



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/210746/>

Version: Published Version

Article:

Roessig, Simon (2024) The inverse relation of pre-nuclear and nuclear prominences in German. *Laboratory Phonology*. pp. 1-43. ISSN: 1868-6354

<https://doi.org/10.16995/labphon.9993>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



The inverse relation of pre-nuclear and nuclear prominences in German

Simon Roessig, Department of Language and Linguistic Science, University of York, York, UK, simon.roessig@york.ac.uk

The importance of pre-nuclear prominences for focus marking has been largely neglected. Recent studies, however, present first evidence that the prosody of the pre-nuclear region indicates its status as part of a broad focus or as pre-focal. This study presents a systematic investigation of the pre-nuclear domain and its relation to the nuclear accent in German. The results show that the realization of the pre-nuclear domain indeed depends on whether it is focal or pre-focal: The pre-nuclear noun is characterized by larger F₀ excursions, higher F₀ maxima, and longer durations when it is in broad focus than when it precedes a narrow focus. Furthermore, the realization of a pre-focal, pre-nuclear domain depends on the following focus: The pre-nuclear noun is produced with smaller F₀ excursions, lower F₀ maxima, and shorter durations before a corrective focus than before a non-corrective narrow focus. The comparison to the nuclear accent suggests an inverse relationship, i.e., the pre-nuclear domain becomes less prominent (i.e., realized with smaller F₀ excursions, lower F₀ maxima and shorter durations) when the nuclear domain becomes more prominent. The findings suggest that the phonetic manifestation of information structure is distributed over larger prosodic domains and entails the modulation of phrasal prominence profiles rather than just local adjustments of the nucleus.



1 Introduction

Structuring or packaging of information is a very important task in communication that speakers perform to align their message with the “hearer’s mental model of the current conversation” (Vallduví & Engdahl, 2013, p. 19; see also Chafe, 1976; Lambrecht, 1994; Prince, 1981; Krifka, 2008). The ways in which speakers structure information vary cross-linguistically, with prosody being the most important means in West-Germanic languages. Most research on the prosodic marking of information structure has been concerned with prominences in the focused part of a sentence and the following post-focal part. Fewer works have looked at the region preceding the focus and its importance has been widely neglected (Büring, 2016). The current paper broadens the perspective on prosodic focus marking in German by taking the focal *and* the pre-focal region into account. It shows that there are systematic effects on pre-nuclear words conditional on focus breadth (broad focus vs. non-corrective narrow focus) and use (non-corrective narrow focus vs. corrective narrow focus). The results suggest that focus marking is distributed over the phrase and focus structures are partially encoded in the relationship between the pre-nuclear and nuclear part. In this introduction, I attempt to introduce the basic concepts of focus and its diverse manifestations that are relevant for the study, and to collect what is known about prosodic effects of focus marking in the pre-nuclear domain.

1.1 Focus, focus breadth, and contrastiveness

The focus of a sentence is said to contribute significant information to the listener’s knowledge and beliefs (Vallduví & Engdahl, 1996). The focus hence contains parts of an utterance that the speaker rates as most important or informative (Halliday, 1967) or that are “unpredictable or pragmatically non-recoverable” (Lambrecht, 1994, p. 207) in the discourse context. Consider example (1): The question Q sets up a context that triggers an answer A of the form “x bought a camera”. The identity of x is needed to differentiate the intended meaning from possible alternatives (e.g., *Kim bought a camera*, *Lee bought a camera*, etc.) The focus of the answer, *Bill*, contributes the important information relating directly to the question and identifies the individual x (Stevens, 2017). The example adapts the widespread convention to indicate focus by square brackets with subscript F (Jackendoff, 1972). The main prominence, or nuclear accent, falls on the focused word *Bill*. In the example this is expressed by the use of capitals. The remainder of A, *bought a camera*, constitutes the background—the part that “anchors the sentence to the previous discourse” (Vallduví & Engdahl, 1996, p. 461). In this case, the background, being post-focal, usually receives no accent.

- (1) Q: Who bought a camera?
A: [BILL]_F bought a camera.

Focus can be characterized by taking various aspects into account. One important aspect is *focus breadth* referring to the size of the focus (Gussenhoven, 2007; Ladd, 1980). The focus can be narrow as in example (1), so that only one constituent is in focus (Breen, Fedorenko, Wagner, & Gibson, 2010) or it can be broader so that it encompasses the whole sentence or the predicate (Lambrecht, 2000), as illustrated in examples (2) and (3) (adapted from Lambrecht, 2000). In (2), the whole sentence is in focus (“sentence focus” following Lambrecht, 2000) while in (3) the predicate is in focus (“predicate focus” following Lambrecht, 2000). In this situation, where the nuclear accent marks a focus that is larger than the nuclear-accented word, we often speak of *focus projection* (Selkirk, 1995). The exact location of the nuclear accent within the focus is determined by language-specific rules (Büring, 2016; Gussenhoven, 2007; Ladd, 2008). In both example (2) and (3), the nuclear accent falls on the last argument which is a typical location in English and German.¹

- (2) Q: What happened?
A: [Mary had an ACCident.]_F
- (3) Q: Why didn't Mary come to work today?
A: She [had an ACCident.]_F

Another important aspect to characterize focus is the notion of *contrastiveness*. Contrastiveness is closely related to the presence of alternatives (Repp, 2016; Rooth, 1992). Some theories regard focus as contrastive when an overt or explicit alternative or set of alternatives is present in the context (Halliday, 1967); see example (4) adapted from Büring (2016), in which *Sam* is said to contrast with the overt alternative *Kim* that was mentioned in the sentence before.

- (4) A: The boss gave Kim a raise.
B: No, the boss gave [SAM]_F a raise.

The relation of a focus to alternatives in the discourse is not always as clear as in this example. Furthermore, the term *contrastive focus* is connected to diverse phenomena and definitions vary widely (Repp, 2016; Zimmermann, 2008). For example, overt alternatives are not considered necessary by all theories (e.g., Rooth, 1992, 2016; Krifka, 2008). Even if there is no consensus about contrastive focus, *corrections* as illustrated in example (4) seem to be widely accepted as cases of contrastive focus (Riester & Baumann, 2013). Repp (2016) hypothesizes that contrastiveness may be a gradable phenomenon (see also Calhoun, 2009). In this view, a corrective focus as in example (4) involves a large degree of contrastiveness while the focus in example (1) may be characterized by a lower degree of contrastiveness. It should be noted that corrective focus may be associated with intonational tunes different from other cases of contrastive focus.

¹ See also Röhr (2016, chapter 3) for an overview of the debate around default accent placement.

The discussion so far has shown focus in various manifestations, differing in focus breadth and whether focus is used to correct (or reject) an explicit alternative preceding. While these manifestations are often labelled *focus types*, the term is not uncontroversial, and other related concepts like “uses of focusing” (Büring, 2007, p. 454) or “focus meanings” (Gussenhoven, 2007, p. 90) have been introduced. In this study, I compare sentences with three different focus structures. They differ in focus breadth (broad focus vs. non-corrective narrow focus) as well as focus use (non-corrective narrow focus vs. corrective narrow focus).

1.2 Focus and (the nuclear) accent

Nuclear accent location is a strong prosodic correlate of focus in West-Germanic languages. As demonstrated in (1), a narrow focus on *Bill* in the sentence *Bill bought a camera* will attract the nuclear accent, the last accent in the phrase. This accent placement is different from the pattern found when the sentence is in broad focus, in which case the nuclear accent will fall on *camera*. The broad focus accent placement has been discussed under the term of neutral or normal accentuation (Ladd, 1980). Nevertheless, the nuclear accent on *camera* does not preclude other focus readings, one of which being the narrow focusing of *camera* (*What did Bill buy? – Bill bought a CAMera*). Thus, with the nuclear accent in a certain position, a sentence is often ambiguous as to different focus readings. For an illustration of this point, consider the three examples in **Table 1** that are based on Halliday (1967) and Ladd (1980). The focus structure in the first case (i) can be described as a broad focus. The other two (ii and iii) have a narrow focus comprising (*the*) *shed*. Additionally, in (iii), the focus involves a correction (discarding the alternative *fence*). Henceforth, I will refer to constructions like (ii) as narrow focus and to constructions like (iii) as corrective focus. Crucially, in all three cases, the nuclear accent is placed on the same word (i.e., on *shed*). With the nuclear accent on *shed*, as Ladd (1980, p. 74) puts it, “the focus can be *the shed*, or *painted the shed* or the whole sentence”.²

	Question	Answer	Focus structure
(i)	What’s new?	[John painted the SHED yesterday.] _F	Broad
(ii)	What did John paint yesterday?	John painted [the SHED] _F yesterday.	Narrow
(iii)	Did John paint the fence yesterday?	John painted [the SHED] _F yesterday.	Corrective

Table 1: Different focus structures with the same nuclear accent placement.

² But see Jabeen et al. (2021) for results indicating that different (mismatching) positions of the nuclear accent may be accepted by German listeners in broad focus statements (even more than mismatching accent types).

Various studies have demonstrated that while the nuclear accent placement is the same for these focus conditions, this does not imply that the prosodic realizations of these sentences are identical. It has been shown that nuclear-accented words are realized with different prosodic patterns depending on the focus condition. In particular, nuclear-accented words in narrow focus are longer than their counterparts occurring in broad focus (e.g., Baumann, Grice, & Steindamm, 2006; Eady & Cooper, 1986; Kügler, 2008) and exhibit higher intensities (Breen et al., 2010; see also Roessig, Winter, & Mücke, 2022 although here the effect size of intensity was rather small). Additionally, higher F0 maxima and means, larger F0 excursions, and later F0 peak alignments have been found for narrow focus compared to broad focus (Baumann et al., 2006; Breen et al., 2010; Eady & Cooper, 1986; Féry & Kügler, 2008; Grice, Ritter, Niemann, & Roettger, 2017; Roessig et al., 2022). In categorical intonation analyses using transcription systems related to ToBI (Silverman, Beckman, Pitrelli, Ostendorf, Wightman, Price, Pierrehumbert, & Hirschberg, 1992; Beckman, Hirschberg, & Shattuck-Hufnagel, 2005; see also other chapters in Jun, 2005), these F0 differences are reflected in lower proportions of downstepped H* accents (Baumann, Becker, Grice, & Mücke, 2007) and falling accents (Grice et al., 2017) in narrow focus compared to broad focus. Similar differences in the nuclear-accented word have also been attested between narrow and corrective focus: Compared to narrow focus, corrective focus is characterized by longer durations (Kügler, 2008), higher intensities (Breen et al., 2010), as well as larger F0 excursions, and higher F0 peaks with later temporal alignment (Baumann et al., 2006; Grice et al., 2017; Roessig et al., 2022). Despite the differences in the marking, which can be used by listeners, a good deal of the ambiguity between focus readings with the same nuclear accent position may remain (Cangemi, Krüger, & Grice, 2015).

1.3 Focus and pre-nuclear accents – focus prosody beyond focus?

As the review so far has shown, the nuclear accent has gained a lot of attention – regarding its position and its realization. The literature also shows that changing the nucleus position goes hand in hand with changing the prosody after it (e.g., in *BILL bought a camera* from (1) the nuclear accent is on the first word to signal narrow focus, and everything that follows the focus will be deaccented). One strong view, however, is that the pre-nuclear region is not affected by focus breadth and use. The example from Büring (2007, p. 462) in **Figure 1** illustrates this view. The idea is that there is some kind of default prosody for the sentence (probably attributable to broad focus, see above) in which version (i) will be realized. The asterisks represent prominence levels, with one asterisk for secondary or pre-nuclear pitch accents and two asterisks for the primary or nuclear accent.³ When the focus is on *letter* as in sentence (ii), the nuclear prominence

³ This description omits some levels of the metrical grid. In a full account, each syllable would receive one asterisk on the lowest level. Primary word stress would be represented by one asterisk on the next higher level. These levels are irrelevant for the present exposition.

shifts to this word, deleting the accent on *government*. The reduction of prominence in the post-focal domain extends to languages beyond the group of West-Germanic languages (Xu, 2011). It is often described as post-focal deaccentuation (the loss of accents where they are expected) or post-focal compression (the reduction of pitch range, duration, and other phonetic parameters). Empirical investigations show a robust phonetic effect of reduction in the post-focal domain in German and English (Breen et al., 2010; Roessig et al., 2022; but see Mücke & Grice, 2014). Interestingly, it is assumed that “the default prosody is retained pre-focally” (Büiring, 2007, p. 464). Observe that there is no change in prominence assignment on *witness* in (ii) of **Figure 1**. Pre-nuclear accents are in this view considered optional or “ornamental” (Büiring, 2007), and if “mentioned at all, they are often assumed to be [...] not meaning-related” (Riester & Baumann, 2013, p. 213). Calhoun (2010a) draws a more nuanced picture showing that rhythmic factors are more important for pre-nuclear prominences while focus (and other semantic factors) can still exert an influence on them.

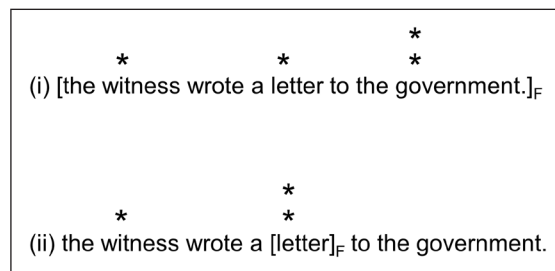


Figure 1: Example illustrating the pre-nuclear default prominence view (Büiring, 2007, p. 462).

