Regular Article

Language *and* Speech

Postfocal Downstep in German

Language *and* Speech
2017, Vol. 60(2) 260–288

DOI: 10.1177/0023830916647204
journals.sagepub.com/home/las
SAGE

Frank Kügler
University of Potsdam, Germany

Caroline Féry
Frankfurt University, Germany

Abstract
This article is a follow-up study of Féry and Kügler (2008. Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics, 36*, 680–703). It reports on an experiment of the F0 height of potential pitch accents in the postfocal region of German sentences and addresses in this way an aspect of the influence of information structure on the intonation of sentences that was left open in the previous article. The results of the experiment showed that, when several constituents are located in this position, they are often in a downstep relation, but are rarely upstepped. In 37% of the cases, the pitch accents are only realized dynamically and there is no down- or upstepping. We interpret these results as evidence that postfocal constituents are phrased independently. The data examined speak against a model of postfocal intonation in which postfocal phrasing is eliminated and all accents are reduced to zero. Instead, the pitch accents are often present, although reduced. Moreover, the facts support the existence of prosodic phrasing of the postfocal constituents; the postfocal position implies an extremely compressed register, but no dephrasing or systematic complete deaccentuation of all pitch accents. We propose adopting a model of German intonation in which prosodic phrasing is determined by syntactic structure and cannot be changed by information structure. The role of information structure in prosody is limited to changes in the register relationship of the different parts of the sentence. Prefocally, there is no or only little register compression because of givenness. Postfocally, register compression is the rule. A model of intonation must take this asymmetry into account.

Keywords
German intonation, postfocal givenness, downstep, pitch register, postfocal compression

1 Introduction

The aim of the present article is to improve our knowledge of the theoretical status and the details of the phonetic correlates of pitch register compression in German. We address this issue from both a phonetic and a phonological perspective.

Corresponding author:
Frank Kügler, Department of Linguistics, University of Potsdam, Karl-Liebknecht-Str. 24-25, D–14476 Potsdam, Germany.
Email: kuegler@uni-potsdam.de

Postfocal compression (PFC)–or postfocal deaccenting–is an indicator of information structural givenness: the words following a narrow focused constituent are realized in an extremely compressed pitch register. From a phonetic perspective, the question is whether the postfocal constituents have phonetic cues in the signal that may be interpreted as pitch accents and, if so, whether they are in a downstep relationship with each other. From a phonological perspective, it is a matter of debate whether the postfocal constituents retain their default prosodic phrasing or whether the phrasing is lost as a consequence of focus.[1] The latter approach has been defended by Liberman and Pierrehumbert (1984) as well as by Beckman and Pierrehumbert (1986, pp. 279, 286), who assume that full pitch accents are definitional of the presence of accent domains and intermediate phrases in English (see also Büring, 2010; Gussenhoven, 1992; Selkirk, 1995; Truckenbrodt, 1995, 2006). The competing approach claims that prosodic domains correspond to syntactic constituents, and are not destructed and/or reconstructed by an early nuclear pitch accent due to a narrow focus. This perspective is implicit in the work of more decidedly syntax-oriented prosodic works, such as Cinque (1993), Rochemont (1986) and Zubizarreta (1998), and it is also defended by Féry (2011, 2016). The current study presents an experiment on the realization of PFC in German, which provides arguments that prosodic phrases are retained in the postfocal domain.

In the autosegmental tradition of intonation research initiated by Bruce (1977) and Pierrehumbert (1980), tone sequences are assigned to prosodic domains mapped to morpho-syntactic constituents or to information structural domains. A sentence is typically mapped to an intonation phrase (henceforth ι-phrase). It was assumed by other authors, see Nespor and Vogel (1986) and Selkirk (1984, 2011) among many others, that smaller syntactic maximal projections are mapped to phonological phrases (henceforth Φ-phrases); see Jun (1998) for a helpful discussion. When all constituents of a German sentence are new, the main sentence accent, called the 'nuclear accent', is assigned to the last Φ-phrase of the sentence (Bierwisch, 1966; Féry, 2011; Jacobs, 1993; Krifka, 1984; Truckenbrodt, 2006). When one of the constituents of the sentence, especially the non-final Φ-phrase of the sentence, is focused or strongly emphasized and all other constituents are given, the focused constituent carries the nuclear accent and the postnuclear constituents are 'deaccented' (cf. Ladd, 1980, 2008).

At first glance, the effect of syntax on phrasing is simply overwritten as a consequence of information structure. This immediate effect on phrasing presupposes an intimate relationship between information structure and prosodic structure. The prosodic phrasing is subject to both syntax and information structure and, in the case of conflict, information structure wins. However, the asymmetry between pitch accents in the pre- and the postnuclear regions of the sentences casts doubt on this simple analysis. Prenuclear pitch accents and prosodic phrasing are preserved even if they are associated with given constituents (Baumann & Grice, 2006; Féry & Kügler, 2008; Féry & Ishihara, 2009). In the prenuclear part of the sentence, pitch accents are not erased as a consequence of information structure, but, at most, are slightly compressed. If pitch accents are interpreted as heads of prosodic phrases, the implication is that prosodic phrasing is not deleted because of information structure in this part of the sentence.

In Féry and Kügler (2008), preservation of pitch accents in the prenuclear part of the sentence was confirmed experimentally. In that study, 18 native speakers of German uttered a total of 2277 sentences of the same syntactic structure, with a varying number of constituents and differing word order and information structure (focus-given). Phonetic correlates of pitch accents were subject to a close investigation in all parts of the sentences. The results showed that, in all-new sentences, pitch accents were in a downstep relation with each other, except for the nuclear one, which was upstepped in approximately half of the realizations. In the remaining cases, the nuclear accent was downstepped as well. In that study 'downstep' was defined as a local pitch lowering, which results in a significantly lower scaling of high tones as compared to declination. We will extend the definition in the current study when describing the intonation contour of the postfocal area. Downstep is a local pitch lowering at specific points in the utterance that pragmatically expresses a distinctive phonological contrast to non-downstepped pitch accent realizations; downstep in this sense is a result of the metrical structure of an utterance (cf. Ladd, 2008, 75ff). Compared to declination as a gradual and global lowering of pitch over the utterance, the magnitude of downstep is independent of the temporal distance between H tones.

A further important effect found in Féry and Kügler's study was that the scaling of high tones, and thus the entire melodic pattern, was influenced by information structure: focus raises tones while givenness lowers them in the prenuclear position and substantially suppresses them or cancels them postnuclearly. In addition to effects due to syntactic and information structure, some purely tonal effects were identified: firstly, an anticipatory effect from L* to a following H tone, and, second, H-raising of the last H* of the sentence, a dissimilatory effect also described by Laniran and Clements (2003) for Yoruba, and Xu (1997) for Mandarin Chinese. The changes in the F0 scaling of accents were explained by the influence of information structure on the reference lines of prosodic domains. In this first experiment, downstep in the postnuclear region was not addressed because the dataset had not been designed with this aim. However, an impressionistic inspection of the data, as well as a theoretical interest for the postfocal region, led us to conduct the follow-up experiment.

