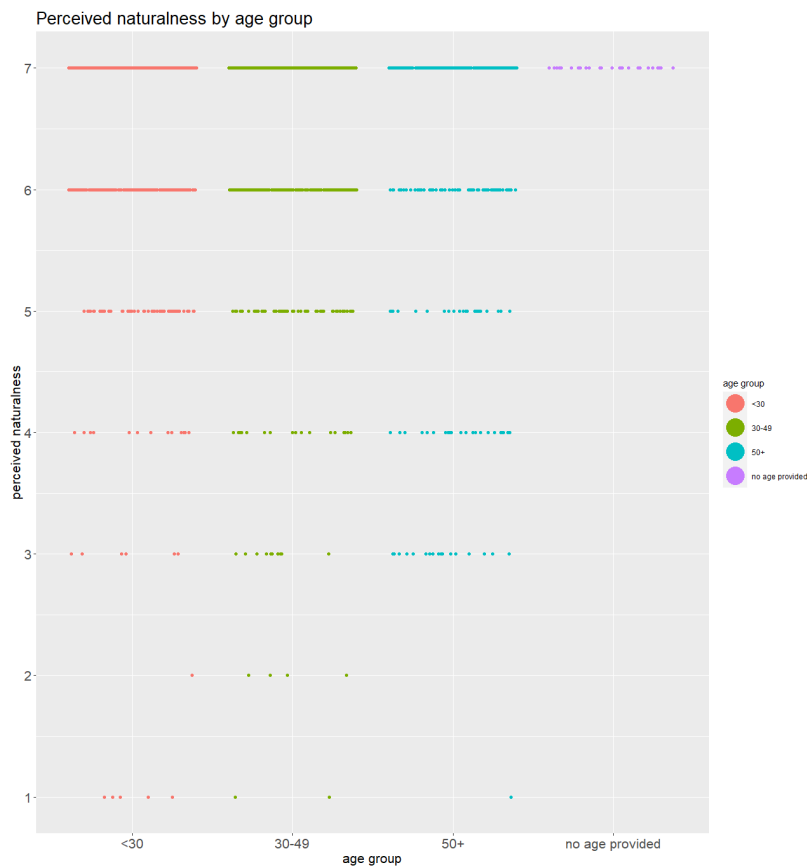
Figure 9: Scatter plot of participants' ratings by age group

As becomes apparent in Figure 9, there are only slight differences between younger and older participants. Much like the findings of Delbar (2019), only minor age-related differences can be observed. Overall, there seems to be a subtle tendency for participants to give lower acceptability ratings as age increases. That is, participants under the age 30 provided an average rating of 6.66. Meanwhile, participants aged between 30 and 49 assigned slightly lower average ratings of 6.54. Similarly, participants aged 50 years or more gave the lowest average rating of 6.43. Nonetheless, the data also indicates a noticeable increase in variability within each age group, as age advances. This, in turn, adds a layer of complexity to making comparisons between these groups. Furthermore, the uneven distribution of participants across age groups further compounds the complexity of intergroup comparisons.

The age breakdown is as follows: There are 42 participants in the group aged below 30, 41 participants aged between 30 and 49, and only 16 participants aged 50 or older.