The empirical evidence on the status of pre-nuclear accents is indeed somewhat scarce. While some investigations do report no consistent prosodic difference between broad focus and narrow focus in the pre-nuclear domain (Eady & Cooper, 1986; Xu & Xu, 2005, both for English)⁴, there is a growing body of evidence supporting the role of the pre-nuclear prominences for focus marking in West-Germanic languages and beyond. For German, Féry and Kügler (2008) showed that pre-nuclear accents exhibited lower F₀ when they appear pre-focally before a narrow focus compared to pitch accents in the same position in a broad focus sentence (i.e., when they are realized on constituents that are in focus). The results presented in Kügler (2008), again in German, indicate that words in pre-focal position preceding corrective focus are shorter compared to the same words in broad focus sentences. Andreeva, Barry, and Koreman (2017) showed that Bulgarian

⁴ Note that Xu and Xu (2005) found that the peaks of pre-focal words are lower than those of the same words in the broad focus for some speakers, although they conclude that, in general, the pitch range of pre-focus words remains constant.

speakers mark focus by modulating the nuclear accent and reducing the phonetic strength of the prenuclear accent in the narrow focus condition. For several varieties of Arabic, it was reported that the pre-focal region may be characterized by compressed F₀, shorter durations and lower intensities (Alzaidi, Xu, Xu, & Szreder, 2023; Alzamil & Hellmuth, 2021; Chahal & Hellmuth, 2014). Royer and Jun (2019) found that some speakers of Kazan Tatar reduce prominence in the pre-focal domain before narrow focus either by deaccentuation or pitch range compression. Yang and Chen (2020) observed lower F₀ and lower intensities in the pre-focal domain in Mandarin Chinese narrow focus structures compared to broad focus.

Related to the question of whether pre-nuclear prominence contains information as to the breadth of the focus is the question when and how pre-nuclear accents can signal a separate focus – such as a contrastive topic that may be seen as a “focus within a topic” (Braun & Biezma, 2019; Steedman, 2000, 2007). Different accent types have been described for contrastive topics (or themes) and contrastive foci (or rhemes) (Bolinger, 1965; Jackendoff, 1972; Steedman, 2000). The findings of Calhoun (2012) for English indicate that this difference is more likely to be a continuous, phonetic difference (rather than categorical, phonological one, i.e., one of accent type choice). This interpretation is in line with other studies which demonstrate that the phonetic realization of pre-nuclear accents is sensitive to the contrastiveness of a sentence topic: For English, Chodroff and Cole (2018) showed that contrastive topics lead to higher f₀ slopes of the pre-nuclear accent with which they are realized. Similarly, Braun (2006) found higher and later peaks on pre-nuclear words associated with contrastive topics in German. In addition, Calhoun (2012) shows that there is a consistent prominence relation between themes and rhemes in her English data, in that themes are relatively less prominent than rhemes in metrical structure (see also Liberman & Pierrehumbert, 1984; note that in Liberman & Pierrehumbert’s work the two accents are both in separate phrases and thus the pre-nuclear/nuclear distinction does not apply in the sense it is used in the present study).

There is no general agreement on the status of pre-nuclear accents in the perceptual differentiation of focus types. While Gussenhoven (1983) and Welby (2003) find no preference for the presence of a pre-nuclear accent in broad and the absence thereof in narrow focus, Bishop (2017) demonstrates that stimuli without prominence in the pre-nuclear domain are preferred in narrow focus contexts by most listeners, while pre-nuclear prominence is acceptable in broad focus (all three studies on English). In an investigation on Dutch by Rump and Collier (1996), a smaller pre-nuclear peak was judged as optimal before corrective focus, while a larger pre-nuclear peak was judged optimal in broad focus sentences. These last two studies suggest that listeners incorporate pre-nuclear prominences into their judgments at least to some extent, although a clear-cut mapping between absence and presence of pre-nuclear accents to focus conditions may not apply.

The literature review above shows that the widespread idea that pre-nuclear accents are not affected by information structure is questionable and that recent research in experimental phonetics and phonology has begun to shed more light on their role. Based on this evidence, there are good reasons to doubt that the pre-nuclear prosodic pattern of broad focus is retained in narrow focus. A related interesting question is whether the pre-nuclear domain contains information about the following focus when comparing different pre-nuclear realizations that are at the same time pre-focal. In their study on German, Baumann et al. (2007) used speech material that includes narrow and corrective focus. Interestingly, there seemed to be a lower probability for the placement of a pre-nuclear accent before corrective focus than before narrow focus, although in both cases the pre-nuclear domain is in the background (i.e., it is pre-focal in both cases). This lower probability of pre-nuclear accents may reflect the tendency to decrease prominence before corrective focus. This decrease of prominence should be reflected in the phonetic realization of the pre-nuclear domain but has not been investigated systematically. The current paper addresses both phenomena: (1) Pre-nuclear words in focus versus pre-nuclear words in the background (i.e., pre-focal), and (2) pre-focal words before narrow versus pre-focal words before corrective focus. The investigation of these conditions is suited to shed more light on the contribution of the pre-nuclear part to the marking of information structure.

1.4 Research questions

The findings outlined in the previous section suggest that the pre-nuclear region may indeed carry information about the information structure of the sentence. The current study has an exploratory character and investigates the role of the pre-nuclear region with the following research questions:

- (1) Does the prosodic realization of a pre-nuclear word depend on whether it is in focus or pre-focal? More specifically, is the pre-nuclear domain realized differently when it is part of the focus (in a broad focus) than when it is pre-focal before a non-corrective narrow focus?
- (2) Does the following focus type have an influence on the pre-focal pre-nuclear domain? More specifically, is the pre-focal pre-nuclear domain realized differently depending on the following focus type (non-corrective narrow vs. corrective narrow focus)?

The predictions concerning the research questions are the following:

- (1) Pre-nuclear words are realized with more prosodic prominence when they are part of a broad focus than when they are pre-focal before a non-corrective narrow focus (i.e., with higher F0, larger F0 excursions and longer durations).

- (2) Pre-focal pre-nuclear words are realized with more prosodic prominence when they precede a non-corrective narrow focus than when they precede a corrective narrow focus (i.e., with higher F0, larger F0 excursions and longer durations).

To investigate these questions, the study analyzes a data set of German sentences elicited in a controlled way with three different focus structures. Although the main interest of the paper is the pre-nuclear domain, the analyses also take the nuclear domain into account in order to shed more light on the relation of the pre-nuclear and the nuclear parts of the utterance. The motivation for looking at both parts comes from the consideration that prosodic prominence, in general, has been described as a relational property of elements in utterances (Lieberman & Prince, 1977). Thus, the description of the prominence of one element is best done with regard to its syntagmatic relations in the same prosodic domain (in this case: Phrase). Consequently, a comprehensive interpretation of the patterns in the pre-nuclear domain hinges on an understanding of the nuclear domain. The results have the potential to contribute to our understanding of the distribution of focus across the phrase and the importance of prominence relations.

From a methodological point of view, the present investigation combines two analysis paradigms: The data are analyzed using traditional static measures (F0 maximum, F0 excursion, and duration) *and* using time-course analyses. The second element was added since, as Wieling (2018) points out, the frequently employed reduction of dynamic phonetic data carries the risk that “potentially interesting patterns in the dynamic data may be left undiscovered.”

The rest of the paper is structured as follows: Section 2 describes the recordings and measures used in the paper; Section 3 presents the results; Section 4 ends the paper with a discussion.

2 Methods

This study was approved by the Local Ethics Committee of the University of Cologne. Each participant gave written informed consent before study participation. The research was conducted in accordance with the Declaration of Helsinki.

2.1 Speech material

The speech material analyzed in this study consists of sentences produced in three different focus conditions. Question-answer pairs were used to elicit these focus conditions. The answers constitute the analyzed target sentences and follow the form *Er hat den/die <A> auf die gelegt* (‘He put the <A> on the ’) with two nouns A and B. The questions served as triggers for the focus structure of the answer. **Table 2** illustrates the speech material with examples. As in the introduction, square brackets and subscript F are used to indicate the focused elements. Since the focus condition labels refer to the two successive nouns A and B in the sentence, the focus

conditions are given as ordered pairs [focus of A, focus of B]. In [broad, broad], both words are in a broad focus construction. In the other two conditions, [background, narrow] and [background, corrective], word A is in the background and thus occurs pre-focally. The difference lies in the focus type of the focal word B: In [background, narrow], word B is in narrow focus. In the second focus condition, [background, corrective], word B is in corrective focus. In all three conditions, word B receives the nuclear accent while word A receives the pre-nuclear accent (if the sentence is produced in one phrase). Therefore, word A will be referred to as pre-nuclear and word B as nuclear in the following.

Focus condition	Example	Status of pre-nuclear word
[broad, broad]	<p><u>Question:</u> <i>Was hat er gemacht?</i> 'What did he do?'</p> <p><u>Answer:</u> <i>Er hat [den Hammer auf die Wohse gelegt.]_F</i> 'He put the hammer on the Wohse.'</p>	in focus
[background, narrow]	<p><u>Question:</u> <i>Wo hat er den Hammer hingelegt?</i> 'Where did he put the hammer?'</p> <p><u>Answer:</u> <i>Er hat den Hammer [auf die Wohse]_F gelegt.</i> 'He put the hammer on the Wohse.'</p>	pre-focal
[background, corrective]	<p><u>Question:</u> <i>Hat er den Hammer auf die Mahse gelegt?</i> 'Did he put the hammer on the Mahse?'</p> <p><u>Answer:</u> <i>Er hat den Hammer auf die [Wohse]_F gelegt.</i> 'He put the hammer on the Wohse.'</p>	pre-focal

Table 2: Example speech material used in the three focus conditions.

The question to elicit [broad, broad] was *Was hat er gemacht?* ('What did he do?'). The question to elicit [background, narrow] followed the scheme *Wo hat er den/die <A> hingelegt?* ('Where did he put the <A>?'). The question to elicit [background, corrective] followed the scheme *Hat er den/die <A> auf die <C> gelegt?* ('Did he put the <A> on the <C>?') where C is a contrasting alternative referent that gets corrected in the following answer.

The full data set comprises the additional focus condition [corrective, background] that is not reported in the current paper. The reason for excluding this condition is that the scope of the

current paper is the comparison of the pre-nuclear (word A) and the nuclear region (word B). In this condition, however, word A receives the nuclear accent while word B is in the post-nuclear region and gets deaccented.

As targets for the pre-nuclear word (word A), 10 German disyllabic nouns denoting common tools with stress on the first syllable were used: *Amboss* ('anvil'), *Besen* ('broom'), *Bohrer* ('drill'), *Bürste* ('scrub brush'), *Hammer* ('hammer'), *Pinsel* ('paint brush'), *Rolle* ('paint roller'), *Säge* ('saw'), *Schere* ('scissors'), and *Zange* ('pliers'). As targets for word B, 20 German sounding disyllabic nonce words with a $C_1V_2.C_2V_2$ structure were created. All nonce words had stress on the first syllable. C_1 was chosen from the set of $\{/n/, /m/, /b/, /l/, /v/\}$, V_1 from $\{/a:/, /o:/\}$, and C_2 from $\{/n/, /m/, /z/, /l/, /v/\}$. V_2 was always $/ə/$. Examples of the nuclear target word (B) are *Nahne* $/na:nə/$, *Mohme* $/mø:mə/$, and *Bahwe* $/ba:və/$. To construct the target sentences, each of the 10 target words in the pre-nuclear position (A) was paired with a target word in the nuclear position (B). Because there were only 10 different target words in the pre-nuclear position, these target words were used twice. This procedure yielded 20 unique target sentences. Including real words in pre-nuclear position and nonce words in nuclear position induces a certain asymmetry in the data. Note that this is because the data set was originally designed to investigate effects on the nuclear word only by using electromagnetic articulography.

2.2 Speakers and recordings

Twenty-seven monolingual native speakers of German were recorded (19-35 years.; 17 of them identified as female, 10 as male). The recordings were conducted at the University of Cologne using a head-mounted condenser microphone. In addition to the acoustic signal, the articulators' movements were recorded using electromagnetic articulography (EMA). This paper only deals with the acoustic data (see Roessig et al., 2022 for an analysis of articulatory parameters). Speaking with EMA sensors can sometimes be perceived as challenging by some speakers. All speakers in this study were able to speak without problems after a training phase.

2.3 Procedure

The participants were prompted to produce the target utterances by involving them in an interactive game on a computer screen. In the game, their task was to help an animated robot retrieve tools. The robot's questions served as triggers for the focus structure of the answer. The questions were recordings of a male native speaker of German (a trained phonetician). The game was developed as an animated browser app. During the recording session, the experimenter sat behind the participant and controlled the flow of the trials with a keyboard.

Each target word in the nuclear position was associated with a fictitious visual object and each target word in the pre-nuclear position was a common tool. In the preparation phase, the

participants were presented with all target words and read the words aloud with the determiner (e.g., “die Nohme” /di: 'no:mə/). A training session with the same focus conditions but different target words preceded the actual recording session (16 trials).

In the experiment, each unique target sentence was produced in all focus conditions without repetition. Thus, each participant was presented with each focus condition 20 times. The trial order was randomized for each participant with the following constraints: Subsequent trials were not allowed to contain the same target words. Only 15% of the trials were allowed to have the same focus condition in the following trial. This constraint was included since it was not possible to alternate between trials with different focus structure *and* different target words. Three subsequent trials with the same focus condition were not allowed. Between trials, a pause of four seconds was included to make sure that the focus structure of the target sentence referred to the current trial only. Every participant received a different randomization but all participants were presented with the same stimuli.