The present study is a follow-up of Féry and Kügler (2008) that we conducted in order to elucidate the phrasing of the postnuclear region of a sentence. Either all pitch accents as well as all correlates of prosodic phrasing are erased in the postfocal region of the sentence, or prosodic phrasing is maintained, but the pitch accents accompanying the heads are extremely reduced due to the enormous compression of the register in this part of the sentence. In both cases, the asymmetry between pre- and postfocal givenness must be accounted for. There is a conspicuous difference between accents on pre- and postfocal given arguments. Prefocal given material can be realized with relatively high prenuclear pitch accents. The postnuclear material, by contrast, is extremely compressed. A number of authors have already commented on the asymmetry between pre- and postfocal givenness in English, for instance Ladd (2008) and Wagner (2005); Xu & Xu (2005) proposed to divide English sentences into three parts: pre-focused, focused and post-focused domains. This is similar to the 'British School' analysis of intonation distinguishing between head, nucleus and tail (e.g., O'Connor & Arnold, 1961).

The remainder of this paper is organized as follows. In the following section, the theoretical issues involved in PFC in German are given a formal account, allowing the formulation of hypotheses as to the phonetic realization of the postfocal region of the sentence. The third section introduces the experiment and its methodology, and the fourth section presents the results. The final section provides answers to the questions and hypotheses raised by the theoretical framework, based on the experimental data.

2 Theoretical background

As already mentioned in the preceding section, we follow mainstream understanding in assuming that prosodic phrases (Φ-phrase) correspond to syntactic constituents in German (Bierwisch, 1966; Féry, 2011; Kratzer & Selkirk, 2007; Krifka, 1984; Truckenbrodt, 1995, 2007, among others). A syntactic maximal projection XP, in particular a sequence predicate + adjacent argument, forms a Φ-phrase. Every additional argument, as well as every adjunct, is phrased independently, in a separate Φ-phrase, at least if they contain lexical material.

When every constituent is new in the sentence, every Φ-phrase has its own pitch accent. The whole sentence forms an ι-phrase, which has different correlates from those of Φ-phrases, such as

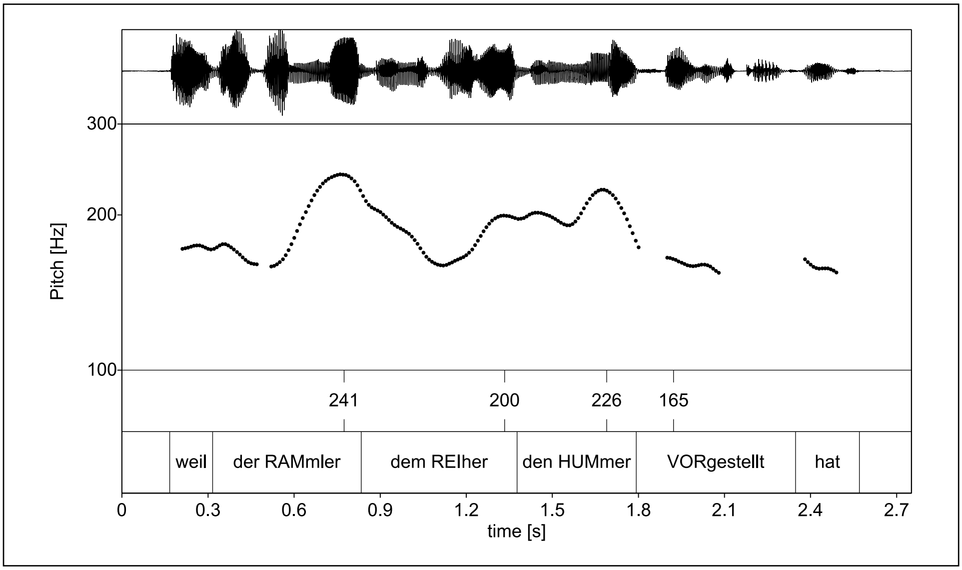

Figure 1. An all-new sentence with three arguments (Nominative, Dative, Accusative, Verb) and an upstepped nuclear accent (from Féry & Kügler, 2008); the figure displays the sound wave, pitch contour, annotation of F0 maxima in Hertz and a transcription of syllables (capitals indicate pitch accented syllables).

final lengthening, optional upstep on the nuclear accent and an additional final boundary tone at the level of the ι-phrase. An example of the prosodic structure of an all-new sentence from the experimental material of Féry & Kügler (2008) is given in (1) with upper case indicating stressed syllables that are associated with pitch accents.

(1) {Why were the animals happy?}
((Weil der RAMMler)$_{\Phi}$ (dem REIher)$_{\Phi}$ (den HUMmer vorgestellt hat)$_{\Phi}$)$_{\iota}$
because the buck.NOM the heron.DAT the lobster.ACC introduced has
'Because the buck introduced the lobster to the heron.'

Figure 1 shows a realization with an upstepped nuclear accent on *Hummer* 'lobster.' The preceding two arguments are in a downstep relationship with each other.

Furthermore, a focused constituent tends to be aligned with the right-hand edge of a prosodic phrase. This is formulated by reference to an optimality theoretic constraint ALIGN-FOCUS-R (Féry, 2013), a constraint inserting a boundary to the right of a focused constituent if there is none by default.[2] The constraint ALIGN-FOCUS-R appears in (2).

(2) ALIGN-FOCUS-R, ι-PHRASE-R (ALIGN-FOC-ι-R):
Align a focus with the right boundary of an intonation phrase.

The pitch accent of the focused constituent forms the head of a phonological phrase. As a result of (2), the pitch accent of the focused constituent becomes the nuclear pitch accent of the intonation phrase and the postfocal material is deaccented, as illustrated in Figure 2 with sentence (3), from Féry and Kügler (2008). Subscripted F indicates focus.

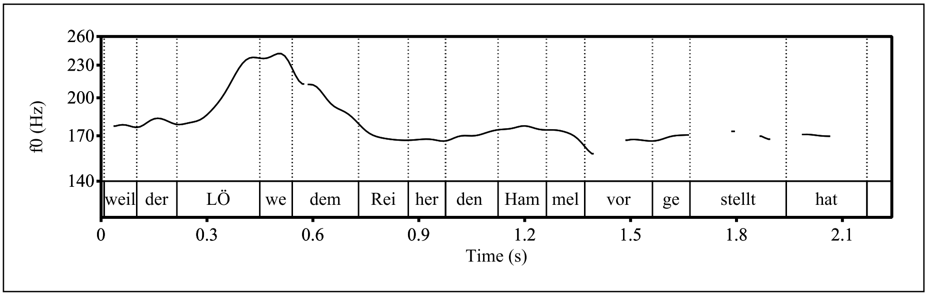

Figure 2. An example pitch track of a sentence with a narrow focus on the first argument (the subject) and postfocal material (from Féry & Kügler, 2008).

(3) Weil der Löwe$_F$ dem Reiher den Hammel vorgestellt hat.
'Because the lion introduced the sheep to the heron.'

The data investigated in the present paper do not have the unmarked word order of the sentences in Féry and Kügler (2008). In order to elicit postfocal material, sentences like (4) with a fronted participle were used. According to the regular phrasing principles postulated above for all-new sentences, the prosodic phrasing of such sentences is as follows: the participle is phrased together with the adjacent argument, in this case the subject. The other arguments are phrased individually.