While the primary focus of this study centers on German native speakers, it's noteworthy that a subset of participants indicated proficiency in languages beyond German. When contrasting the acceptability ratings between monolingual and bilingual German speakers, a notable pattern emerges. That is, overall, bilingual participants tend to provide lower ratings. Specifically, bilingual participants yielded an average rating of 6.43, whereas monolinguals exhibited slightly higher ratings at 6.59.
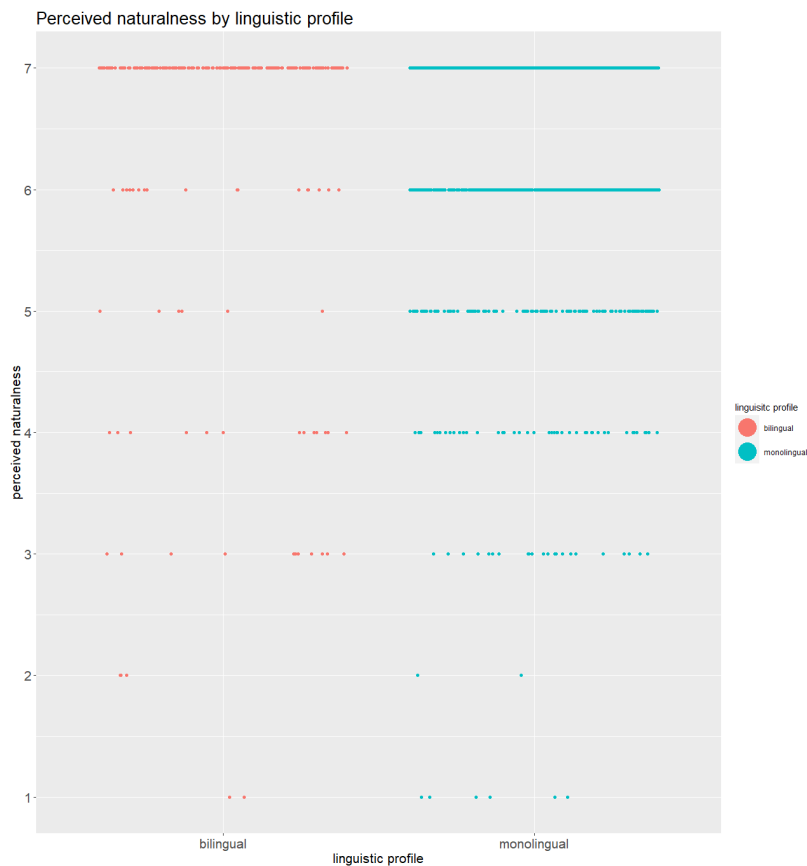
Figure 10: Scatter plot of participants' ratings by linguistic profile

However, it's crucial to exercise caution when drawing broad conclusions from these findings. The higher variability observed in the ratings provided by bilinguals, coupled with their smaller representation in the study necessitates careful consideration. That is, the data set includes only 9 participants compared to the significantly larger group of 91 monolinguals. Further investigation is warranted to explore whether bilingualism truly impacts the perception of fragmentary answers involving contrastive focus.

In terms of regional variations in the acceptance of fragments, the findings of this study should be interpreted within the context of the study's limitations, which restrict the ability to draw comparisons across different regions. The limited number of potential participants available on Prolific has hindered the exploration of potential regional differences. Additionally, the specific setup and methodology employed in this study were designed to assess the effectiveness of conducting acceptability judgment tasks, with a specific focus on factors such as modality, emphasis, and fragment type and have not been configured for capturing nuanced regional variations.

Addressing biases inherent in the study's methodology, it's crucial to recognize several factors that might have influenced participants' responses. Firstly, participants were well aware that their task involved assessing language structures based on their perceived naturalness. While this awareness aligns with the study's objectives, such explicit evaluation might lead to a heightened sensitivity, potentially affecting the participants' judgments.

Furthermore, it is worth noting that participating in a research study inherently deviates from the spontaneous communication that characterizes everyday interactions. The conditions inherent to the research setting may potentially trigger different reactions and alter participants' perception of such linguistic phenomena. This alternation might introduce a level of artificiality that contrasts with natural language use. However, the prominence of ratings assigned with a value of 7, indicative of complete naturalness, lends credence to the inference that, despite the study's controlled framework, participants did not perceive the provided dialogues as inherently unnatural.

This high frequency of ratings indicating a complete naturalness could be attributed to a plausible linguistic phenomenon. That is, the prevalence of occasional linguistic errors in real-life conversations. Native speakers, despite their proficiency, may inadvertently use ungrammatical or unconventional structures in their speech, suggesting that such structures, though not adhering to formal grammar, are still recognizable and accepted as part of natural language use.

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.

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

(own example)

(29)  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), speaker B's 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 what the preposition *until* refers to based on the preceding clause, 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. However, further studies are required to check those hypotheses.

Next, the context might also influence how contrastive, fragmentary answers are perceived. That is, the prominence of the correlate in the discourse possibly impacts how natural the fragmentary answer sound to them. Since no previous context was given when any of the dialogues were presented, it can be ruled out that this affected participants' judgements. Furthermore, all preceding sentences in the critical stimuli had the same syntactic structure and did not include any subordinate clauses. Therefore, it was ensured that all correlates in the preceding sentence had the same degree of prominence in the discourse.