2.4 Annotations, measurements, and data processing

The boundaries of the two target words of a phrase (pre-nuclear and nuclear) and their stressed syllables were annotated by hand. Additional segmental annotations were obtained from forced alignment using Kaldi (Povey, Ghoshal, Boulianne, Burget, Glembek, Goel, Hannemann, Motlicek, Qian, Schwarz, Silovsky, Stemmer, & Vesely, 2011) through the Montreal Forced Aligner (McAuliffe, Socolof, Mihuc, & Wagner, 2017). Additional annotations were used to get the start and end of the phrase (i.e., start of the first word, end of the last word). Furthermore, the low boundary tone at the end of each sentence was labelled, referred to as L-% in the following. An example TextGrid is shown in **Figure A.1** in the appendix.

Using these annotations, the durations of the lexically stressed syllables of the target words in pre-nuclear and nuclear position (word A and B) were extracted. In addition, F0 was calculated over the whole sentence using Praat (Boersma & Weenink, 2001) through the Python interface parselmouth (Jadoul, Thompson, & de Boer, 2018). For each speaker, the floor for the F0 calculation in Praat was set separately as the F0 value of the lowest L-% boundary tone in all productions of that speaker minus 10 Hz. The ceiling parameter was held constant at 500 Hz. With the obtained pitch track, the following measures were obtained: F0 excursion was calculated locally as the difference between the F0 maximum and the F0 minimum in a word expressed in semitones. In addition, F0 maximum in each target word was calculated in semitones relative to the 5th percentile of the distribution of all L-% boundary tones of each speaker. In addition, the differences of the nuclear and the pre-nuclear target words with respect to the F0 maximum, F0 excursion, and stressed syllable duration within one phrase are analyzed. In each phrase, the value for each measure of the pre-nuclear word is subtracted from that of the nuclear word in the same utterance (nuclear minus pre-nuclear).

F0 trajectories over the intervals of the two target words (A and B) and the whole sentence were calculated. The trajectories are time-normalized in 49 equal time steps over the words in positions A and B, and in 149 equal time steps over the whole sentence. The values in the time-normalized trajectories are again calculated in semitones relative to the 5th percentile of the distribution of all L-% boundary tones of each speaker.

To summarize, the following phonetic quantities are analyzed:

- (1) F0 maximum in pre-nuclear and nuclear target words
- (2) F0 excursion in pre-nuclear and nuclear target words
- (3) Duration of the lexically stressed syllable in pre-nuclear and nuclear target words
- (4) Difference in F0 maximum between nuclear and pre-nuclear target words
- (5) Difference in F0 excursion between nuclear and pre-nuclear target words
- (6) Difference in stressed syllable duration between nuclear and pre-nuclear target words
- (7) Time-normalized F0 trajectory over whole sentence
- (8) Time-normalized F0 trajectories over prenuclear and nuclear target words

2.5 Data exclusion and number of data points

1601 complete recordings could be obtained. Productions that had a clear phrase boundary between word A and word B were excluded to ensure that word A was always pre-nuclear. This exclusion was based on a perceptual judgement following the GToBI annotation scheme (Grice et al., 2005). After this exclusion, the data set comprised 1460 recordings. This means that 8.8% (141 productions) of the data had to be excluded due to phrase boundaries. The exclusions due to phrase boundaries were distributed across the focus conditions as follows: 58 productions (11%) of [broad, broad], 52 productions (9.7%) of [background, narrow], and 31 productions (5.8%) of [background, corrective] had to be excluded.

Visual inspection of the distributions of F0 maximum and F0 excursion of the remaining 1460 production revealed some outliers. To deal with the outliers the productions were removed that contributed F0 maximum and F0 excursion data points below 0.01% or above 99.9% of the data (i.e., data between 0.001-quantile and 0.999-quantile was accepted). This procedure led to the exclusion of an additional nine productions (0.62% of the remaining data; four in [broad, broad], three in [background, narrow], and two in [background, corrective]).

The final data set that entered the analysis contains 1451 data points. The data set is almost balanced for focus conditions: 465 productions are [broad, broad] (32.1% of the data); 483 productions are [background, narrow] (33.3% of the data); 503 productions are [background, corrective] (34.7% of the data).

2.6 Statistical analyses

The effects on the phonetic measures F0 maximum, F0 excursion, and stressed syllable duration as well as the differences between the target words with regard to these parameters within the phrase are assessed statistically by using Bayesian linear mixed models for each phonetic parameter and position (pre-nuclear vs. nuclear) with brms (Bürkner, 2018), an interface to Bayesian inference in Stan (Carpenter, Gelman, Hoffman, Lee, Goodrich, Betancourt, Brubaker, Guo, Li, & Riddell, 2017). The goal of Bayesian regression is to obtain the probability of the parameters of the model (including the regression coefficients) given the data analyzed. This technique does not attempt to find single optimal values for the model parameters but instead the posterior probability *distributions* of these parameters (Franke & Roettger, 2019). As a consequence, Bayesian modeling allows us to discuss the probability of a parameter directly and thus fits researchers' intuitive understanding of statistical results better than frequentist approaches (Nalborczyk, Batailler, Løvenbruck, Vilain, & Bürkner, 2019). Among others, Vasishth, Nicenboim, Beckman, Li, and Kong (2018), Franke and Roettger (2019), and Nalborczyk et al. (2019) offer very useful tutorial introductions using linguistic data (see also McElreath, 2020).

All models used FOCUS CONDITION as fixed effect and include random intercepts for SPEAKER and TARGET WORD, as well as random slopes for the effect of FOCUS CONDITION by SPEAKER and by TARGET WORD. In the case of differences within the phrase, the variable TARGET WORD represents the combination of the pre-nuclear and nuclear target word. For example, in the sentence *Er hat den Hammer auf die Wohse gelegt*, this value of TARGET WORD is "HammerWohse".

All models ran four chains with 9000 iterations. Weakly informative prior for regression coefficient with a mean of zero and a standard deviation of 10 were used. All other priors were the default priors of brms. Convergence of the models was checked by ensuring no model yielded *Rhat* values larger than 1. The *Rhat* statistic considers the variance of the samples of each sampling chain to that from all samples across all chains. Its value indicates whether the different chains reached a similar outcome (Franke & Roettger, 2019). The model fits were assessed by visual inspection of the predictive posterior checks. For plotting and data processing, tidyverse (Wickham, Averick, Bryan, Chang, McGowan, François, Golemund, Hayes, Henry, Hester, Kuhn, Pedersen, Miller, Bache, Müller, Ooms, Robinson, Seidel, Spinu, Yutani, 2019), emmeans (Lenth, 2022), zoo (Zeileis & Grothendieck, 2005), and cowplot (Wilke, 2020) were used.

The F0 trajectories are analyzed using Generalized Additive Mixed Models (GAMMs) in R (R Core Team, 2021) mgcv (Wood, 2011) and tidymv (Coretta, 2022) for fitting and visualizing. Prior to performing this analysis, the F0 trajectories were interpolated linearly. GAMMs were fitted to the data of each position, pre-nuclear (word A) and nuclear (word B). As fixed effects, the models included FOCUS CONDITION as a parametric term and smooths over TIME for FOCUS CONDITION. In addition, random factor smooths per focus condition were included for

the individual levels of SPEAKER and TARGET WORD. Following the model scheme outlined in Sós-kuthy (2017) and Wieling (2018), the models were fitted such that the smooth over TIME for the condition [background, narrow] represents the reference smooth, and the model contains difference smooths for the conditions [broad, broad] and [background, corrective]. This approach allows us to conveniently assess the differences relevant for the research questions. That is, the difference smooth for [broad, broad] reflects the effect of producing the pre-nuclear word as part of the focus versus producing the word pre-focally. The difference smooth for [background, corrective] reflects the effect of the following focus type on the pre-focal pre-nuclear word. The full models were compared against two types of null models using the *compareML()* function of the *itsadug* package (Rij, Wieling, Baayen, & Rijn, 2022): (a) A null model without the smooth for FOCUS CONDITION over time, and (b) a null model without the smooth for FOCUS CONDITION over time *and* without the parametric term for FOCUS CONDITION. The syntax of all models is given in the appendix.

3 Results

The presentation of the results is structured as follows: First, average F0 contours are presented to obtain a better general impression of the prosodic realization of the analyzed material. Second, scalar phonetic parameters, F0 maximum, F0 excursion and stressed syllable duration, are analyzed. Third, the F0 trajectories are assessed more formally by presenting the results of the GAMM analysis complementing the analysis of the scalar parameters.

3.1 Average F0 contours

The upper panel of **Figure 2** depicts scatterplots of the F0 points in the window of the entire sentence. The black lines represent contours that were calculated by averaging over all F0 measures of a sample in normalized time. Only those time samples for which more than 25% of measures existed entered the calculation of the average contour. Regions where voiceless sounds are located appear as disruptions in the contours (e.g., /f/ of *auf*, English ‘on’, in the central region of the contour). The average contours were smoothed by applying a rolling mean with a window size of three points. The lower panel of the figure “zooms” into the F0 trajectories by giving the average contours for the pre-nuclear (left) and nuclear word (right).

In the pre-nuclear word, the F0 trajectory for [broad, broad] (green circles) reaches the highest F0 peak while the trajectory of [background, corrective] (red squares) reaches the lowest peak. The condition [background, narrow] (blue triangles) is located in between the two other conditions. In addition, the contour of [background, corrective] exhibits a plateau-like shape. The difference between [broad, broad] and [background, narrow] appears somewhat larger than between [background, narrow] and [background, corrective] in the pre-nuclear domain.

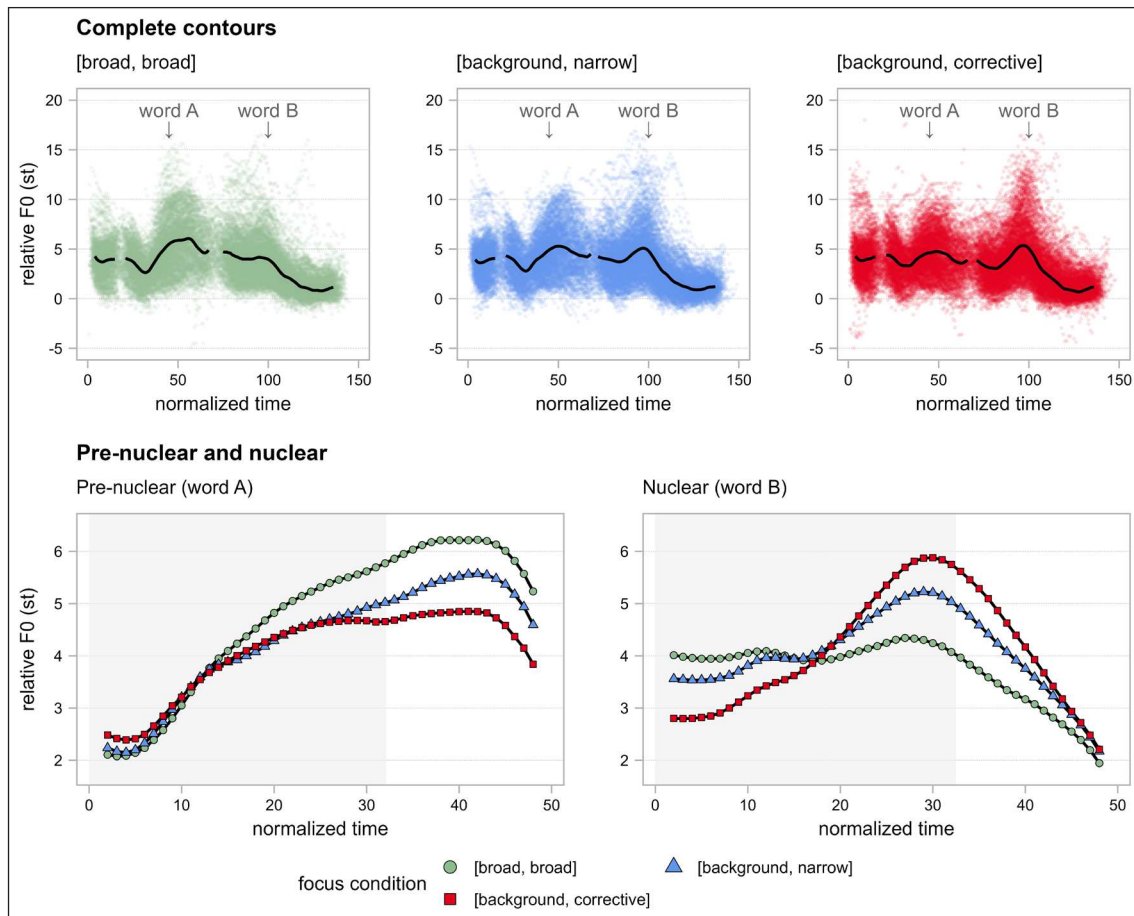


Figure 2: F0 contours. Top: Contours over the complete sentence, raw F0 points and average contours. Bottom: Average contours of the pre-nuclear and nuclear target words. The grey box indicates the average time boundaries of the stressed syllable on the normalized time scale.

In the nuclear word, the image appears to be reversed: The highest peak is found in [background, corrective] and the lowest peak in [broad, broad]. It is also apparent that the contours do not start at the same height in the nuclear word. The beginning of the trajectory is lowest for [background, corrective] and highest for [broad, broad] – with [background, narrow] in between again. Consequentially, F0 excursions seem to follow the pattern [broad, broad] < [background, narrow] < [background, corrective] in the nuclear domain.

Looking at the entire contour, it is evident that the relations between the pre-nuclear and the nuclear F0 peak are different in the three focus conditions. On average, the peak of the pre-nuclear word exceeds the peak of the nuclear word in [broad, broad]. By contrast, the peak of the nuclear word is higher in [background, corrective]. In [background, narrow] the peaks seem to be at roughly the same height.