(4) ((Vorgestellt hat der HUMMER)$_{\Phi}$ (den REIher)$_{\Phi}$ (dem HAMMEL)$_{\Phi}$)$_{\iota}$
introduced has the lobster.NOM the heron.ACC the sheep.DAT
'The lobster introduced the heron to the sheep.'

Such sentences were elicited in a context in which only the participle was focused and all arguments were given and thus postfocal; see (5) and (6) below.

The question that we address here is how the postfocal arguments are phrased in German. There are some options that we discuss here. The first option, illustrated in (5), consists in deleting all Φ-phrase boundaries in the postfocal area. If all the postfocal material is included in the Φ-phrase containing the focus of the sentence, the head of the unique Φ-phrase is aligned to the right of the Φ-phrase, because no other head intervenes up to the end of the Φ-phrase (see Truckenbrodt, 1995).

(5) {Did the lobster show the heron to the sheep?}
Nein. (VORgestellt$_F$ hat der Hummer den Reiher dem Hammel)$_{\Phi}$

The second option, illustrated in (6), is to insert an additional Φ-phrase boundary to the right of the focused constituent. When the focus is early in the sentence, it is separated from the remainder of the sentence with a Φ-phrase boundary. All arguments may remain phrased as in the default option in (4).[3]

(6) {Did the lobster show the heron to the sheep?}
Nein. (VORgestellt$_F$)$_{\Phi}$ (hat der HUMMER)$_{\Phi}$ (den REIher)$_{\Phi}$ (dem HAMMEL)$_{\Phi}$

In the following, we only compare these two possibilities that fulfil exhaustive phrasing at the Φ-phrase level and ALIGN-FOCUS-R. At first glance, the first option may appear to be the best one, as it deletes all phrases and all pitch accents in the postfocal domain and thus seems to account for the PFC in a straightforward way. The second option preserves the phrasing mapped to syntax, and thus presupposes that each Φ-phrase has a pitch accent. In order to understand the distribution of pitch accents more precisely, one component still needs to be added to the model, namely the metrical grid regulating relational prominence among the pitch accents. The metrical grid expresses metrical prominence in a purely relational manner: a grid mark at a higher level of the prosodic hierarchy means a higher metrical prominence than a grid mark at a lower level. A higher metrical prominence is assumed to result in higher prosodic prominence. Relational prominence has two sources. Firstly, morpho-syntax determines default phrasing with default pitch accents. In the present article, a conservative view of the relationship between prosodic domains and grid marks is assumed: every ω-word projects a grid mark at the level of the ω-word, the lowest level of the prosodic hierarchy considered here. Every Φ-phrase projects a grid mark at the level of the Φ-phrase, on the rightmost ω-word of the domain, as illustrated in (7). Finally, in an all-new sentence, nuclear stress is assigned to the rightmost Φ-phrase in the ι-phrase. The prosodic structure and metrical grid in all-new sentences are illustrated in (7).

	(			x)	ι-phrase
	(	x)	(x)	(x)	Φ-phrase
	(x)	(x)	(x)	(x)	ω-word
(7)	(vorgestellt	hat der Hummer)$_{\Phi}$	(den Reiher)$_{\Phi}$	(dem Hammel)$_{\Phi}$	

The second source of relational prominence comes from information structure. Focus and givenness can change the prominence relationship among pitch accents and this change is reflected in the metrical grid. A focused constituent gets the highest grid mark in its domain, which is always the whole sentence in the cases considered in this paper (see Selkirk, 1995; Truckenbrodt, 1995, among others). In the region of the sentence coming after the grid mark assigned on the focus constituent, no other grid mark can be assigned at the same level–the level of the ι-phrase. This means that the focused constituent always carries the nuclear accent, the last one of its domain, and is thus the most prominent element prosodically.

Returning to the two possible phrasing options that were illustrated in (5) and (6), we now augment them with metrical grids. In Option 1, shown in (8), given material in the postfocal area has no prosodic structure of its own above the level of the ω-word. It forms a single prosodic phrase with the focused and accented word, the participle. There is a single Φ-phrase and the focused constituent has the single pitch accent of the sentence.

	(x			)	ι-phrase
	(x			)	Φ-phrase
	(x)	(x)	(x)	(x)	ω-word
(8)	(VORgestellt$_{F}$	hat der Hummer	den Reiher	dem Hammel)$_{\Phi}$	

In Option 2, illustrated in (9), the postfocal material is phrased independently at the level of the Φ-phrase and the postfocal given material has a prosodic structure of its own. The projection of the focus licenses the metrical prominence at the ι-phrase level.

	(x			)	ι-phrase
	(x)	(x)	(x)	(x)	Φ-phrase
	(x)	(x)	(x)	(x)	ω-word
(9)	(VORgestellt$_{F}$)$_{\Phi}$	(hat der HUmmer)$_{\Phi}$	(den REIher)$_{\Phi}$	(dem HAmmel)$_{\Phi}$	

In summary, Option 1 assumes that pitch accents and phrasing are deleted in the postfocal region, and Option 2 predicts that the default phrasing and all pitch accents are mostly kept, albeit with changed prosodic prominence.

Usually, the pitch register is characterized by declination of register top and bottom lines, that is, a gradual lowering of pitch throughout the intonation phrase (e.g., Bruce & Gårding, 1978; Ladd, 2008). As a gradual effect, we expect declination also to take place after a focus, that is, within the compressed pitch register. As opposed to declination, downstep represents an F0 lowering that targets lower pitch scaling than one would expect for declination and, more importantly, the pitch lowering of downstep is independent of the temporal distance between H tones. Since downstep is a regular pattern in an all-new sentence in German (cf. Féry & Kügler, 2008) and pitch accents, whether downstepped or not, represent the head of a Φ-phrase, we expect these pitch accents in postnuclear position also to be downstepped if the postfocal phrasing is intact (cf. (9)).

Other phrasing options are not compatible with exhaustive parsing and we do not consider them as serious alternatives to the phrasing options given in examples (8) and (9). One of them is to assume Φ-phrases without heads (without pitch accents). This is problematic because German does not have much further evidence for the existence of Φ-phrases apart from pitch accents as heads (see Hyman, 2006, and his notion of culminativity, applicable to all languages with lexical stress, of which German is a prototypical example). Such headless Φ-phrases have been assumed for other languages, as for example for Chichewa by Samek-Lodovici (2005) and for Japanese by Ishihara (2011). However, these languages have other correlates of Φ-phrases–as for instance penultimate lengthening in Chichewa or a high tone on the second mora of each prosodic phrase in Japanese–and prosodic prominence is thus optional or even non-existent for them. Further options are non-exhaustive phrasing, extrametrical phrasing or recursive phrasing. Non-exhaustive or extrametrical phrasing in the postfocal region implies that Φ-phrases are irrelevant in the postfocal region. This option is not really different from the phrasing option given in example (8). Recursive phrasing can also be associated with (8), or be a variant of (9). We return to these options in the discussion section.