5.2 Comparison with previous studies and theoretical predictions

In this subchapter, the similarities and disparities that emerge between previous research and theoretical conjectures on one hand, and the empirical outcomes of the current investigation on the other are explored.

First, in accordance with the hypotheses drawn from previous research findings outlined in chapter 2.3, the role of modality in identifying correlate-remnant pairings is apparent, evidenced by the higher acceptability ratings assigned to auditory stimuli. Notably, the auditory stimuli encompassed recordings both with and without emphasis on the contrasting words, necessitating a consideration of the role of emphasis. For an in-depth exploration of this aspect, refer to the latter part of this subchapter.

Returning to the theme of modality, it is important to acknowledge that the manner in which the contrasting words were emphasized depended on the modality, thereby potentially impacting the ratings. To elaborate, the natural occurrence of pitch accent on contrasting words, as explained in chapter 2.3, contrasts with the less common practice of emphasizing words through uppercase formatting in written language. This distinction could account for the lower ratings received by both emphasized and non-emphasized stimuli in comparison to their auditory counterparts.

Second, as implied by the CRC, dialogues that lack pitch accent on the contrasting words should be rated as less natural. This theoretical supposition aligns with the empirical results, which reveal that stimuli without emphasis indeed receive lower naturalness ratings, and this difference is statistically significant. Despite the statistical significance, it was believed that the difference between ratings assigned to the stimuli with and without emphasis exhibit a more substantial discrepancy.

One could explain this finding as follows. As has already been mentioned in chapter 2.3, it is established that pitch accent inherently occupies the default position, even in cases of silent reading (cf. Rasekhi & Harris, 2021). Consequently, it stands to reason that the same principle applies to pitch accent on contrasting words. Certainly, the inferred pitch accents present during silent reading cannot be equated with those produced in spoken language. Therefore, it is reasonable to conclude that if intonational patterns are somewhat inherent in written language, it is logical to presume that disparities between modalities exist, although they might not be as substantial as previously believed.

Furthermore, the comparison of auditory and written stimuli underscores the nuanced interaction between emphasis and contrastive focus. Auditory recordings without emphasis still encompass natural intonation, albeit without explicit emphasis on contrasting words. The brevity of the stimuli potentially aids participants in comprehending the underlying contrastive focus even in the absence of explicit emphasis.

Regarding parallelism, it is worth noting that the stimuli exhibited contrasting words that shared both morphological and phonetic characteristics. These encompassed morphological features such as number, gender, case, as well as their prosodic weight. Drawing from the hypothesis on Morphological Parallelism discussed in chapter 2.3, these shared properties contribute to the ease of identifying correlate-remnant pairings. For instance, the case marking ensured the unambiguity of all dialogues, while the similar syntactic structures of the preceding clause and the fragmentary answer further enhanced the comprehensibility of the dialogues. This parallelism could potentially explain the prevalence of high acceptability ratings

observed across the dialogues. However, in order to truly validate the morphological parallelism hypothesis, it is imperative to conduct further research that involves a direct comparison between the outcomes of the current study and dialogues featuring dissimilar correlate-remnant pairings.

Delving into the locality aspect, the design of stimuli intentionally challenges the anticipated correlation between position and accent. By positioning the appropriate correlate away from the final position, which traditionally carries the nuclear sentence accent, the stimuli deviate from the expected norm. Remarkably, despite this intentional violation, acceptability ratings remain consistently high and exhibit minimal variability across conditions. The unexpectedness of these ratings may arise from a complex interplay of linguistic elements and cognitive processing.