3.2 Scalar parameters: F0 maximum, F0 excursion, and duration

This subsection is concerned with the analysis of scalar measures in the pre-nuclear and nuclear word. Two of the measures, F0 maximum and F0 excursion, pertain to the F0 contour and reflect the patterns observed in the average contours above. The third measure, stressed syllable duration, extends the analysis perspective to include the temporal domain of prosody. First, the pre-nuclear and nuclear domain are considered separately and compared across focus conditions. Subsequently, the phonetic parameters are analyzed regarding the difference within the same prosodic phrase to gain a better understanding of the prominence relations between pre-nuclear and nuclear words.

3.2.1 Pre-nuclear and nuclear region separately

Table 3 lists the descriptive means and standard deviations for the three measures F0 maximum, F0 excursion, and stressed syllable duration as measured in the pre-nuclear and nuclear word. In the prenuclear position, the values in **Table 3** show the order [broad, broad] > [background, narrow] > [background, corrective]. In nuclear position, on the contrary, the values show the opposite order [broad, broad] < [background, narrow] < [background, corrective].

Pre-nuclear						
	F0 maximum (st)		F0 excursion (st)		Stressed syllable duration (ms)	
<i>Focus condition</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
[broad, broad]	7.30	2.67	5.68	2.47	251.68	59.40
[background, narrow]	6.49	2.27	4.87	2.36	238.16	56.61
[background, corrective]	5.97	2.03	4.12	1.95	231.63	56.40
Nuclear						
	F0 maximum (st)		F0 excursion (st)		Stressed syllable duration (ms)	
<i>Focus condition</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
[broad, broad]	5.63	2.40	4.31	2.20	248.55	42.39
[background, narrow]	6.23	2.75	4.85	2.34	258.30	44.67
[background, corrective]	6.62	2.81	5.42	2.62	261.95	46.75

Table 3: Means and standard deviations for the F0 maximum, F0 excursion and stressed syllable duration.

Table 4 presents the fixed effects results of the Bayesian mixed models in the form of the posterior mean estimates for the correlation coefficient β along with the lower (Q5) and upper (Q95) boundary of the 90% equal-tailed credible interval. The models use [background, narrow] as the reference level for focus condition. Therefore, in **Table 4**, [background, narrow] represents the Intercept, and the estimates β for the other levels indicate a change in comparison to this reference level. **Figure 3** plots the predicted mean values for the three focus conditions with standard errors (conditional effects) in both pre-nuclear and nuclear position.

Pre-nuclear									
	F0 maximum (st)			F0 excursion (st)			Stressed syllable duration (ms)		
<i>Focus condition</i>	β	Q5	Q95	β	Q5	Q95	β	Q5	Q95
[background, narrow] (Intercept)	6.67	5.91	7.43	4.85	4.21	5.50	238.75	211.08	266.28
[broad, broad]	0.89	0.50	1.27	0.85	0.40	1.29	13.43	7.91	18.79
[background, corrective]	-0.64	-0.89	-0.40	-0.75	-1.16	-0.34	-7.13	-11.08	-3.14
Nuclear									
	F0 maximum (st)			F0 excursion (st)			Stressed syllable duration (ms)		
<i>Focus condition</i>	β	Q5	Q95	β	Q5	Q95	β	Q5	Q95
[background, narrow] (Intercept)	6.36	5.49	7.23	4.84	4.22	5.46	261.74	247.28	276.02
[broad, broad]	-0.57	-0.98	-0.16	-0.48	-0.84	-0.13	-8.36	-11.65	-5.06
[background, corrective]	0.28	0.05	0.51	0.53	0.22	0.84	2.15	-0.81	5.08

Table 4: Results of Bayesian models for F0 maximum, F0 excursion and stressed syllable duration. β indicates the mean estimate, Q5 and Q95 indicate the upper and lower boundaries of the 90% equal-tailed credible interval.

In the pre-nuclear position, the models speak in favor of a higher F0 maximum, larger F0 excursion and longer stressed syllable duration in [broad, broad] compared to [background, narrow] (i.e., when the word is in focus compared to when it is out of focus (pre-focal)). The estimates for [broad, broad] are positive and none of the 90% credible intervals includes zero.

Likewise, the models provide evidence for a lower F0 maximum, smaller F0 excursion, and shorter stressed syllable duration in [background, corrective] compared to [background, narrow] (i.e., depending on the type of the following focus). The estimates for the regression coefficients for all three parameters are negative and none of the 90% credible intervals includes zero. Thus, the models support the ordering [broad, broad] > [background, narrow] > [background, corrective] in the pre-nuclear position.

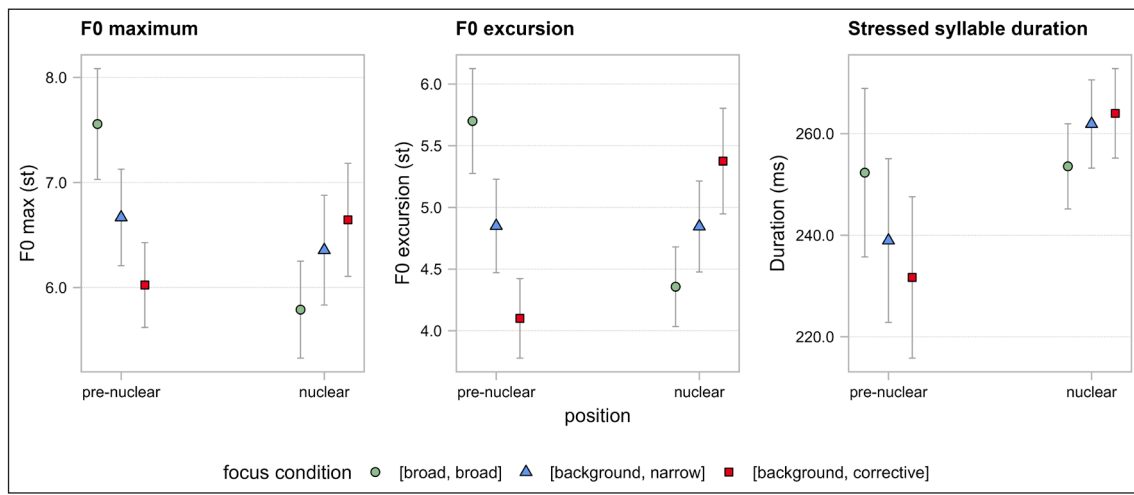


Figure 3: Predicted means and standard errors of F0 maximum, F0 excursion and stressed syllable duration for the three focus conditions and two positions.

In the nuclear position, the models provide evidence for a lower F0 maximum, smaller F0 excursion and shorter stressed syllable duration in [broad, broad] compared to [background, narrow]. The estimates of the regression coefficient for this condition are all negative and none of the 90% credible intervals includes zero. For [background, corrective], the estimates are all positive. However, the estimate for stressed syllable duration is very close to zero and the 90% credible interval includes zero. Overall, the results suggest a higher F0 maximum and larger F0 excursion in [background, corrective] compared to [background, narrow] in the nuclear position. Thus, for F0 maximum and F0 excursion, the models support the order [broad, broad] < [background, narrow] < [background, corrective], while for duration it may be [broad, broad] < [background, narrow] = [background, corrective].

To summarize the analyses in this subsection: Pre-nuclear words that are part of a broad focus exhibit on average higher F0 maxima, larger F0 excursions and longer stressed syllable durations than pre-nuclear words that are in the pre-focal background. Pre-nuclear words in the pre-focal background before corrective focus show on average lower F0 maxima, smaller F0 excursion and shorter stressed syllables than pre-nuclear words before narrow focus. In the

nuclear region, F0 maxima and F0 excursions are larger in narrow compared to broad focus, and in corrective compared to narrow focus. With regard to duration in the nuclear domain, the results present evidence for longer stressed syllables in narrow focus than in broad focus.

3.2.2 Differences between pre-nuclear and nuclear within the phrase

So far, the presentation of the results has concentrated on averages over the pre-nuclear and the nuclear words separately. The analyses suggest that there is an inverse relationship between the pre-nuclear and the nuclear domain. While the ranking [broad, broad] > [background, narrow] > [background, corrective] is found for all measures in the prenuclear domain, the reversed ranking [broad, broad] < [background, narrow] < [background, corrective] is attested in the nuclear domain (with the exception of [background, narrow] vs. [background, corrective] with regard to syllable duration). This subsection presents an exploratory analysis in zooming in to the level of the individual utterance and looking at the distributions of differences within one phrase. That is, the values for the pre-nuclear and nuclear domain are no longer analyzed separately. Instead, their difference for each utterance is calculated by subtracting the value for the pre-nuclear word from that of the nuclear word in the same phrase (nuclear minus pre-nuclear).⁵ Note that all utterances only consist of one phrase. This intra-phrasal difference is denoted with Δ_{ip} , e.g., “ Δ_{ip} F0 maximum”. The descriptive means and standard deviations are given in **Table 5**. Each measure shows the same ranking of Δ_{ip} through the focus conditions: [broad, broad] < [background, narrow] < [background, corrective]. This indicates that the value of the nuclear word increases on average *in relation* to the value of the pre-nuclear word.

<i>Focus condition</i>	Δ_{ip} F0 maximum (st)		Δ_{ip} F0 excursion (st)		Δ_{ip} Stressed syllable duration (ms)	
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
[broad, broad]	-1.67	3.31	-1.37	3.12	-3.13	58.56
[background, narrow]	-0.26	3.11	-0.02	3.19	20.14	56.81
[background, corrective]	0.65	2.77	1.30	3.34	30.32	58.55

Table 5: Means and standard deviations for the intra-phrasal difference Δ_{ip} in F0 maximum, F0 excursion and stressed syllable duration.

Table 6 presents the fixed effect results of the Bayesian mixed model as mean estimates β with the lower (Q5) and upper (Q95) boundaries of the 90% equal-tailed credible interval. Again, the models use [background, narrow] as the reference level for FOCUS CONDITION. Therefore, this level represents the Intercept, and the estimates β for the other levels indicate a

⁵ Subtraction was chosen because it is a simple and transparent operation.

change in comparison to [background, narrow]. The model results support the ranking found in the descriptive means with regard to the intra-phrasal differences in all phonetic parameters: [broad, broad] < [background, narrow] < [background, corrective]. For [broad, broad], the model results show negative estimates β for all differences indicating that Δ_{ip} is smaller in this condition than in [background, narrow]. In all cases, the 90% credible interval does not include zero. On the contrary, the estimates β for [background, corrective] are positive indicating that Δ_{ip} is greater in this condition than in [background, narrow]. In all cases, the 90% credible interval does not include zero. **Figure 4** plots the predicted mean values for each of the three focus conditions with standard errors (conditional effects).

Focus condition	Δ_{ip} F0 maximum (st)			Δ_{ip} F0 excursion (st)			Δ_{ip} Stressed syllable duration (ms)		
	β	Q5	Q95	β	Q5	Q95	β	Q5	Q95
[background, narrow] (Intercept)	-0.31	-1.21	0.61	0.01	-0.78	0.79	16.09	-0.66	32.44
[broad, broad]	-1.39	-2.12	-0.66	-1.36	-2.07	-0.64	-21.45	-26.75	-15.89
[background, corrective]	0.92	0.52	1.32	1.28	0.78	1.77	9.62	5.39	13.71

Table 6: Results of Bayesian models for intra-phrasal difference in F0 maximum, F0 excursion and stressed syllable duration.

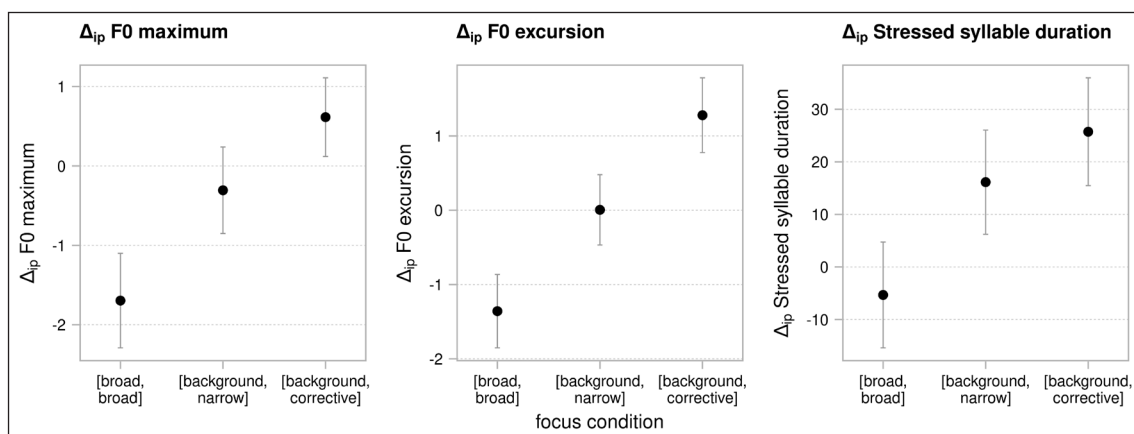


Figure 4: Predicted means of intra-phrasal differences Δ_{ip} in F0 maximum, F0 excursion and stressed syllable duration and standard errors of each focus condition. The intra-phrasal difference Δ_{ip} is calculated by subtracting the value of the pre-nuclear word from the value of the nuclear word (nuclear minus pre-nuclear).

To summarize, the analysis of the intra-phrasal differences revealed the following ranking: [broad, broad] < [background, narrow] < [background, corrective] for all three phonetic measures. This pattern indicates that the value of the nuclear word increases on average *in relation* to the value of the pre-nuclear word when comparing [broad, broad] to [background, narrow], and [background, narrow] to [background, corrective]. Thus, the focus conditions are characterized by relational patterns between the pre-nuclear and nuclear words such that the nuclear word gains prominence relative to the pre-nuclear word in [background, narrow] compared to [broad, broad], and in [background, corrective] compared to [background, narrow]. Interestingly, while the evidence for a difference in syllable duration between narrow and corrective focus with regard to the nuclear word alone was weak, the relational pattern of duration shows a stronger, more robust effect with a 90% credible interval that does not include zero.