Postfocal prosodic prominences have been addressed in intonational phonology in terms of phrase accents (Grice, Ladd, & Arvaniti, 2000) and are also discussed in the case of second occurrence focus (SOF). Grice et al. (2000) argued for the existence of postfocal prosodic prominences and proposed to model these in terms of a phrase accent that captures the F0 contour between the last (nuclear) pitch accent and the final intonation phrase boundary tone. Their argument was based on an analysis of the Eastern European Question Intonation that contains a low pitch accent (L*), a low boundary tone (L%) and a high tone in between, of which the exact location depends on the language or variety under consideration, often a stressed syllable. The fact that, in some languages, phrase accents are associated with stressed syllables let Grice et al. conclude that although the primary association of a phrase accent is the structural position of an intonation phrase boundary, the secondary association to a metrically stressed syllable does not result in fully fledged pitch accents. This last observation also holds for the postfocal prosodic prominences discussed in the present paper.

Another kind of study that has investigated postfocal prosodic prominences is related to SOF, see Beaver, Clark, Flemming, Jaeger, and Wolters (2007) for English and Féry and Ishihara (2009) and Baumann, Mücke, and Becker (2010) for German. SOF is defined as a secondarily focused constituent by means of a focus particle and located in the vicinity of a primarily focused one. The first focus receives the strongest prosodic prominence in a sentence, thus the nuclear pitch accent, and SOF is both focused and given. As a result, it does not carry the main pitch accent of the sentence. Studies on SOF found only minimal differences between simply given postfocal constituents and SOF ones. Even though it was not marked by a pitch accent, the SOF constituent had greater acoustic prominence cues as compared to a simple postfocal given constituent. Baumann et al. (2010) therefore suggested analysing the acoustic prominence cues of SOF phonologically as a phrase accent similar to Grice et al. (2000). This analysis cannot be maintained here: the SOF structures discussed in the literature differ from the kind of structures investigated here in having maximally one postfocal constituent. Our data allow for a more precise relational analysis showing the effects of postfocal prosodic prominence between one, two and three postfocal constituents.

In order to decide between the two phrasing options illustrated in (8) and (9) and to investigate postfocal prosodic prominence in more detail, we report the results of a production experiment in the next sections. The aim of the experiment is to provide arguments in favour of one or the other phrasing option. We thus formulate two mutually exclusive sets of hypotheses as to the realization of postfocal given arguments. H1 is compatible with Phrasing Option 1 in (5) and (8) and H2 with Phrasing Option 2 in (6) and (9).

(10) First hypothesis (H1, Option 1): Φ-phrases in the postfocal and postnuclear region are deleted and the postnuclear constituents–regardless of length–are included in the Φ-phrase of the last pitch-accented word, which is the participle in the sentences used here. In this case, the following predictions should be confirmed.

P1a. No increases in intensity and duration as a reflex of the presence of a pitch accent are present on the given constituents, and only intensity and duration, reflecting correlates of lexical stress, are preserved on the lexically stressed syllable of each argument.

P1b. After the focused constituent declination is observed on the postfocal constituents, the lowest F0 value of the sentence is reached at the end of the sentence.

(11) Second hypothesis (H2, Option 2): Φ-phrases in the postfocal and postnuclear region are preserved, even though the register used is extremely compressed as compared to the new part of the sentence. In this case, the following predictions should be confirmed.

P2a. The individual postfocal arguments are pitch-accented, albeit in a narrow pitch range.

P2b. The individual pitch accents are downstepped to each other, that is, F0 scaling is locally determined and not gradually declining as it was due to declination. The lowest F0 value of the sentence is reached at the end of every declarative sentence, regardless of the number of arguments following the focused participle.

The next section presents an experiment aimed at testing the hypotheses.

3 Experimental procedure

3.1 Speakers

Eleven female native speakers of German participated in the experiment. All were undergraduate students at the University of Potsdam speaking the standard variety of German spoken in the Berlin-Brandenburg region. None of them reported any speech or hearing impairments. They either received course credit or were paid for their participation.

3.2 Speech materials

In order to test the prosodic realization of postfocal constituents, test sentences were constructed with a sentence-initial participial verb followed by one to three arguments. The test sentences were embedded into contexts that contrast the participial verb, hence the sentence-initial verb was in focus, and all following constituents were given by the context; this construction has been called focus fronting in the syntactic literature (see, for instance, Fanselow & Lenertová, 2011).[4] In the experimental sentences, focus is always the element in a sentence that answers a *wh*-question; a given constituent has been mentioned in the preceding context question, as shown in (12)–(14).

(12) A verb and a single argument

{Did the lobster end the fight?}

Nein,	angefangen	hat	der	Hummer.
no,	started	has	the.NOM	lobster

'No, the lobster started.'

(13) A verb and two arguments

{Did the heron punish the sheep?}

Nein,	eingeladen	hat	der	Reiher	den	Hammel.
no,	invited	has	the.NOM	heron	the.ACC	sheep

'No, the heron invited the sheep.'

(14) A verb and three arguments

{Did the lobster introduce the sheep to the heron?}

Nein,	gezeigt	hat	der	Hummer	den	Hammel	dem	Reiher.
no	showed	has	the.NOM	lobster	the.ACC	sheep	the.DAT	heron

'No, the lobster showed the sheep to the heron.'

Sentence length was varied as a function of the number of arguments of the verb. The short sentences contained a focused intransitive verb and a single given argument (nominative), as in (12). The medium sentence length contained a transitive verb and two given arguments (nominative and accusative), as in (13). The long sentence length contained a ditransitive focused verb, followed by three given arguments (nominative, accusative and dative), as in (14). The underlined verb was always the only focused constituent of the sentence, with a context inducing correction of the verb, followed by given arguments. The sentences were presented with the focal verb underlined, in order to minimize errors of interpretation (cf. Féry & Kügler, 2008). For each sentence length two different verbs were used: two intransitive (*anfangen* 'to begin', *anrufen* 'to call'), two transitive (*einladen* 'to invite', *besuchen* 'to visit') and two ditransitive ones (*vorstellen* 'to introduce', *zeigen* 'to show'). Because it was not possible to find highly frequent ditransitive verbs with the same number of syllables, we had to be content with one disyllabic and one trisyllabic verb. All other verbs were trisyllabic and consisted of a verbal stem and a particle.

The arguments were chosen from three nouns (names of animals: *Hummer* 'lobster', *Hammel* 'sheep' and *Reiher* 'heron'), all three trochaic with a final schwa syllable. Three tokens per verb and sentence length were constructed, resulting in six individual test sentences per sentence length (three combinations of the animal names × 2 verbs × 3 sentence lengths = 18 target sentences).

3.3 Recordings

The speakers were digitally recorded at a sampling rate of 22.05 kHz with a 16-bit resolution in a soundproof booth at the University of Potsdam. Participants answered questions that elicited corrective focus on the participial verb (cf. (12)–(14)).

The material was presented on slides in a PowerPoint presentation. Each sentence was presented in two steps. In the first step, a slide served to introduce the context. It contained a question presented both visually and acoustically over headphones. The context questions were Yes–No Questions realized with a final rise, and inverted verb–subject syntax. After understanding the question, the speakers pressed the return key and a target sentence appeared on the screen. The participants were asked to read the sentence as naturally as possible, respecting the previous context. The rate of the presentation was self-paced, and participants could correct themselves if they thought that their production was incorrect or unnatural. The sentences were pseudo-randomized so that the participants had to concentrate on the context, since it varied in each sentence. This procedure was used to avoid monotony.