Regarding emphasis, as hypothesized, stimuli featuring emphasized contrasting words received higher ratings than those without such emphasis. This trend possibly arises from participants still perceiving a degree of emphasis even in the condition without pitch accent on the correlate. In the written stimuli, on the one hand, participants might have construed it as emphasized due to their mental visualization of a situation and how the dialogue would unfold in such a scenario, although the dialogue lacked explicit emphasis on the contrasting words. In the auditory stimuli, on the other hand, the recordings without emphasis maintained a natural intonation. Here, the emphasis was not directed at the contrasting words but rather at the preverbal position, as exemplified in chapter 3.3. It is plausible that the stimuli sentences were short enough to be understood. That is, participants could understand what the contrastive focus relates to in the preceding clause, without needing the emphasis on the contrasting words. This could provide an explanation for why the ratings of the stimuli that lack emphasis exhibit a statistically significant decrease compared to those featuring emphasis. However, even with this decrease, the ratings for the non-emphasized stimuli do not show a significant drop when compared to the ratings of the emphasized stimuli.

Conducting a comparison of the time participants took to provide their ratings could offer insights into whether understanding the dialogues without emphasis required more time compared to those with emphasis. Extended

reading durations might imply that the absence of pitch accent on the contrasting words makes it comparatively more challenging to establish correlate-remnant relations. However, interpreting the outcomes of studies that rely on reaction times presents a challenge. Longer reaction times can arise from various factors, making the interpretation of cognitive processes less straightforward (cf. Schütze, 2016).

As a last point, the study delved into the potential influence of semantics on the perception of fragmentary answers arising from contrasting words. This investigation aimed to discern whether distinctions between lexical and functional fragments impact their perceived naturalness among native speakers. As explained in chapter 2.3, it is commonly assumed that lexical terms hold a more significant role in conveying meaning and are more likely to carry stress in speech compared to functional words. However, surprisingly, the findings of the current study challenge this convention. The results suggest that contrastive, functional fragments are perceived as more natural in comparison to their lexical counterparts. This outcome could be attributed to the study design, in which prepositions were chosen to represent functional words. In most of the dialogues, those prepositions informed the hearer or reader about specific times or locations. As the dialogues lacked contextual cues, participants might have inferred that details regarding time or location were more important in the conversation than the information given by the lexical words. Therefore, the functional words seemed more prominent in the discourse than the lexical contrasting words. Nevertheless, it's crucial to note that this interpretation remains speculative, necessitating further research to corroborate these findings and delve deeper into the underlying dynamics.

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

The present study

- Explored the factors that influence how native speakers perceive fragments in dialogues involving contrastive focus in German
- Therefore, the first study that empirically tests the hypotheses regarding the pitch accent placement in dialogues involving contrastive focus. Hence, pathing the way for future research that uses acceptability judgment tasks in this field
- Also adds to the existing body of research in this field by exploring contrasting fragments in German, therefore adding to the existing body of research by testing previous hypotheses and previous research findings on an additional language
- Moreover, as outlined in chapter 2.4, there is a need for formally obtained data from acceptability judgement tasks that are line with scientific research standards and fulfill the requirements for reliable research.
- Also adds to the existing body of research on the importance of orthographic marking. While this method is already established and used in other fields of linguistics, this study is the first to explore the role of orthography in the comprehension of contrastive focus
- Results indicated that contrastive fragments seem more natural if the contrasting words are either prosodically or orthographically emphasized, if the dialogues are presented as auditory recordings rather than as written dialogues, and if the answers in the dialogues contain functional contrastive fragments instead of lexical contrastive fragments.
- Results confirm previous findings saying that pitch accent on contrasting words is needed

- Present study gives new insights into the processing of functional fragments as the results suggests that functional words play a more crucial role than has been believed so far. That is, while lexical words were assumed to be crucial for conveying messages, the present study indicates that functional fragments are more prominent, therefore are easier to comprehend, indicated by the higher acceptability ratings for such dialogues, compared to the ones with lexical fragments.
- Study gives insights into the role of the medium in which the dialogues are presented. While conducting acceptability judgement tasks containing recordings instead of written copies of the dialogues requires more resources, the results of the present study highlights the need for that and why studies using auditory stimuli deliver much more reliable results in linguistics.
- Results somewhat confirm the hypothesis regarding the parallelism of both clauses, insofar elliptical structures that are morphologically and syntactically very similar were easier to the preceding sentence received high acceptability ratings. However, the present study did not investigate dissimilar structure and therefore could not test the parallelism hypothesis

The present study delves into the factors affecting how native speakers perceive fragments in dialogues featuring contrastive focus in German. This pioneering investigation represents the first empirical examination of hypotheses concerning the placement of pitch accents in dialogues with contrastive focus, setting the stage for future research employing acceptability judgment tasks in this domain. It not only contributes to the existing body of cross-linguistic research by exploring contrasting fragments in German but also adds to the pool of knowledge by validating earlier hypotheses and research findings in the context of an additional language.