3.3 F0 trajectory analysis

In this subsection, the differences in the time course of the F0 contours are assessed statistically using GAMMs. Thus, the same research questions are investigated using a different methodological tool. This analysis is carried out to shed more light on the F0 profiles of the productions in the present data set without a prior reduction of the dynamic data to static parameters (for a discussion of this general issue in phonetics and laboratory phonology, see Wieling, 2018). Models were fit such that the smooth over TIME for the condition [background, narrow] represents the reference smooth, and the model contains difference smooths for the conditions [broad, broad] and [background, corrective] (for details refer to the methods section). This subsection presents the smooths and difference smooths obtained from the models. **Table A.1** in the appendix lists the summary of the model output for the fixed effects smooths over time. The results between the conditions are presented using difference smooths of the form “X – Y”. These difference smooths can be viewed as the result of the subtraction of smooth Y from smooth X. If a negative difference remains in a certain region, this means that Y is higher than X in that region. If a positive difference remains, this means that Y is lower than X in that region. The difference smooths are plotted with their 95% confidence intervals. Red shaded areas along the x-axis indicate where the 95% confidence interval does not include zero (often regarded as the windows of significant differences).

First, the model for the pre-nuclear word is discussed. The comparison of the full model against the null models reveals that the full model was superior both to the simpler model without the slopes for FOCUS CONDITION over time [$\chi^2(12) = 680.23$, $p < 0.001$], and the simpler model without both the parametric term and the slopes for FOCUS CONDITION over time [$\chi^2(14) = 794.50$, $p < 0.001$]. **Figure 5** presents the smooths and differences smooths. The smooths are shown with their 95% confidence intervals in the top panel. The difference

smooths are visualized in the bottom panel. The difference smooth “[background, narrow] – [broad, broad]” on the left side can be conceptualized as the result of subtracting the smooth of [broad, broad] from that of [background, narrow]. In the last two thirds of the pre-nuclear word, a negative difference remains. This indicates that the two trajectories start similarly but the trajectory of [broad, broad] rises higher than that of [background, narrow] starting before the midpoint of the word. The difference smooth “[background, narrow] – [background, corrective]” on the right side can be conceptualized as the result of subtracting the smooth of [background, corrective] from that of [background, narrow]. Again, the start is similar but a positive difference emerges: From roughly the midpoint of the pre-nuclear word onwards, the F0 contour of [background, corrective] takes a lower course than that of [background, narrow].

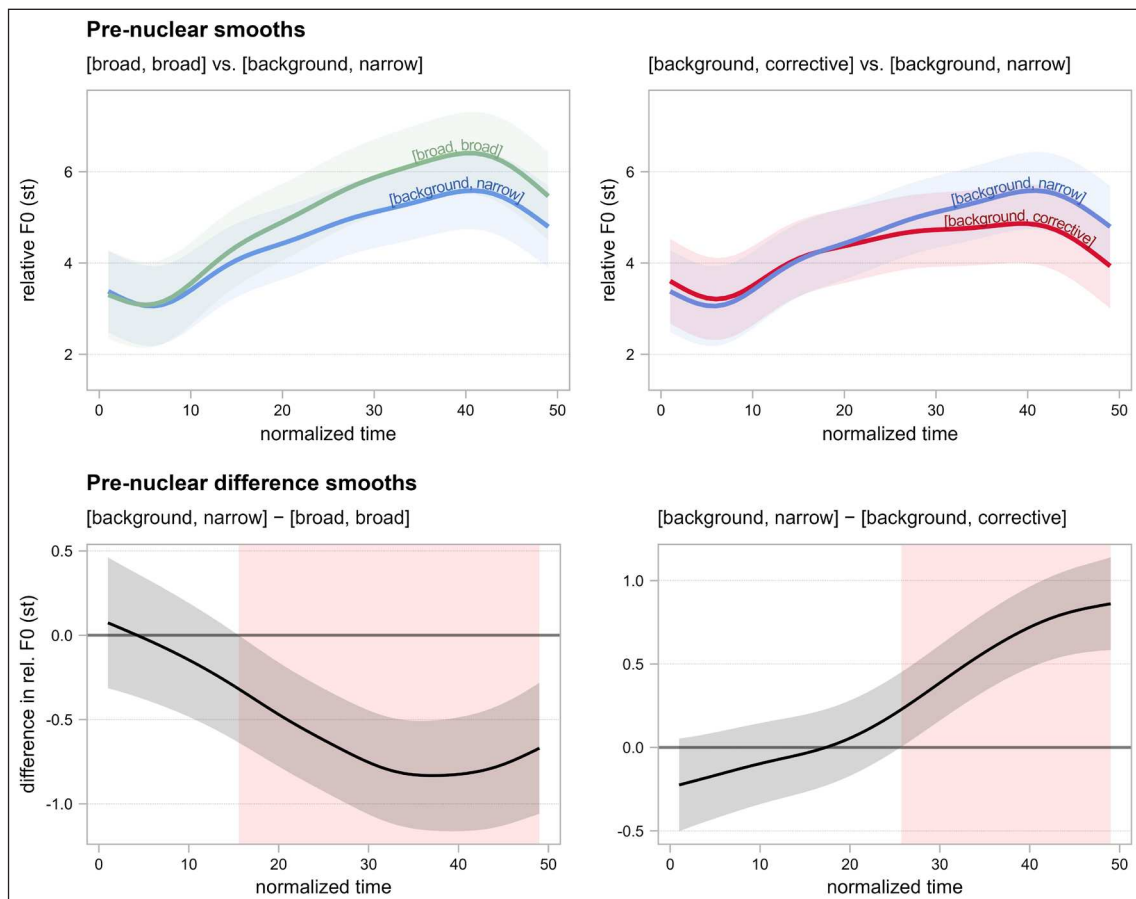


Figure 5: GAMM smooths with 95% confidence interval (top) and difference smooths with 95% confidence interval (bottom) for the pre-nuclear word. Red shaded area in difference smooths indicate where the 95% confidence interval does not include zero.

Second, the model for the nuclear word is discussed. As in the case the pre-nuclear region, the comparison of the full model against the null models reveals that the full model was superior both to the simpler model without the slopes for FOCUS CONDITION over time [$\chi^2(12) = 1471.35$, $p < 0.001$], and the simpler model without both the parametric term and the slopes for FOCUS CONDITION over time [$\chi^2(14) = 1479.05$, $p < 0.001$]. **Figure 6** presents the smooths and different smooths (again smooths at the top; difference smooth at the bottom). The visualizations of the difference smooths in the bottom panel reveal two interesting regions in the contours: At the beginning and around the peak. The difference smooth “[background, narrow] – [broad, broad]” shows that the start of [background, narrow] is lower while the region of the peak is higher compared to [broad, broad]: At the beginning, the difference is negative; in the region of the peak it is positive. Likewise, the difference “[background, narrow] – [background, corrective]” shows that the start of [background, corrective] is lower and its peak is higher than that of [background, narrow]. As reflected in the scalar phonetic parameters analyzed above, the F0 excursion is largest in [background, corrective] and smallest in [broad, broad].

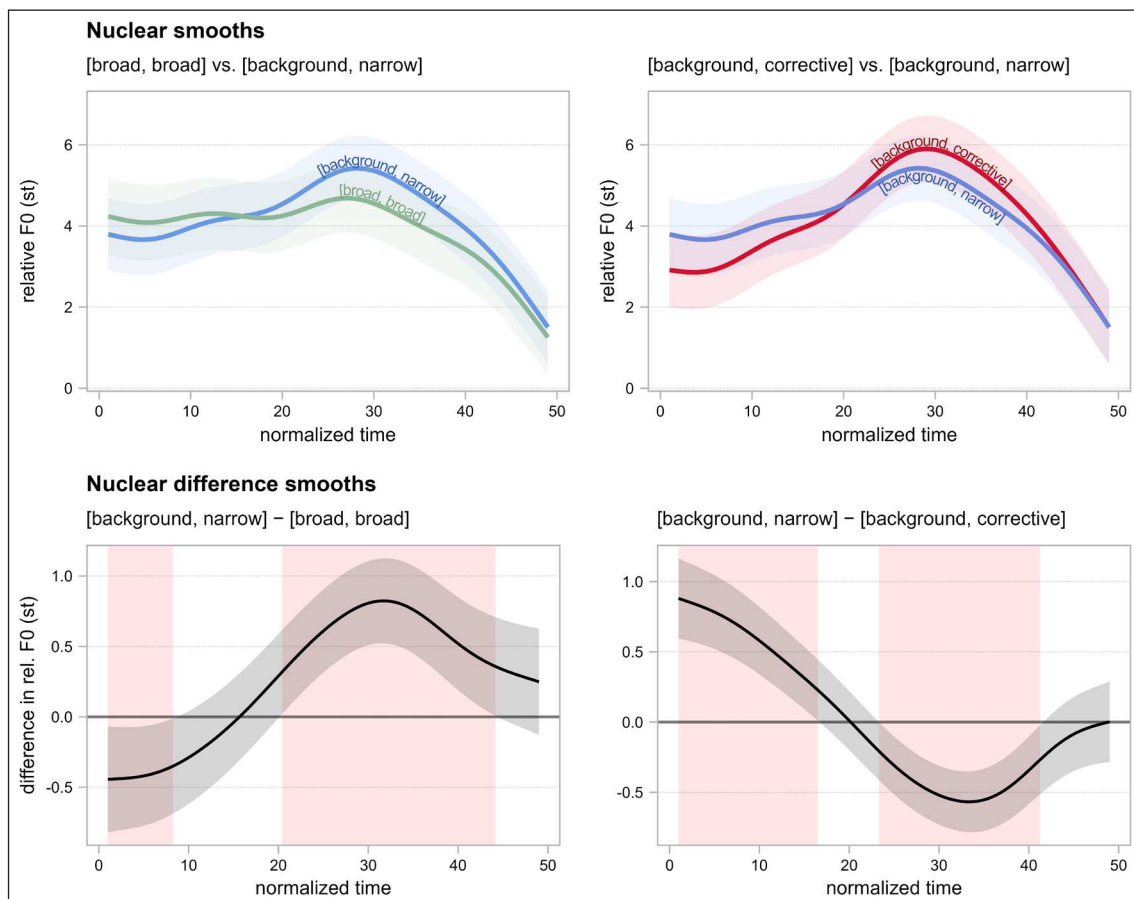


Figure 6: GAMM smooths (top) and difference smooths (bottom) for the nuclear word.

To summarize the analyses of the F0 trajectories, the time course of the F0 contours differs with focus condition in the nuclear *and* pre-nuclear domain. In pre-nuclear position, the following pattern is found:

- (1) There is an effect of “in focus vs. out of focus”: The trajectory for [broad, broad] diverges from that of [background, narrow] at approximately one-third of the word to reach a higher peak.
- (2) There is an influence of the following focus on the pre-focal / pre-nuclear word: The trajectory of [background, corrective] diverges at approximately the midpoint of the word and stays lower than that of [background, narrow].
- (3) The beginnings of the three contours are very similar, which means that the difference in terms of pre-nuclear contours is primarily with regard to the high F0 target later in the word with the ranking [background, corrective] < [background, narrow] < [broad, broad] in the pre-nuclear domain.

In nuclear position, the picture is (partially) reversed. The contours differ in the beginning and the region of the peak:

- (1) The peaks of the contours increase from broad to narrow and from narrow corrective focus.
- (2) The starting points of the contours decrease from broad to narrow and from narrow corrective focus.
- (3) Taken together these differences lead to larger F0 excursions with the ranking [background, corrective] > [background, narrow] > [broad, broad] in the nuclear domain.

4 Discussion

This study presented data on the prosodic realization of words in pre-nuclear and nuclear position in varying focus structures. The first research question targeted the differences between pre-nuclear words in focus (as part of a broad focus) versus pre-nuclear words that occur pre-focally (before narrow focus). The results show that the F0 trajectories reach a higher peak when the pre-nuclear part is in broad focus. The F0 maximum and excursion show higher values and the lexically stressed syllables exhibit longer durations in the pre-nuclear position when the word is in broad focus compared to when it is pre-focal before narrow focus. The data presented here confirm previous findings that the information structural distinction between focus and pre-focal background affects the prosody of utterances in the pre-focal domain (e.g., Féry & Kügler, 2008; Kügler, 2008) using different speech material and employing time-course analyses in addition

to static parameter analysis. The second research question asked whether information about the following focus type is contributed by pre-nuclear words that are in the pre-focal background. The results support the idea that the pre-focal region contains information about the focus structure of the sentence: The F0 trajectories reach lower peaks before corrective focus compared with narrow focus; F0 excursions are smaller and lexically stressed syllable durations are shorter before corrective focus than before narrow focus. The analysis of both research questions shows that the prosodic realization of the pre-nuclear word is reduced when the information structural weight and prosodic prominence of the nuclear word is increased.

One possible interpretation is that *both* prenuclear accent and the nuclear accent are controlled to mark focus structure resulting in a more holistic encoding of information structure. From this perspective, the results contribute evidence against the view that pre-nuclear accents are merely optional and retain the default prosody regardless of information structure. Furthermore, the findings show that information structure marking is distributed over the phrase, and not merely localized in a single accent. These findings mesh well with observations from other studies about the interrelatedness of prosodic elements in a phrase. For example, Braun, Asano, and Dehé (2019) show that a following L- is needed after a L+H* accent to evoke contrastive focus, and that neither L+H* nor L- alone are sufficient. The authors conclude that not the accent alone but the accent plus its *tonal environment* influence the interpretation as contrastive focus. Rump and Collier (1996) show that the relative heights of two successive peaks are used by listeners in deciding whether the sentence signals a single contrast on the second accented word or a double contrast (contrastive topic + contrastive focus).