Each sentence length condition was intended to come in six tokens, with different verbs and names of animals in each case. However, only five tokens appeared in the two-argument conditions because of a mistake in the design. Most of the 11 speakers assigned pitch accents as expected: they produced the postfocal material in an extremely reduced pitch register. Some speakers did not considerably reduce the register in the postfocal material, but instead pronounced the sentence with full pitch accents on the postverbal arguments, as if the sentence had been presented in an all-new context. When this happened, it was always in sentences with three postfocal arguments. These realizations were retained in the analysis, and this had no influence on the significance of the results.

Altogether 187 sentences were recorded and retained for analysis (11 speakers × 3 sentence lengths × 6 tokens–11 two-argument sentences (one missing token) = 187).

3.4 Analysis

The recorded sentences were analysed using the acoustic speech analysis software Praat (Boersma & Weenink, 2013). The recordings were partly automatically and partly manually divided into labelled substrings with the help of spectrograms and acoustic inputs. Obvious errors due to the F0 algorithm (for instance, octave jumps) were corrected by hand, and the contours were smoothed using the Praat smoothing algorithm (frequency band 10 Hz) to minimize micro-prosodic perturbations.[5] All frequency measurements were performed semi-automatically using a script that detects the highest and surrounding lowest F0 values within a given domain. The domains for measurements were the participle and each noun, for example, *eingeladen*, *Reiher* and *Hummer* in (15). In this example '#' stands for a word boundary, and boldface for a measurement domain.

(15)	# **Eingeladen**	# hat # der	#	**Reiher** # den	#	**Hummer**	#
	# invited	# has # the.NOM	#	heron # the.ACC	#	lobster	#
	'The heron invited the lobster.'						

The analysis was done in a number of steps. Firstly, a Praat script located F0 maxima as well as F0 minima to the left and right of the F0 maximum in each target domain. In (15), there are three domains and thus three F0 maxima and six F0 minima. Secondly, the results of the Praat script were hand-edited to correct spurious labelling. Both authors individually evaluated the F0 labels against the F0 track, the substring divisions, an auditory impression and the spectrogram. Where the Praat script had assigned an F0 maximum or minimum label that was not in the correct position (because of obvious errors due to the algorithm), the label was manually moved. In the third step, another Praat script recovered the F0 values at the positions of the F0 labels, characterizing each value as either F0 maximum or minimum, and compiled them in a table.

On the basis of the F0 measurements four different calculations were performed.

(i) To obtain the intra-domain F0 range of the individual arguments (prediction P1a/P2a), the F0 rise was calculated from the F0 minimum to the F0 maximum of each argument; the F0 fall was calculated from the F0 maximum to the following F0 minimum of each argument. These measurements reflect the absence or presence of F0-induced prominence. If postfocal deaccentuation deletes all pitch accents, an F0 range near zero is expected. If, on the other hand, the F0 range shows values well above just noticeable differences (JNDs) of approximately 3 Hz (cf. Kollmeier, Brand, & Meyer, 2008, p. 65), plus a buffer of intrinsic F0 of the vowels of about 4 Hz, this measure is an indicator of postfocal prominences.

(ii) To assess the presence of the downward trend in F0 after the focused verb, the F0 minimum after the F0 maximum of the participial verb and that of the final argument of the sentence were compared.

(iii) To decide whether declination only or declination and downstep occur in the postfocal area, the time between the F0 peaks and the amount of F0 drop between the F0 peaks was correlated.

(iv) To evaluate the scaling of the F0 maxima in a cross-sentence comparison, the F0 heights of the first and second arguments were calculated as a function of the number of arguments; in a within-sentence comparison, the scaling of each argument was compared to the scaling of the following argument to assess whether the arguments were downstepped relative to each other or not.

For statistical analyses, we used R (R Core Team, 2013) and lme4 (Bates, Maechler, Bolker, & Walker, 2013) to perform a linear mixed-effects analysis of the relationship between F0 scaling and sentence length in terms of the number of arguments. Linear mixed-effects models were fitted that relied on the dependent variable ‘F0 value.’ The dependent variable contained measurements of the verb and all arguments, and the actual measurement of ‘F0 value’ depended on the three expectations of the hypothesis. For the analysis of the position of the lowest value of the sentence, F0 minimum represents the ‘F0 value’; for the analysis of F0 range on the postfocal constituents, F0 rise and fall (i.e., the difference between F0 maximum and F0 minimum) represent the ‘F0 value’; for the analysis of scaling, F0 maximum represents the ‘F0 value.’ All dependent measures were log-transformed to meet the assumption of the regression model. The models applied crossed random factors ‘speaker’ and ‘item’, and ‘position of argument’ or ‘argument type’ as fixed factors. Random slopes (Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013) for speakers and items were integrated into the models assuming that differences exist for each speaker’s individual pitch range, and that each speaker realized the postfocal F0 values in a slightly different way. Backward modelling (Barr et al., 2013) of random slopes for speaker and item revealed no significant differences between the models, suggesting that the lesser complex model fits the data more precisely, which in our case is the model with random slopes for speakers only. For some comparisons specified below in the respective result sections we ran mixed models with random intercepts for speakers and items only. The reason for omitting random slopes of speakers is that the models including random slopes were over-parameterized, referring to the correlation of the random effects near or exactly as 1. The assumption then is that speakers realize differences in the postfocal F0 values in the same way, and there is indeed no reason to assume that speakers vary in their realizations.

For illustration purposes, the obtained Hz values were aggregated within each participant and each condition. The figures and tables below show the aggregated scores per condition averaged across speakers.

4 Results

The presentation of the results is divided into three parts. Firstly, the results for all data are presented. Secondly, since we observed a considerable amount of variation of postfocal F0 realization, we decided to divide the data into a group that showed a clear downward trend above 7 Hz difference and another group showing no downward or even an upward trend in F0. The results for these two groups are presented subsequently.

4.1 All data

Figure 3 presents the normalized and averaged pitch contours of all the data collected in the experiment; thus, 66 sentences with one argument, 55 sentences with two arguments and 66 sentences with three arguments (recall from Section 2 that one token was missing in the sentences with two arguments). A visual inspection of Figure 3 reveals that postfocal arguments were indeed realized in a compressed pitch register, as compared to the scaling of the nuclear H*L pitch accent on the contrastively focused verb. The F0 fall of the nuclear accent resembles that established in Féry and Kügler (2008). Visual inspection of Figure 3 also reveals that the sentences reached the same final point in all three cases (about 166 Hz), but only at the end of the sentence.