As discussed in chapter 2.4, previous instances of acceptability judgment tasks have often relied on informal approaches. Consequently, a need

arises for formally gathered data through acceptability judgment tasks that meet the criteria essential for scientific research. This study addresses this demand by recruiting 100 participants, ensuring robust statistical power, and employing a 7-point Likert scale to encompass an optimal array of response options. In doing so, this study diligently aligns with the prerequisites established for conducting structured and formal acceptability judgment tasks.

Moreover, the study contributes to the existing research on the significance of orthographic marking. While this approach is already established and utilized in other linguistic fields, this study stands out as the first to investigate the role of orthography in comprehending contrastive focus.

The findings revealed that contrastive fragments are perceived as more natural when the contrasting words are emphasized through prosody or orthography. Furthermore, presenting dialogues as auditory recordings, rather than written text, and including functional contrastive fragments in responses, as opposed to lexical ones, enhanced the naturalness of contrastive fragments. These outcomes validate prior assertions that pitch accents on contrasting words are indeed necessary.

The study introduces fresh insights into the processing of functional fragments. Results indicate that functional words play a more pivotal role than previously believed. While lexical words were traditionally considered vital for conveying messages, this study highlights the prominence of functional fragments. This is evidenced by higher acceptability ratings for dialogues containing functional fragments compared to those with lexical fragments.

Furthermore, the study sheds light on the impact of the medium used to present dialogues. While conducting acceptability judgment tasks with auditory recordings demands more resources, the study underscores the necessity of this approach and its capacity to yield more reliable results in linguistics.

Finally, the results partially support the hypothesis regarding clause parallelism. Elliptical structures with morphological and syntactic similarity to the preceding sentence garnered higher acceptability ratings. However, it's important to note that dissimilar structures were not investigated in this study, and thus, the parallelism hypothesis couldn't be fully tested.

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
- Only contrastive fragments, not necessarily general findings for all sorts of correlate-remnant pairings
- Only limited research on functional words, as only three different contrasting preposition pairs were researched
- There might be confounding factors that influenced the results and weren't accounted for in the statistical analysis. One might be age differences in the acceptability ratings and another one might be regional varieties. As age and dialect differences weren't the focus of the present study, it just shows that further research is needed.

6.4 Suggestions for future research

X

What aspects need further research?

Discuss the value of this pilot study for future research

- Results did not confirm the hypothesis that lexical words are more prominent in a discourse and that contrastive lexical fragments are therefore easier to process. Instead, functional fragments received higher acceptability ratings. This is somewhat contradicting to previous research and makes future research that investigates the differences between lexical and functional words in fragmentary answers necessary
- As hypothesized, dialogues in which the correlate and remnant bore pitch accent were rated as more natural than those in which the correlate was not prosodically emphasized. However, the latter dialogues still received relatively high acceptability ratings. Therefore, future research should investigate in more depth whether it is required to emphasize both contrasting words or whether might be sufficient to only have pitch accent on the remnant, as described in the CRC, to identify correlate-remnant pairings.
- Results indicate that bilingualism might influence the acceptability judgement ratings. Further research is needed to explore whether having more than one native language impacts the perception of what language structures seem natural.
- Further investigations should explore whether it is important that there is a complete prepositional phrase in the fragmentary answer is required or whether native speakers could also comprehend fragmentary answers if they only consist of the contrastive preposition.
- All in all, more research is needed but the present study already gives insights into the comprehension of contrastive fragments in German

7. References

Akinnaso, F. N. (1982). On The Differences Between Spoken and Written Language. *Language and Speech*, *25*(2), 97–125. https://doi.org/10.1177/002383098202500201

Beckman, M. E., & Ayers. (1997). *Guidelines for ToBI labelling, version 3*. The Ohio State University Research Foundation. http://www.ling.ohio-state.edu/phonetics/ToBI/ToBI.0.html

Carlson, K., Frazier, L., & Clifton, C. (2009). How prosody constrains comprehension: A limited effect of prosodic packaging. *Lingua*, *119*(7), 1066–1082. https://doi.org/10.1016/j.lingua.2008.11.003

Chomsky, N., & Halle, M. (1991). *The sound pattern of English* (1st MIT Press paperback. ed). MIT Press.