A slightly different interpretation is that the differences found in the pre-nuclear accent are merely a reflex of the preparation of the nuclear accent. In this case, speakers would not manipulate the pre-nuclear accent intentionally. Instead, a lower peak in the pre-nuclear region could be due to a lower L target at the beginning of the nuclear rise (i.e., a coarticulation of the H tone of the pre-nuclear accent and the L leading tone of the nuclear accent). The study remains inconclusive regarding this question and the exact nature of a potential preparation or coarticulation effect. However, it should be noted that even if the pattern found the pre-nuclear region is just a reflex of preparing the nuclear accent, this region still holds information about the nucleus (and hence the focus marking).

The next two subsections discuss the present findings, in particular the phrasal relationship of the pre-nuclear and nuclear prominences, with respect to an interpretation in intonational phonology and potential perceptual effects.

4.1 Categorical distinctions of prominence relations?

The results are particularly interesting when taking into account the relation of the pre-nuclear word to the nuclear word because prosodic prominence must be seen as an inherently

relational characteristic (Hayes, 1995; Ladd, 2008; Liberman & Prince, 1977). The relational patterns found in the data form prominence profiles spanning larger stretches and entail more than just local adjustments of the nuclear accent. The question arises how we can model these data. Metrical phonology presents a symbolic approach to prominence relations (Hayes, 1995). In this view, stress patterns of words are modeled as an alternation of strong (s) and weaker (w) syllables, as illustrated in **Figure 7A** for the contrasting stress patterns of “permit” as a verb and “permit” as a noun. This way of thinking was extended to prominence relations on the sentence level by assigning strong and weak prominence status to words in the sentence (e.g., Ladd, 2008). The main idea is that information structural differences are reflected in the prominence assignment of words in a sentence, as shown in (i) and (ii) of **Figure 7B**. In (i), broad focus is manifested in that *coffee* is strong while *cup* is weak. This is a consequence of focus projection by which the accent *cup* licenses focus on the whole phrase (Ladd, 1980; Selkirk, 1995). When *cup* is in narrow focus as in (ii), the prominence pattern is reversed. Importantly, the prominence succession w-s, i.e. weak *cup* and strong *coffee*, will be ambiguous between broad and narrow focus as illustrated in **Figure 7B** (iii) (Calhoun, 2010b). The same applies to the case with corrective focus on *coffee* (iv) that is equal to broad and narrow focus.

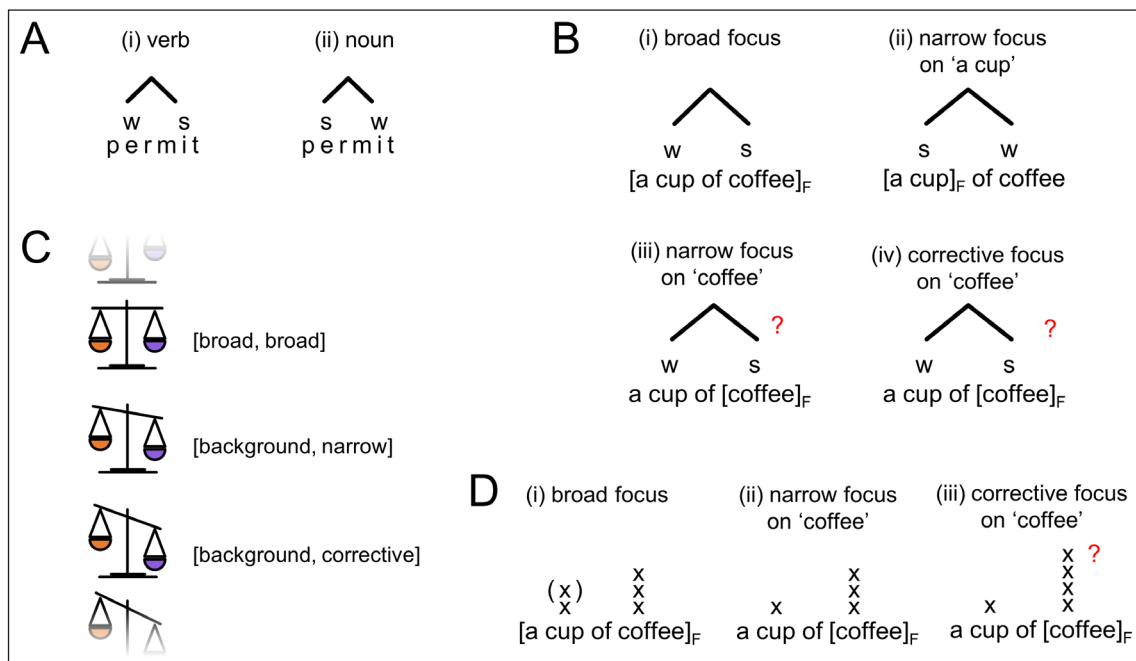


Figure 7: A: Examples of binary prominence relations in metrical phonology on the level of the word. B: Extension of binary prominence relations to the level of the sentence following (Ladd, 2008). C: Scales metaphor illustrating the continuous nature of prominence relations; D: Differentiation of focus types following Selkirk (1995) and extrapolating.

The results presented in this study raise the question whether a binary modeling of phrasal prominence is satisfactory. It may be true that the nuclear accent is stronger than the pre-nuclear accent in all three focus conditions and thus a prominence succession of the form w-s may hold in general. But when comparing the patterns of multiple focus conditions, the data show different prominence relations. It seems as though the nuclear word gains gradually more weight when going from [broad, broad] to [background, narrow] and further to [background, corrective], while the pre-nuclear word loses weight at the same time – as illustrated by the balance scales in **Figure 7C**. Note that a higher degree of prominence results in larger weight on the scale. With larger weight the scale tilts more to one side (given that the other side remains constant or loses weight). The binary modeling of prominence relations adequately represents the strength relations in the sentence when looking at an isolated focus structure. When considering different focus structures, it appears too coarse and falls short of capturing the extent of the imbalance between the two prominences. The specific profiles of balance or imbalance may, however, play an important role for the prosodic expression of information structure.

A possible differentiation of the prosodic patterns of broad focus and narrow focus can be derived from the theory of Selkirk (1995) as depicted in **Figure 7D** (i) and (ii). The approach builds upon the observation that the word *cup* is new in broad focus but given in narrow focus. This difference is deemed to be responsible for different accentual patterns: In the narrow focus case, the pre-nuclear accent is unexpected while it is optional in broad focus (Bishop, 2017). The use of metrical grids allows for some form of differentiation. The nuclear accent is represented by three x, a pre-nuclear accent is represented by two x (in addition, a stressed syllable without an accent receives one x). The question arises how the corrective focus case could be modeled in this theory. Here, the word *cup* is given as in the narrow focus case. One possibility would be to add an extra layer of prominence and adding another x to the nucleus to represent the prominence relation between the pre-nuclear part and the nuclear part. This approach is not without problems. The first problem is that it posits that no pre-nuclear accents are possible before narrow and corrective focus. While pre-nuclear accents are very hard to reliably detect (Ladd, 2008), the F0 trajectories presented here speak against categorical deaccentuation of the pre-nuclear word. There is considerable movement rather than a flat stretch of F0 that would be expected for unaccented pre-nuclear words. Results from other studies (Baumann et al., 2007; Féry & Kügler, 2008) also suggest that pre-nuclear accents are rather common before narrow and/or corrective focus (Baumann, Mertens, Kalbertodt, 2021). A possible solution could be a probabilistic approach in which different probabilities for pre-nuclear accents are assigned to the focus structures (e.g., broad focus has a probability of 0.9 for pre-nuclear accentuation, while narrow focus has a probability 0.7 to be accented in the pre-nuclear region). The importance of probabilistic approaches to the relation of prosody and meaning have been pointed out by various authors (e.g., Calhoun, 2010b; Cangemi & Grice, 2016; Kurumada & Roettger, 2021).

Another potential problem might be that an extra prominence level to differentiate corrective focus opens up the possibility to add an unrestricted number of prominence levels if a new prominence relation is detected in future investigations (a fifth or sixth level may be necessary). Furthermore, we must ask whether there is evidence for a categorical distinction between the prominence levels. Consequentially, the system to describe the relations in a grid format may grow into a ‘digitization’ of what may essentially be a continuous phenomenon. In addition to the importance of probability for prosody, the role of the interplay between categorical (e.g., accentuation/deaccentuation or accent type) and continuous aspect (e.g., peak height, alignment) has been outlined (e.g., Cangemi & Baumann, 2020; Grice et al., 2017; Ladd, 2022). This perspective cannot be fully explored in the context of the present study since the data lack categorical, symbolic annotations of accentuation and accent types. The investigation of the categorical-continuous interplay in the context of the present study has thus to be left for future research.

Another important aspect to bear in mind is that AM phonology views metrical structure and tonal structure as separate. Hence, the differentiation of prominence profiles associated with focus structures could take place at the level of tonal string *only*, while leaving the metrical relations intact. This differentiation could then manifest in different pitch accent choices and / or in terms of continuous modifications within one pitch accent category. Furthermore, the probabilistic approach mentioned above does not need to be restricted to accentuation (whether the pre-nuclear region gets accented or not) and can be extended to accent type choice (e.g., a probability of 0.7 for H* and 0.3 for L* + H). Finally, the discussion here was limited to the relation of two accents. More complex patterns may arise from the interaction of more than two accents.

4.2 The role of pre-nuclear prominence and phrasal prominence relations for perception

Whether the differences found in the present production data are used by listeners in perception remains open. It has been shown that listeners can to some extent differentiate broad focus from narrow and/or corrective focus although the detection rate is not perfect (Breen et al., 2010; Cangemi et al., 2015; Grice et al., 2017). In the data used in these experiments, the realization of the whole contour including the nuclear accent was different in the various focus conditions. Therefore, a perceptual difference between focus types may not be driven by the nuclear accent alone.

One question is thus whether listeners use pre-nuclear information at all. As outlined in the introduction, the importance of pre-nuclear accents has been neglected by some authors. This view seems to be supported by Kapatsinski, Olejarczuk, and Redford (2017), who show in a learning experiment with artificial intonation contours that older children and adults pay more

attention to later parts of intonation contours while younger children may be equally sensitive to earlier and later parts. Concentration on the nuclear part of the contour may be a learned pattern that facilitates the processing of prosodic information. Nonetheless, there is also evidence that prosodic information before the nuclear accent does play a role in processing. Bishop (2017) demonstrates in a series of associative priming experiments that the absence of the pre-nuclear accent makes a narrow focus interpretation more likely than a broad focus interpretation. More generally, research employing eye-tracking or gating paradigms indicates that the processing of prosodic information is incremental (i.e., that early prosodic information is used before the entire sentence is heard and the intonation contour is complete) (Braun & Biezma, 2019; Ito & Speer, 2008; Kurumada, Brown, Bibyk, Pontillo, & Tanenhaus, 2014; Petrone & D'Imperio, 2011; Petrone & Niebuhr, 2014; Roettger, Turner, Cole, 2020; Weber et al., 2006).

The role of a probabilistic mapping between prosodic realization and focus structure has been outlined above for production findings. It needs to be considered in the context of perception as well. Calhoun (2010b), for instance, proposes that the probability of interpreting a word as part of the focus is influenced by various factors, including its position but also whether it is realized more or less prominently than expected.

Another question is whether listeners exploit the relative prominence patterns of the pre-nuclear and nuclear accents. The findings of Rump and Collier (1996) indicate that listeners are sensitive to the connection between focus and the relative scaling of the pre-nuclear and nuclear peaks. Their results show that corrective focus is more acceptable with a lower first peak and a higher second peak compared with broad focus. In terms of the processing of prosodic information, there is evidence that later acoustic information is integrated to update the interpretation of earlier events. In other words, earlier prosodic events are re-interpreted in relation to prosodic events downstream (Dennison & Schafer, 2010; Heim & Alter, 2006; Roettger et al., 2020). Thus, accents may not be (only) meaningful on their own, but their relative pattern can carry important information.

Related to this idea is the question as to how exactly the phonetic production patterns relate to the perception of prominence (relations). So far, I have assumed that higher F₀ peaks, larger F₀ excursions and longer durations mean more prominence of a word. Increasing these parameters on the nuclear word while decreasing them on the pre-nuclear word may thus tilt the prominence scale towards the nuclear word (i.e., the nuclear word becomes more prominent). For paradigmatic prominence relations, there is good evidence for the idea that increases in F₀ peaks, F₀ excursions, and durations lead to increase perceived prominence (e.g., Baumann & Winter, 2018; Bishop, Kuo, & Kim, 2020; Cole, Hualde, Smith, Eager, Mahrt, & Napoleão de Souza, 2019; Turk & Sawusch, 1996; e.g., a word with a longer duration is found to be more prominent than a word with a shorter duration in the same position). When it comes

to the perception of syntagmatic prominence relations, the situation may be more complex. The simplest assumption would be that prominence is syntagmatically relational and hence decreasing a phonetic parameter in one position leads to an increase of perceived prominence in another position in the same phrase. For example, lowering the F0 peak in a pre-nuclear word makes the nuclear word later in the phrase more prominent, even if the F0 peak of that nuclear word is held constant.

The *Gussenhoven-Rietveld effect* challenges this expectation (Gussenhoven & Rietveld, 1988). The effect describes the unexpected finding that raising the pre-nuclear peak boosts the perceived prominence of the nuclear peak in Dutch. In an attempt to replicate the effect in English, Ladd and colleagues found that the effect holds for moderate levels of the second peak (Ladd, Verhoeven, & Jacobs, 1994). However, when the second peak is raised above 140 Hz for a male speaker, the relation is as expected: Raising the first peak leads to lower prominence ratings of the second peak. The authors' interpretation holds that a second peak below the threshold is not evaluated individually. In this setting, listeners rather rate the prominence of the whole utterance. Given that both peaks contribute to the prominence rating of the whole utterance, raising the first peak is expected to increase the prominence response of the nuclear accent (and the whole phrase). Whether or not this explanation is plausible, the findings show that the relation between two F0 peaks may not be straightforward in perception. Future research will have to evaluate how the pattern found in the present production study is reflected perceptually.