To test our first prediction about the presence or absence of postfocal pitch accents we calculated the F0 range within each of the nouns (the difference between F0 maximum and F0 minimum; see Figure 3). Table 1 presents the average F0 rise towards an F0 peak on the verb and each of the following arguments in the sentence. The lowest F0 rises with means of 5.0, 6.2 and 7.3 Hz were found in the last argument of each sentence independently of sentence length. In non-final arguments, however, the rise was larger: in the three-argument sentences, average rises of 12.5 and 18.2 Hz were found in the first and second arguments, respectively. That is, the F0 rise increased for three-argument sentences. The F0 fall from the F0 peak of the arguments presented in Table 2 was in general even larger, with means ranging from 13.0 to 20.6 Hz. The F0 differences in general are well above the JNDs (Kollmeier et al., 2008). The analysis of F0 range indicates that in all sentences, postfocal prominences caused by F0 were realized. Visual inspection of the data also reveals that the F0 peaks were located on the stressed syllables in each case. Hence, the individual arguments were pitch-accented, albeit in a narrow pitch register, which confirms hypothesis P2a.

To assess the intrinsic F0 differences between vowels, we calculated the mean F0 rise and mean F0 fall separately for each target word. Table 3 presents the results for each target word in all sentence positions. The effect of intrinsic F0 is visible such that the rise in [hʊm] (Hummer) and [ham] (Hammel), for instance, differs from 2 to 5 Hz on average, and the fall between 3 and 4 Hz on average. We may interpret this difference as a microprosodic effect between low /a/ and high /u/. This effect is consistent through the data (see Table 3), but it does not affect the general pattern of rises and falls.

To test our second prediction about declination and downstep in the postfocal domain, we used different measures. Firstly, we calculated the difference between the F0 minimum after the maximum of the verb and the final F0 minimum in the final constituent; the particular F0 minima were entered as the dependent variable in the statistical model. These aggregated means are displayed in Table 4. Fitting a linear mixed model with 'position of F0 minimum' and 'number of arguments' as fixed factors and 'speaker' and 'item' as random factors revealed a significant effect of 'position of F0 minimum' (cf. Table 5), showing that the sentence low was reached only at the very end of each sentence and not after the focused constituent (cf. Table 4 for aggregated means). The model also revealed a significant effect for 'number of arguments' (cf. Table 5), showing that one-argument sentences differed both from two-argument sentences and from three-argument sentences (second and third contrast in Table 5). In addition, both interactions were significant, showing that each of the individual comparisons of F0 minimum position per number of arguments differed significantly, which is also shown in the interaction plot in Figure 4. Note that the fact that the F0 minimum after the verb was highest for two-argument sentences and not for three-argument sentences (cf. Table 4) is spurious and does not bear on the point that the low point of the sentence was reached only at the very end of the sentence. Taken together, the lowest value of the sentence was reached only at the end of every declarative sentence, regardless of the number of postfocal arguments, and not directly after the focused constituent. This result shows that some

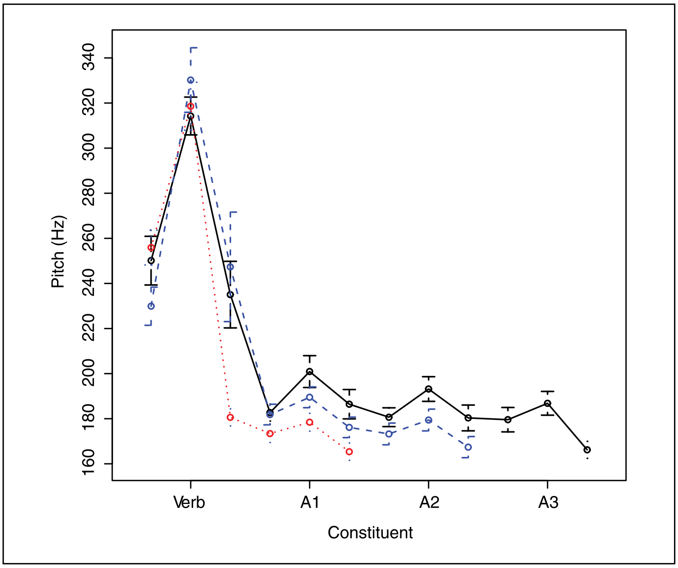

Figure 3. Comparison of the averaged values of the three sentence lengths for all data. Sequence of F0 minimum, F0 maximum and F0 minimum in Hertz per constituent measured on the verb or noun (*x*-axis) for one-argument (dotted line), two-argument (dashed line) and three-argument sentences (solid line).

Table 1. F0 range–average F0 rise in Hertz to F0 peak and standard deviation in parenthesis within the corresponding domain.

Arguments	V	A1	A2	A3
1	62.7 (41.1)	5.0 (4.3)		
2	100.3 (62.4)	7.7 (6.5)	6.2 (5.2)	
3	64.1 (47.8)	18.2 (24.5)	12.5 (14.4)	7.3 (8.8)

Table 2. F0 range–average F0 fall in Hertz from F0 peak and standard deviation in parenthesis within the corresponding domain.

Arguments	V	A1	A2	A3
1	138.0 (44.5)	13.0 (7.0)		
2	82.9 (75.1)	13.3 (5.7)	12.0 (6.3)	
3	79.2 (67.6)	14.5 (8.4)	12.8 (7.5)	20.6 (15.0)

F0 downtrend after the focus occurs and that the final low is reached constantly at the minimum of a speaker's range in German also in the case of an earlier focused constituent (Féry & Kügler, 2008; Grabe, 1998; Truckenbrodt, 2002, 2007).

Secondly, we computed a Pearson product-moment correlation coefficient to assess whether there is a correlation between the size of the downward F0 step between F0 peaks and the temporal distance between the two peaks. Given that declination is a continuous effect over time we expect a positive correlation between the size of the downward F0 step and the temporal distance between

Table 3. Mean F0 rise and mean F0 fall in Hertz on the individual target words in each position of the sentence averaged across speakers.

Arguments		Rise			Fall		
		A1	A2	A3	A1	A2	A3
1	Hummer	6.7	–	–	17.1	–	–
	Reiher	3.5	–	–	8.9	–	–
	Hammel	4.8	–	–	13.2	–	–
2	Hummer	11.3	10.1	–	16.1	14.7	–
	Reiher	2.8	2.3	–	8.2	9.0	–
	Hammel	6.4	8.1	–	13.1	13.8	–
3	Hummer	20.5	18.1	13.5	19.3	16.9	26.1
	Reiher	18.3	6.0	2.0	9.0	8.3	14.6
	Hammel	15.8	13.5	6.3	15.1	13.2	21.2

Table 4. F0 minimum in Hertz and standard deviation in parenthesis (i) at the end of the verb (L1), and (ii) the end of the sentence (Final Low), aggregated over speakers, items.

No. of arguments	L1 (Hz)	Final Low (Hz)
1	181 (15.4)	166 (15.8)
2	247 (89.8)	168 (17.6)
3	235 (60.1)	166 (15.4)

Table 5. Linear mixed-effects model with *speaker* and *item* as random factors, and *position of F0 minimum* and *number of arguments* as fixed factors. * = significant at the $p < 0.05$ level.