Delbar, N. A. (2019). *Swiping in English and Dutch: The Interaction between R-Pronouns and Modal Particles*.

Featherston, S. (2008). Thermometer judgements as linguistic evidence. In C. M. Riehl & A. Rothe (Eds.), *Was ist linguistische Evidenz?* Shaker Verlag.

Féry, C. (2011). German sentence accents and embedded prosodic phrases. *Lingua*, *121*(13), 1906–1922. https://doi.org/10.1016/j.lingua.2011.07.005

Féry, C., & Herbst, L. (2004). German Sentence Accent Revisited. In S. Ishihara, M. Schmitz, & A. Schwarz (Eds.), *Working Papers of the SFB632, Interdisciplinary Studies on Information Structure (ISIS) 1* (pp. 43–75). Universitätsverlag Potsdam.

Franck, J., Bowers, J., Frauenfelder, U. H., & Vigliocco, G. (2003). Orthographic influences on agreement: A case for modality-specific form effects on grammatical encoding. *Language and Cognitive Processes*, *18*(1), 61–79. https://doi.org/10.1080/01690960143000452

Frazier, L., & Clifton, C. (1998). Comprehension of Sluiced Sentences. *Language and Cognitive Processes*, *13*(4), 499–520. https://doi.org/10.1080/016909698386474

Frazier, L., Taft, L., Roeper, T., Clifton, C., & Ehrlich, K. (1984). Parallel structure: A source of facilitation in sentence comprehension.

*Memory & Cognition*, *12*(5), 421–430. https://doi.org/10.3758/BF03198303

Griffiths, J., Güneş, G., & Lipták, A. (2023). Reprise fragments in English and Hungarian: Further support for an in-situ Q-equivalence approach to clausal ellipsis. *Language*, *99*(1), 154–191. https://doi.org/10.1353/lan.2023.0000

Griffiths, J., & Lipták, A. (2014). Contrast and Island Sensitivity in Clausal Ellipsis. *Syntax*, *17*(3), 189–234. https://doi.org/10.1111/synt.12018

Harris, J. A. (2015). Structure Modulates Similarity-Based Interference in Sluicing: An Eye Tracking study. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01839

Harris, J. A., & Carlson, K. (2016). Keep it local (and final): Remnant preferences in "let alone" ellipsis. *Quarterly Journal of Experimental Psychology*, *69*(7), 1278–1301. https://doi.org/10.1080/17470218.2015.1062526

Harris, J. A., & Carlson, K. (2018). Information Structure Preferences in Focus-Sensitive Ellipsis: How Defaults Persist. *Language and Speech*, *61*(3), 480–512. https://doi.org/10.1177/0023830917737110

Jasso, J. (2022). *How Does Parent Input Influence Bilingual Children's Knowledge and Use of Spanish Subjunctive? A Dyadic Study* [Dissertation]. University of Texas at Austin.

Juzek, T. S. (2016). *Acceptability Judgement Tasks and Grammatical Theory*. University of Oxford.

Kayali, N. (2023). "Does this make sense?": The effect of matching guise in regional accent on grammatical acceptability judgments. *Proceedings of the Linguistic Society of America*, *8*. https://doi.org/10.3765/plsa.v8i1.5525

Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, *55*(3–4), 243–276. https://doi.org/10.1556/ALing.55.2008.3-4.2

Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge University Press.

Lobeck, A. C. (1995). *Ellipsis: Functional heads, licensing, and identification.* Oxford University Press.