4.3 Limitations and future directions

There are several limitations of the current study that should be addressed in future work. First, the speech material contained a single syntactic structure (*Er hat die <A> auf die gelegt*, "He put the <A> on the "). In contrast to this study, empirical investigations of focus often employ sentences like "Mary wants to see Paul" with a direct object and no prepositional phrase. The question arises whether the results presented here generalize to this and other syntactic structures. A second issue with the current speech material may be that a certain asymmetry arises through the use of a real word in the pre-nuclear region and a nonce word in the nuclear region. It should be noted that at least this is consistent across all conditions. Furthermore, future research should look more closely at prominence relations in sentences with more than two pitch accents.

Second, the effects presented in this paper are overall rather small. This is interesting for future research because of at least two reasons: (1) The data presented here are elicited in a controlled scenario resulting in what is often called "lab speech" (although the interactive scenario does not elicit merely read speech). It is unclear whether the effects are expected to be greater or smaller in spontaneous speech. (2) As mentioned earlier, an important question

of future research is whether the present differences are relevant in perception. Progress on answering these questions will help to better understand the role of the pre-nuclear region and the relation between pre-nuclear and nuclear prominence for focus marking.

5 Conclusion

The present study has shown that the pre-nuclear part is affected by focus marking. Its realization depends on whether it is focal or pre-focal and, if it is pre-focal, which kind of focus follows. The results add evidence questioning the view that default prosodic patterns are retained in the pre-nuclear domain. The data also show that the pre-nuclear and the nuclear part stand in an inverse relationship. This in turn suggests that information structure does not merely affect local prominence but leads to prominence profiles distributed over larger stretches of speech.

Data Accessibility Statement

The data table and analysis scripts are available on Open Science Framework: <https://osf.io/gvtx7/>.

Acknowledgements

This work was supported by the German Research Foundation (DFG) as part of the Walter Benjamin project RO 6767/1-1. I would like to thank Sam Tilsen, Stefan Baumann, and Janne Lorenzen for comments on an earlier version of this work.

Competing Interests

The author has no competing interests to declare.

References

- Alzaidi, M. S. A., Xu, Y., Xu, A., & Szreder, M. (2023). Analysis and computational modelling of Emirati Arabic intonation – A preliminary study. *Journal of Phonetics*, 98, 101236. DOI: <https://doi.org/10.1016/j.wocn.2023.101236>
- Alzamil, A., & Hellmuth, S. (2021). The realization of different structural focus conditions in Saudi Arabic dialects. *Proceedings of the 1st International Conference on Tone and Intonation (TAI)*, 196–199.
- Andreeva, B., Barry, W. J., & Koreman, J. (2017). Local and Global Cues in the Prosodic Realization of Broad and Narrow Focus in Bulgarian. *Phonetica*, 73(3–4), 256–278. DOI: <https://doi.org/10.1159/000448044>
- Baumann, S., Becker, J., Grice, M., & Mücke, D. (2007). Tonal and Articulatory Marking of Focus in German. *Proceedings of the 16th International Congress of Phonetic Sciences*, 1029–1032.
- Baumann, S., Grice, M., & Steindamm, S. (2006). Prosodic Marking of Focus Domains—Categorical or Gradient? *Proceedings of Speech Prosody*, 301–304.
- Baumann, S., Mertens, J., & Kalbertodt, J. (2021). The influence of informativeness on the prosody of sentence topics. *Glossa: A Journal of General Linguistics*, 6(1). DOI: <https://doi.org/10.16995/glossa.5871>
- Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics*, 70, 20–38. DOI: <https://doi.org/10.1016/j.wocn.2018.05.004>
- Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The Original ToBI System and the Evolution of the ToBI Framework. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 9–54). Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199249633.003.0002>

- Bishop, J. (2017). Focus projection and prenuclear accents: Evidence from lexical processing. *Language, Cognition and Neuroscience*, 32(2), 236–253. DOI: <https://doi.org/10.1080/23273798.2016.1246745>
- Bishop, J., Kuo, G., & Kim, B. (2020). Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from Rapid Prosody Transcription. *Journal of Phonetics*, 82, 100977. DOI: <https://doi.org/10.1016/j.wocn.2020.100977>
- Boersma, P., & Weenink, D. (2001). PRAAT, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341–345.
- Bolinger, D. L. (1965). *Forms of English: Accent, morpheme and order*. Harvard University Press.
- Braun, B. (2006). Phonetics and phonology of thematic contrast in German. *Language and Speech*, 49(4), 451–493. DOI: <https://doi.org/10.1177/00238309060490040201>
- Braun, B., Asano, Y., & Dehé, N. (2019). When (not) to Look for Contrastive Alternatives: The Role of Pitch Accent Type and Additive Particles. *Language and Speech*, 62(4), 751–778. DOI: <https://doi.org/10.1177/0023830918814279>
- Braun, B., & Biezma, M. (2019). Prenuclear L* + H Activates Alternatives for the Accented Word. *Frontiers in Psychology*, 10. <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01993>. DOI: <https://doi.org/10.3389/fpsyg.2019.01993>
- Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, 25(7–9), 1044–1098. DOI: <https://doi.org/10.1080/01690965.2010.504378>
- Büring, D. (2007). Intonation, Semantics and Information Structure. In G. Ramchand & C. Reiss (Eds.), *The Oxford Handbook of Linguistic Interfaces* (pp. 445–474). Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780199247455.013.0015>
- Büring, D. (2016). *Intonation and Meaning*. Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199226269.001.0001>
- Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, 10(1), 395–411. DOI: <https://doi.org/10.32614/RJ-2018-017>
- Calhoun, S. (2009). What Makes a Word Contrastive? Prosodic, Semantic and Pragmatic Perspectives. In D. Barth-Weingarten, N. Dehé, & A. Wichmann (Eds.), *Where Prosody Meets Pragmatics* (pp. 53–77). BRILL. DOI: https://doi.org/10.1163/9789004253223_004
- Calhoun, S. (2010a). How does informativeness affect prosodic prominence? *Language and Cognitive Processes*, 25(7–9), 1099–1140. DOI: <https://doi.org/10.1080/01690965.2010.491682>
- Calhoun, S. (2010b). The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language*, 86(1), 1–42. DOI: <https://doi.org/10.1353/lan.0.0197>
- Calhoun, S. (2012). The theme/rheme distinction: Accent type or relative prominence? *Journal of Phonetics*, 40(2), 329–349. DOI: <https://doi.org/10.1016/j.wocn.2011.12.001>
- Cangemi, F., & Baumann, S. (2020). Integrating phonetics and phonology in the study of linguistic prominence. *Journal of Phonetics*, 81, 100993. DOI: <https://doi.org/10.1016/j.wocn.2020.100993>

- Cangemi, F., & Grice, M. (2016). The Importance of a Distributional Approach to Categoricity in Autosegmental-Metrical Accounts of Intonation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1), 1–20. DOI: <https://doi.org/10.5334/labphon.28>
- Cangemi, F., Krüger, M., & Grice, M. (2015). Listener-specific perception of speaker-specific productions in intonation. In S. Fuchs, D. Pape, C. Petrone, & P. Perrier (Eds.), *Individual Differences in Speech Production and Perception* (pp. 123–145). Peter Lang.
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, 76(1), 1–32. DOI: <https://doi.org/10.18637/jss.v076.i01>
- Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and Topic* (pp. 25–55). Academic Press.
- Chahal, D., & Hellmuth, S. (2014). The intonation of Lebanese and Egyptian Arabic. In S.-A. Jun (Ed.), *Prosodic Typology II: The Phonology of Intonation and Phrasing* (pp. 365–404). Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199567300.003.0013>
- Chodroff, E., & Cole, J. (2018). Information Structure, Affect and Prenuclear Prominence in American English. *Proceedings of Interspeech 2018*, 1848–1852. DOI: <https://doi.org/10.21437/Interspeech.2018-1529>
- Cole, J., Hualde, J. I., Smith, C. L., Eager, C., Mahrt, T., & Napoleão de Souza, R. (2019). Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics*, 75, 113–147. DOI: <https://doi.org/10.1016/j.wocn.2019.05.002>
- Coretta, S. (2022). *tidymv: Tidy Model Visualisation for Generalised Additive Models*. <https://CRAN.R-project.org/package=tidymv>
- Dennison, H. Y., & Schafer, A. J. (2010). Online construction of implicature through contrastive prosody. *Proc. Speech Prosody 2010*, paper 338.
- Eady, S. J., & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *The Journal of the Acoustical Society of America*, 80(2), 402–415. DOI: <https://doi.org/10.1121/1.394091>
- Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, 36(4), 680–703. DOI: <https://doi.org/10.1016/j.wocn.2008.05.001>
- Franke, M., & Roettger, T. B. (2019). *Bayesian regression modeling (for factorial designs): A tutorial*. PsyArXiv. DOI: <https://doi.org/10.31234/osf.io/cdxv3>
- Grice, M., Baumann, S., & Benz Müller, R. (2005). German Intonation in Autosegmental-Metrical Phonology. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 55–83). Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199249633.003.0003>
- Grice, M., Ritter, S., Niemann, H., & Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics*, 64, 90–107. DOI: <https://doi.org/10.1016/j.wocn.2017.03.003>
- Gussenhoven, C. (1983). Testing the Reality of Focus Domains. *Language and Speech*, 26(1), 61–80. DOI: <https://doi.org/10.1177/002383098302600104>

- Gussenhoven, C. (2007). Types of Focus in English. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and Focus* (Vol. 82, pp. 83–100). Springer Netherlands. DOI: https://doi.org/10.1007/978-1-4020-4796-1_5
- Gussenhoven, C., & Rietveld, A. C. M. (1988). Fundamental frequency declination in Dutch: Testing three hypotheses. *Journal of Phonetics*, 16, 355–369. DOI: [https://doi.org/10.1016/S0095-4470\(19\)30509-1](https://doi.org/10.1016/S0095-4470(19)30509-1)
- Halliday, M. A. K. (1967). *Intonation and grammar in British English*. De Gruyter. DOI: <https://doi.org/10.1515/9783111357447>
- Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. University of Chicago Press.
- Heim, S., & Alter, K. (2006). Prosodic pitch accents in language comprehension and production: ERP data and acoustic analyses. *Acta Neurobiologiae Experimentalis*, 66, 55–68. DOI: <https://doi.org/10.55782/ane-2006-1587>
- Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, 58(2), 541–573. DOI: <https://doi.org/10.1016/j.jml.2007.06.013>
- Jabeen, F., Wagner, P., & Hartmann, J. (2021). Creativity and Variability in the Perception of Prosody and Focus Marking in German. *1st International Conference on Tone and Intonation (TAI)*, 142–146. DOI: <https://doi.org/10.21437/TAI.2021-29>
- Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. MIT Press.
- Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, 71, 1–15. DOI: <https://doi.org/10.1016/j.wocn.2018.07.001>
- Jun, S.-A. (2005). *Prosodic Typology*. Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199249633.001.0001>
- Kapatsinski, V., Olejarczuk, P., & Redford, M. A. (2017). Perceptual Learning of Intonation Contour Categories in Adults and 9- to 11-Year-Old Children: Adults Are More Narrow-Minded. *Cognitive Science*, 41(2), 383–415. DOI: <https://doi.org/10.1111/cogs.12345>
- Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, 55(3–4), 243–276. DOI: <https://doi.org/10.1556/ALing.55.2008.3-4.2>
- Kügler, F. (2008). The role of duration as a phonetic correlate of focus. *Proceedings of Speech Prosody 2008*, 591–594.
- Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, 133(2), 335–342. DOI: <https://doi.org/10.1016/j.cognition.2014.05.017>
- Kurumada, C., & Roettger, T. B. (2021). Thinking probabilistically in the study of intonational speech prosody. *WIREs Cognitive Science*, e1579. DOI: <https://doi.org/10.1002/wcs.1579>
- Ladd, D. R. (1980). *The structure of intonational meaning: Evidence from English*. Indiana University Press.
- Ladd, D. R. (2008). *Intonational Phonology*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511808814>

- Ladd, D. R. (2022). The Trouble with ToBI. In J. Barnes & S. Shattuck-Hufnagel (Eds.), *Prosodic Theory and Practice* (pp. 247–257). The MIT Press. DOI: <https://doi.org/10.7551/mitpress/10413.003.0009>
- Ladd, D. R., Verhoeven, J., & Jacobs, K. (1994). Influence of adjacent pitch accents on each other's perceived prominence: Two contradictory effects. *Journal of Phonetics*, 22, 87–99. DOI: [https://doi.org/10.1016/S0095-4470\(19\)30268-2](https://doi.org/10.1016/S0095-4470(19)30268-2)
- Lambrecht, K. (1994). *Information Structure and Sentence Form*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511620607>
- Lambrecht, K. (2000). When subjects behave like objects: An analysis of the merging of S and O in Sentence-Focus Constructions across languages. *Studies in Language*, 24(3), 611–682. DOI: <https://doi.org/10.1075/sl.24.3.06lam>
- Lenth, R. V. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. <https://CRAN.R-project.org/package=emmeans>
- Lieberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. Oehrle, *Language sound structure* (pp. 157–233). MIT Press.
- Lieberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8, 249–336.
- McAuliffe, M., Socolof, M., Mihuc, S., & Wagner, M. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *Proceedings of INTERSPEECH, 20–24 August, Stockholm, Sweden*, 498–502. DOI: <https://doi.org/10.21437/Interspeech.2017-1386>
- McElreath, R. (2020). *Statistical rethinking* (Second edition). CRC Press. DOI: <https://doi.org/10.1201/9780429029608>
- Mücke, D., & Grice, M. (2014). The effect of focus marking on supralaryngeal articulation – Is it mediated by accentuation? *Journal of Phonetics*, 44, 47–61. DOI: <https://doi.org/10.1016/j.wocn.2014.02.003>
- Nalborczyk, L., Batailler, C., Løevenbruck, H., Vilain, A., & Bürkner, P.-C. (2019). An Introduction to Bayesian Multilevel Models Using brms: A Case Study of Gender Effects on Vowel Variability in Standard Indonesian. *Journal of Speech, Language, and Hearing Research*, 62(5), 1225–1242. DOI: https://doi.org/10.1044/2018_JSLHR-S-18-0006
- Petrone, C., & D'Imperio, M. (2011). From Tones to Tunes: Effects of the f0 Prenuclear Region in the Perception of Neapolitan Statements and Questions. In S. Frota, G. Elordieta, & P. Prieto (Eds.), *Prosodic Categories: Production, Perception and Comprehension* (pp. 207–230). Springer Netherlands. DOI: https://doi.org/10.1007/978-94-007-0137-3_9
- Petrone, C., & Niebuhr, O. (2014). On the Intonation of German Intonation Questions: The Role of the Prenuclear Region. *Language and Speech*, 57(1), 108–146. DOI: <https://doi.org/10.1177/0023830913495651>
- Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., & Vesely, K. (2011, December). The Kaldi Speech Recognition Toolkit. *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*.
- Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical Pragmatics* (pp. 223–255). Academic Press.

- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Repp, S. (2016). Contrast: Dissecting an Elusive Information-structural Notion and its Role in Grammar. In C. Féry & S. Ishihara (Eds.), *The Oxford Handbook of Information Structure* (pp. 270–289). Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780199642670.013.006>
- Riester, A., & Baumann, S. (2013). Focus Triggers and Focus Types from a Corpus Perspective. *Dialogue & Discourse*, 4(2), 215–248. DOI: <https://doi.org/10.5087/dad.2013.210>
- Rij, J. van, Wieling, M., Baayen, R. H., & Rijn, H. van. (2022). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs* [Computer software].
- Roessig, S., Winter, B., & Mücke, D. (2022). Tracing the Phonetic Space of Prosodic Focus Marking. *Frontiers in Artificial Intelligence*, 5, 842546. DOI: <https://doi.org/10.3389/frai.2022.842546>
- Roettger, T. B., Turner, D., & Cole, J. (2020). *Intonational processing is incremental and holistic*. PsyArXiv. DOI: <https://doi.org/10.31234/osf.io/nhbgs>
- Röhr, C. T. (2016). *The Information Status of Nominal and Verbal Expressions: Intonational Evidence from Production and Perception in German* [Doctoral Thesis, Universität zu Köln]. <http://www.uni-koeln.de/>
- Rooth, M. (1992). A Theory of Focus Interpretation. *Natural Language Semantics*, 1(1), 75–116. DOI: <https://doi.org/10.1007/BF02342617>
- Rooth, M. (2016). Alternative Semantics. In C. Féry & S. Ishihara (Eds.), *The Oxford Handbook of Information Structure* (pp. 19–40). Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780199642670.013.19>
- Royer, A. J., & Jun, S.-A. (2019). Prominence marking in Kazan Tatar declaratives. *Proceedings of the 19th International Congress of Phonetic Sciences*.
- Rump, H. H., & Collier, R. (1996). Focus Conditions and the Prominence of Pitch-Accented Syllables. *Language and Speech*, 39(1), 1–17. DOI: <https://doi.org/10.1177/002383099603900101>
- Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. In J. A. Goldsmith, *The handbook of phonological theory*. Blackwell.
- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C. C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). ToBI: A Standard for Labeling English Prosody. *Second International Conference on Spoken Language Processing*. DOI: <https://doi.org/10.1515/cllt-2012-0011>
- Sóskuthy, M. (2017). *Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction*. arXiv. DOI: <https://doi.org/10.48550/arXiv.1703.05339>
- Steedman, M. (2000). Information Structure and the Syntax-Phonology Interface. *Linguistic Inquiry*, 31(4), 649–689. DOI: <https://doi.org/10.1162/002438900554505>
- Steedman, M. (2007). Information-Structural Semantics for English Intonation. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation* (pp. 245–264). Springer Netherlands. DOI: https://doi.org/10.1007/978-1-4020-4796-1_13
- Stevens, J. S. (2017). Pragmatics of Focus. In J. S. Stevens, *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. DOI: <https://doi.org/10.1093/acrefore/9780199384655.013.207>

- Turk, A. E., & Sawusch, J. R. (1996). The processing of duration and intensity cues to prominence. *The Journal of the Acoustical Society of America*, 99(6), 3782–3790. DOI: <https://doi.org/10.1121/1.414995>
- Vallduví, E., & Engdahl, E. (1996). The linguistic realisation of information packaging. *Linguistics*, 34, 459–519. DOI: <https://doi.org/10.1515/ling.1996.34.3.459>
- Vallduví, E., & Engdahl, E. (2013). The linguistic realization of information packaging. *Linguistics*, 51(s1), 19–20. DOI: <https://doi.org/10.1515/ling-2013-0041>
- Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, 71, 147–161. DOI: <https://doi.org/10.1016/j.wocn.2018.07.008>
- Weber, A., Braun, B., & Crocker, M. W. (2006). Finding Referents in Time: Eye-Tracking Evidence for the Role of Contrastive Accents. *Language and Speech*, 49(3), 367–392. DOI: <https://doi.org/10.1177/00238309060490030301>
- Welby, P. (2003). Effects of Pitch Accent Position, Type, and Status on Focus Projection. *Language and Speech*, 46(1), 53–81. DOI: <https://doi.org/10.1177/00238309030460010401>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. DOI: <https://doi.org/10.21105/joss.01686>
- Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, 70, 86–116. DOI: <https://doi.org/10.1016/j.wocn.2018.03.002>
- Wilke, C. O. (2020). *cowplot: Streamlined Plot Theme and Plot Annotations for “ggplot2.”* <https://CRAN.R-project.org/package=cowplot>
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1), 3–36. DOI: <https://doi.org/10.1111/j.1467-9868.2010.00749.x>
- Xu, Y. (2011). Post-focus compression: Cross-linguistic distribution and historical origin. *Proceedings of the 17th International Congress of Phonetic Sciences*.
- Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, 33(2), 159–197. DOI: <https://doi.org/10.1016/j.wocn.2004.11.001>
- Yang, Y., & Chen, S. (2020). Revisiting focus production in Mandarin Chinese: Some preliminary findings. *Speech Prosody 2020*, 260–264. DOI: <https://doi.org/10.21437/SpeechProsody.2020-53>
- Zeileis, A., & Grothendieck, G. (2005). zoo: S3 Infrastructure for Regular and Irregular Time Series. *Journal of Statistical Software*, 14(6), 1–27. DOI: <https://doi.org/10.18637/jss.v014.i06>
- Zimmermann, M. (2008). Contrastive focus and emphasis. *Acta Linguistica Hungarica*, 55(3–4), 347–360. DOI: <https://doi.org/10.1556/ALing.55.2008.3-4.9>

Appendix

Annotations

Figure A.2 shows the annotations used in this study with a screenshot of one sample sound file and TextGrid.

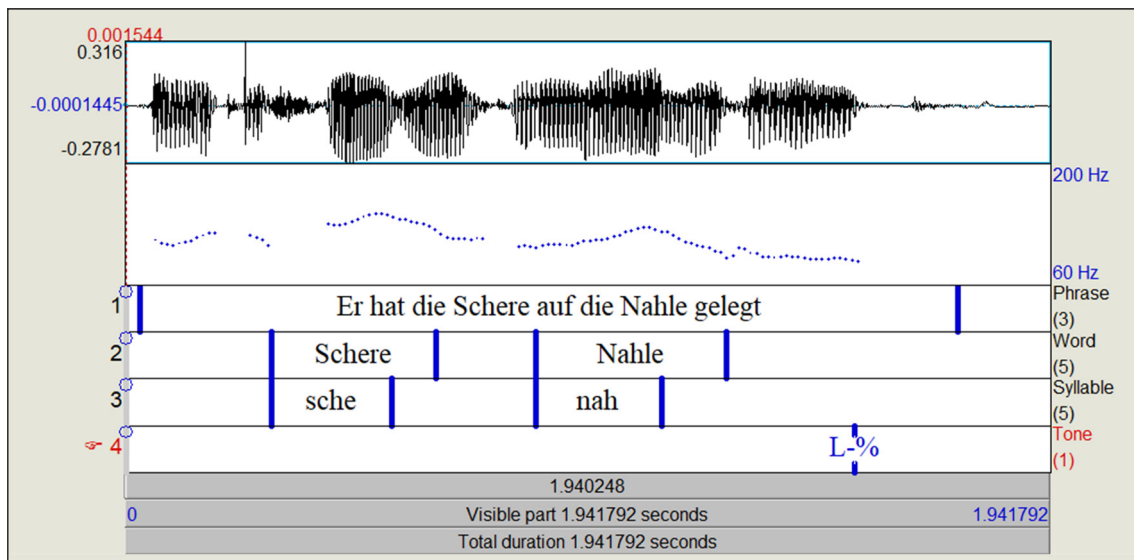


Figure A.1: Example sound file and TextGrid illustrating the annotations.

Statistical modeling

Analysis of static parameters (F0 max, F0 excursion, duration)

The following formula (given here in pseudo-R-code) was used to model the static parameters. The variable *phonetic variable* stands as a placeholder for F0 maximum, F0 excursion and stressed syllable duration.

$$\text{phonetic variable} \sim \text{focus condition} + \\ (1 + \text{focus condition} \mid \text{speaker}) + \\ (1 + \text{focus condition} \mid \text{target word})$$

Time-course analysis (GAMMs)

The following formula (given here in pseudo-R-code) was used to model the F0 trajectories in the pre-nuclear and nuclear region:

$$f_0 \sim \text{focus condition} + \\ s(\text{time}) + \\ s(\text{time}, \text{by} = \text{focus condition}) +$$

s(time, speaker, bs = “fs”, m = 1) +
 s(time, speaker, by = focus condition, bs = “fs”, m = 1) +
 s(time, target word, bs = “fs”, m = 1) +
 s(time, target word, by = focus condition, bs = “fs”, m = 1)

The null model excluding the effect of FOCUS CONDITION for the model comparison was fit with the following formula (this model lacks the smooths over time *and* the parametric term for FOCUS CONDITION):

f0 ~ s(time) +
 s(time, speaker, bs = “fs”, m = 1) +
 s(time, target word, bs = “fs”, m = 1)

The null model excluding the smooths for FOCUS CONDITION over time for the model comparison was fit with the following formula:

f0 ~ focus condition +
 s(time) +
 s(time, speaker, bs = “fs”, m = 1) +
 s(time, target word, bs = “fs”, m = 1)

Table A.1 presents the summary of the model output for the smooths over time in the model for the pre-nuclear word. The first line refers to [background, narrow] which represents the reference smooth. The following lines refer to the difference smooths for [broad, broad] and [background, corrective].

Pre-nuclear			
<i>Smooth</i>	<i>edf</i>	<i>F-value</i>	<i>p-value</i>
s(time): [background, narrow]	8.01	22.86	< 0.001
s(time): [broad, broad]	4.27	4.76	< 0.001
s(time): [background, corrective]	4.27	9.22	< 0.001
Nuclear			
<i>Smooth</i>	<i>edf</i>	<i>F-value</i>	<i>p-value</i>
s(time): [background, narrow]	8.29	107.31	< 0.001
s(time): [broad, broad]	6.68	8.81	< 0.001
s(time): [background, corrective]	6.91	12.47	< 0.001

Table A.1: Results of the fixed effect smooths in the GAMM analysis.

Table A.2 presents the means and standard deviations for all conditions including the condition [corrective, background] that was not included in the data set analyzed in the main part of the paper. This condition is triggered with a question that contains a competitor for the first target noun in the target sentence. For example, the question could be *Hat er die Säge auf die Wohse gelegt?* (“Did he put the saw on the Wohse?”) to trigger the target sentence *Er hat den Hammer auf die Wohse gelegt* (“He put the hammer on the Wohse”). In this case, we cannot speak of pre-nuclear and nuclear in the same sense as in the main analysis of this paper. This is because the nuclear accent falls on the *first* noun, Hammer (“hammer”) in the example above. Therefore, the words are labelled as Word A and Word B. The F0 contours for all four conditions are given in **Figure A.2** as scatterplots and average contours.

Word A						
	F0 maximum (st)		F0 excursion (st)		Stressed syllable duration (ms)	
<i>Focus condition</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
[background, corrective]	8.23	3.50	7.07	3.42	256.38	56.85
[broad, broad]	7.30	2.67	5.68	2.47	251.68	59.40
[background, narrow]	6.49	2.27	4.87	2.36	238.16	56.61
[background, corrective]	5.97	2.03	4.12	1.95	231.63	56.40
Word B						
	F0 maximum (st)		F0 excursion (st)		Stressed syllable duration (ms)	
<i>Focus condition</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
[background, corrective]	2.79	1.39	2.33	1.33	231.34	56.85
[broad, broad]	5.63	2.40	4.31	2.20	248.55	42.39
[background, narrow]	6.23	2.75	4.85	2.34	258.30	44.67
[background, corrective]	6.62	2.81	5.42	2.62	261.95	46.75

Table A.2: Means and standard deviations for the phonetic parameters in both target word positions.