Contrasts	*SE*	*t*	Sign.
F0 minimum verb–sentence-final F0 minimum	0.02584	−3.38	*
one argument–two arguments	0.02739	9.85	*
one argument–three arguments	0.02584	9.12	*
Interaction position F0 min–one/two arguments	0.03833	−6.47	*
Interaction position F0 min–one/three arguments	0.03655	−6.32	*

the two peaks. On the other hand, the absence of such a correlation would show that the downward step is categorical and thus local and independent of temporal distances. In this case, we would interpret the downtrend as downstep. Figure 5 displays the correlation plot, which shows that there is no such correlation between the downward step of the peaks and their time distance. The solid line in Figure 5 shows the correlation for the two-argument sentences, and the dashed grey line shows the correlation for the three-argument sentences, including both the downward step and time distance between the first and second argument and between the second and third argument. Both correlations are not significant (two-argument sentence, $r = 0.098$, $p > 0.05$; three-argument sentence, $r = 0.148$, $p > 0.05$). We can thus conclude that the downward steps can be interpreted as local effects and thus as downstep.

Thirdly, we tested the scaling of F0 maxima in postfocal constituents, applying two different calculations: firstly a cross-sentence comparison of the individual arguments and secondly a within-sentence comparison of adjacent arguments in two- and three-argument sentences.

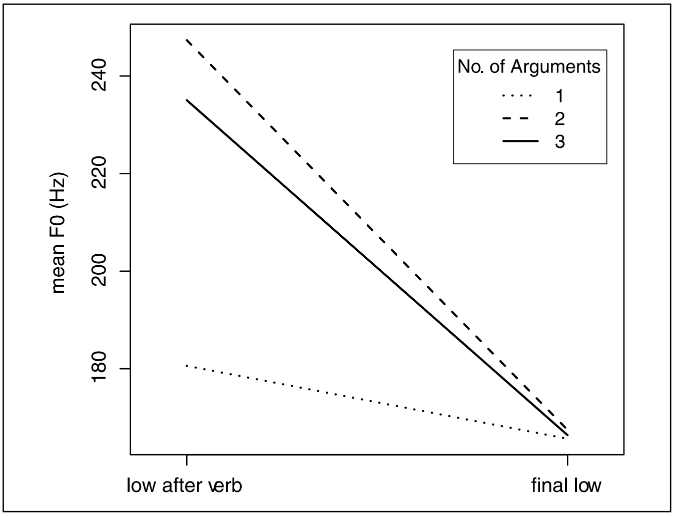

Figure 4. Interaction plot of the linear mixed-effects model with treatment contrasts comparing the F0 minimum after the verb (left-hand side) with the sentence-final F0 minimum (right-hand side) for each sentence length (one argument–dotted line, two arguments–dashed line, three arguments–solid line).

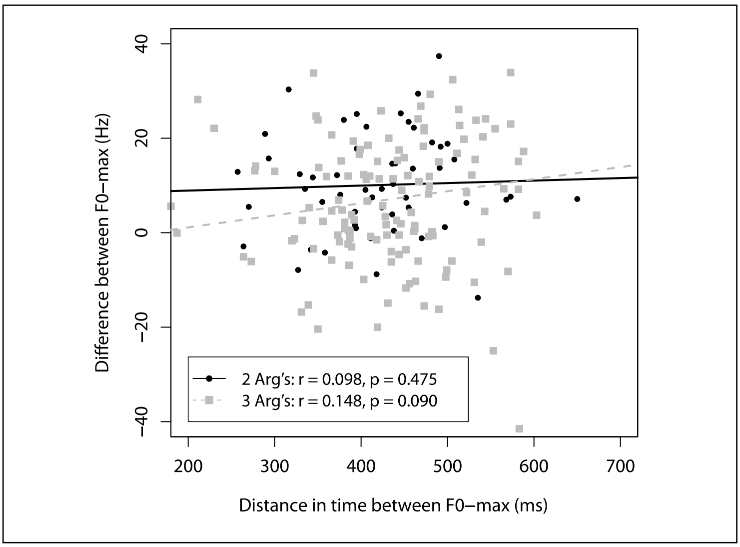

Figure 5. Correlation plot between the downward step of the F0 peaks and the distance in time between the F0 peaks split by sentence type; the solid black line represents the data point for two-argument sentences and the dashed grey line for three-argument sentences.

Figure 6 displays the F0 maximum of each postfocal argument and illustrates downstep-like F0 lowering across the arguments in all data points. Each point stands for the averaged highest F0 level (i.e., F0 maximum) of the arguments. It can be seen that the height of the first argument's (A1's) F0 maximum depends on the number of arguments following it. The first argument's F0 value was higher in two-argument sentences than in one-argument sentences, and was even higher

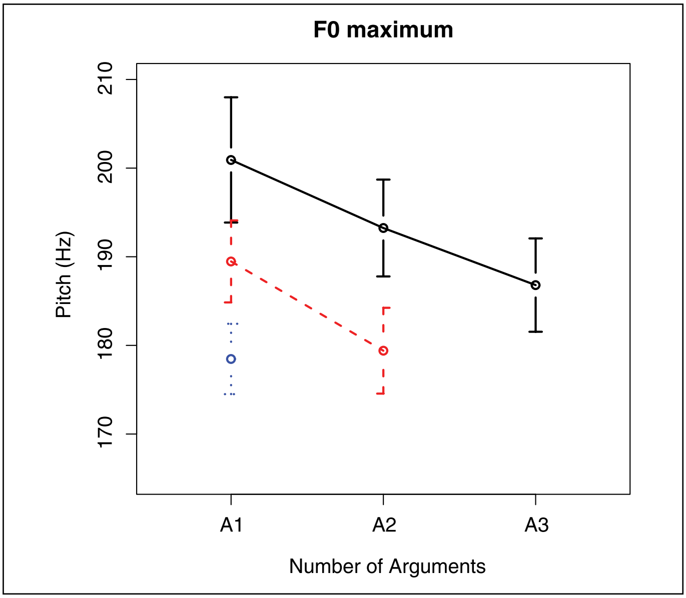

Figure 6. F0 maximum in Hertz for all data on the first, second or third argument (x-axis) for one-argument (dotted line), two-argument (dashed line) and three-argument sentences (solid line).

Table 6. Means of F0 maximum per argument and verb in Hertz with SD in parentheses.

No. of arguments	Verb	A1	A2	A3	*n*
1	319 (43)	179 (16)			66
2	330 (53)	190 (17)	179 (18)		55
3	314 (34)	200 (29)	193 (22)	187 (21)	66

than that in three-argument sentences. The last F0 maximum of the one- and two-argument sentences was largely equivalent, but was lower than that of the three-argument sentences, which was slightly lower than F0 maximum of the first argument in two-argument sentences. Table 6 shows the average F0 maximum values in each case.

An important observation to be gained from Table 6 is that in all cases, the highest tone (F0 maximum) of the first (nominative) argument was considerably lower relative to that of the preceding participle. This amounted to more than 100 Hz in all three cases, and is comparable to the postfocal final drop found by Féry and Kügler (2008).