Merchant, J. (2005). Fragments and ellipsis. *Linguistics and Philosophy*, *27*(6), 661–738. https://doi.org/10.1007/s10988-005-7378-3

Merchant, J. (2019). Ellipsis: A survey of analytical approaches. In J. Van Craenenbroeck & T. Temmerman (Eds.), *The Oxford Handbook of Ellipsis* (1st ed., pp. 19–45). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198712398.013.2

Meurers, D., Ziai, R., Amaral, L., Boyd, A., Dimitrov, A., Metcalf, V., & Ott, N. (2010). Enhancing Authentic Web Pages for Language Learners. *Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications*.

Opitz, A., & Bordag, D. (2022). THE IMPACT OF ORTHOGRAPHY ON LEXICAL ACCESS: THE CASE OF CAPITALIZATION AND WORD CATEGORY INFORMATION IN L1 AND L2 GERMAN. *Studies in Second Language Acquisition*, *44*(4), 1194–1209. https://doi.org/10.1017/S0272263121000711

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Phillips, C., & Parker, D. (2014). The psycholinguistics of ellipsis. *Lingua*, *151*, 78–95. https://doi.org/10.1016/j.lingua.2013.10.003

Rasekhi, V., & Harris, J. A. (2021). Resolving ambiguous polarity stripping ellipsis structures in Persian. *Glossa: A Journal of General Linguistics*, *6*(1), 1–31. https://doi.org/10.16995/glossa.5881

Schütze, C. T. (2016). The empirical base of linguistics: Grammaticality judgments and linguistic methodology [Application/pdf]. *Classics in Linguistics*, 1.01 MB. https://doi.org/10.17169/LANGSCI.B89.100

Sedarous, Y., & Namboodiripad, S. (2020). Using audio stimuli in acceptability judgment experiments. *Language and Linguistics Compass*, *14*(8), 1–21. https://doi.org/10.1111/lnc3.12377

Sprouse, J., Schütze, C. T., & Almeida, D. (2013). A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001–2010. *Lingua*, *134*, 219–248. https://doi.org/10.1016/j.lingua.2013.07.002

Wagner, M. (2012). Focus and givenness: A unified approach. In I. Kučerová & A. Neeleman (Eds.), *Contrasts and Positions in Information Structure* (1st ed., pp. 102–147). Cambridge University Press. https://doi.org/10.1017/CBO9780511740084.007

Wagner, P. S. (1999). The synthesis of German contrastive focus. *Proceedings of the 14th ICPhS*, 1529–1532.

Wierzba, M., Brown, J. M. M., & Fanselow, G. (2023). The syntactic flexibility of German and English idioms: Evidence from acceptability rating experiments. *Journal of Linguistics*, 1–38. https://doi.org/10.1017/S0022226723000105

Winkler, S. (2019). Ellipsis and Prosody. In J. Van Craenenbroeck & T. Temmerman (Eds.), *The Oxford Handbook of Ellipsis* (1st ed., pp. 357–386). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198712398.013.15

# 8. Appendix

## 8.1 Abbreviations, symbols and other notational conventions

| | |
|---|---|
| ? | questionable/marginal acceptability |
| # | infelicitous |
| * | ungrammatical |
| $_1 \ldots t_1$ | syntactic movement |
| _ | omitted linguistic material |
| $XP^i \ldots YP^i$ | coreference |
| // | intonational phrase boundary |
| UPPERCASE | emphasis (orthographical marking or pitch accent) |
| $[\ldots]_F$ | focused position |
| ⟦…⟧ | denotation/semantic representation |
| L+H* | pitch accent |
| A, B | speaker |
| ACC | accusative |
| AJT | acceptability judgment tasks |
| A-movement | argument movement |
| CLMM | Cumulative Link Mixed Models |
| CRC | Contrastive Remnant Condition |
| DAT | dative |
| DP | determiner phrase |
| iff | if and only if |
| MaxQUD | maximal QUD |
| NOM | nominative |
| p-omission | preposition-omission |
| p-stranding | preposition-stranding |
| QUD | Question Under Discussion |
| SVO | subject-verb-object |
| SQA | syntactic question approach |

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 vorge-
stellt.
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.

(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 [MS27]:** Link to github, however repository is private! Publish repository?

**Kommentiert [MS28]:** 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 ge
　　　kündigt.

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 23rd, 2023

*Miriam Schiele*

Miriam Schiele