Next, we fitted a linear mixed-effects model for F0 maxima on the first and second arguments with 'number of arguments' as the fixed factor and 'speaker' and 'item' as random factors. The results are presented in Table 7. The cross-sentence comparison revealed a significant difference for F0 maximum on the first argument as a function of the number of arguments: it was higher when the sentence was longer. The comparison between one- and two-argument sentences (first contrast in Table 7), the comparison between one- and three-argument sentences (second contrast in Table 7) and the comparison between two- and three-argument sentences (third contrast in Table 7) were all significant. The same is true when comparing the F0 maximum on the second argument as a function of number of arguments. The difference between two- and three-argument sentences also yielded significance (fourth contrast in Table 7). This result shows that the scaling of initial and later

Table 7. Treatment contrast of linear mixed-effects model for F0 maximum on first (A1) and second (A2) argument as a function of number of arguments. * = significant at the $p < 0.05$ level.

Contrasts	*SE*	*t*	Sign.
A1: one argument–two arguments	0.01130	4.83	*
A1: one argument–three arguments	0.02496	4.55	*
A1: two arguments–three arguments	0.02009	2.85	*
A2: two arguments–three arguments	0.01450	5.68	*

Table 8. Treatment contrasts of linear mixed-effects models for two-argument (2Arg) and three-argument sentences (3Arg) comparing the individual argument's scaling (F0 maximum) within a sentence. * = significant at the $p < 0.05$ level.[6]

Contrasts	*SE*	*t*	Sign.
2Arg: A1–A2	0.00832	−6.68	*
3Arg: A1–A2	0.01259	−2.87	*
3Arg: A1–A3	0.01257	−5.58	*
3Arg: A2–A3	0.01022	−3.34	*

postfocal arguments differed as a function of the number of postfocal arguments. In other words, the more arguments follow, the higher the scaling of the individual arguments.

As for the within-sentence comparison, when all values were averaged, postfocal arguments showed a clear downward step, which we interpret as downstep. The non-final arguments reached higher F0 maxima the more distant they were from the end of the sentence. The observed differences in F0 values were in an extremely compressed register, as visible in the comparison between the maximal value of the initial participle, which was on average around 320 Hz, and the maximal values of the postfocal given arguments, which on average lie between 180 and 200 Hz.

Table 8 displays the treatment contrasts of the linear mixed-effects models showing that there was a significant difference between the F0 maximum of the first and the second argument in two-argument sentences (first contrast in Table 8). There was also a significant difference between the arguments in three-argument sentences: there was downstep between the first and the second argument (second contrast in Table 8), and between the first and the third argument (third contrast in Table 8), as well as between the second and third argument (fourth contrast in Table 8). F0 scaling differed between postfocal arguments, and negative *t*-values indicate lower F0 values for the later constituents, which means that they were in a downstep relation.

To sum up this section, we showed, firstly, evidence for postfocal prosodic prominences in terms of F0 range, secondly, that the F0 scaling of postfocal arguments depended on the number of upcoming arguments, thirdly, that the difference between F0 peaks, measured on the F0 maximum in each argument, was significant in postfocal positions, which we interpret as a downstep relation between the F0 peaks and, fourthly, that the absence of a correlation between the downward steps between F0 peaks and their time distance speaks in favour of an interpretation of downstep.

4.2 Data analysis of the values with a downward step on the arguments larger than 7 Hz

During visible inspection of the data we observed considerable variation as to the realization of postfocal arguments: the amount of downward step varied and some realizations showed no clear

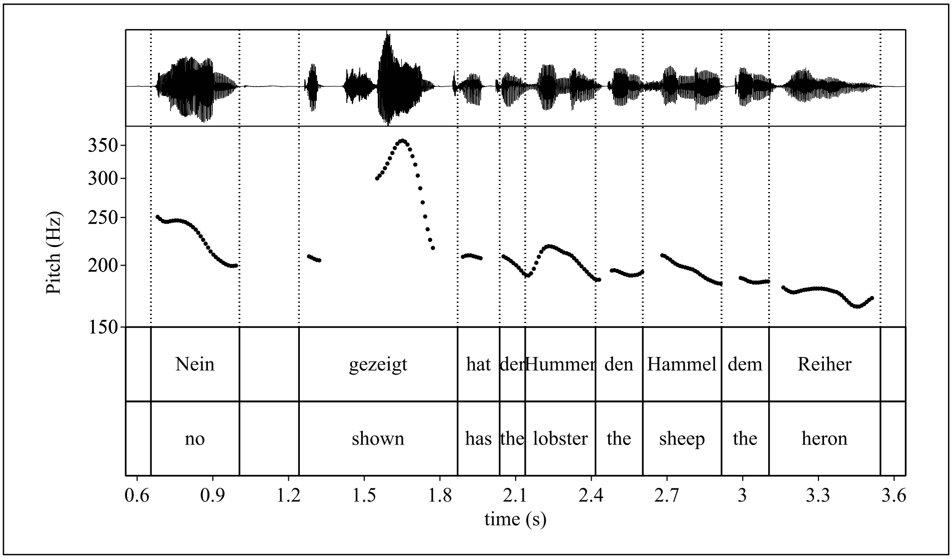

Figure 7. Pitchtrack of an experimental three-argument sentence with downstep and pitch accents on all three words.

downstep, and some even showed a small upstep (a variation that is similar to the one in prenuclear position and in all-new sentences reported by Féry & Kügler, 2008). Although the aggregated overall pattern reported in the previous section showed clear and significant F0 range and significant differences between F0 peaks on postfocal arguments, we grouped the data according to their postfocal realizations. In this section, we therefore concentrate on the data that show a clear downward step between two arguments. The threshold to divide the data was a downward step of 7 Hz between the two accents. This value corresponds roughly to 0.5 semitones, a perceivable step. This value is based on robust perceptual detectability such that 7 Hz is above the JND, including a small extra buffer that roughly includes microprosodic variation of about 2–4 Hz (see Table 3 for the differences in F0 rise and F0 fall between the target words). Sixty per cent of the two-argument sentences and a little less than 50% of the three-argument sentences fell into this category and are examined in the present section. The remaining sentences are the subject of the next section. An example of a sentence with this clear downward step is shown in Figure 7.

(16) Nein. <u>Gezeigt</u> hat der Hummer den Hammel dem Reiher.
no. showed has the.NOM lobster the.ACC sheep the.DAT heron
'No. The lobster showed the sheep to the heron.'

Table 9 sums up the absolute number of data considered in this section. One hundred per cent for A1 means that the difference between the preceding verb and A1 was always larger than 7 Hz. This was the case in one- and in two-argument sentences. The result of 97% for A1 in three-argument sentences was a consequence of the fact that there were two realizations in the data in which speakers realized a higher F0 on the argument than on the verb. Further, it can also be seen in Table 9 that 60% of the sentences in two-argument sentences and 46% in three-argument sentences showed a difference of more than 7 Hz between A1 and A2, and the difference between the second and third argument was subject to downstep in 48% of the sentences with three arguments. It is important to realize that a three-argument sentence can have downstep in one transition from

Table 9. Number of data points per argument in which its F0 peak is lower than that of the previous argument by more than 7 Hz (i.e., downstep).

No. of arguments	No. of downstepped cases (%)		
	A1	A2	A3
1	66 (100%)		
2	55 (100%)	33 (60%)	
3	64 (97%)	31 (46%)	32 (48%)

Table 10. Means of F0 maximum per argument and verb in Hertz with SD in parentheses; only cases of clear downstep (difference between arguments ⩾7 Hz) are included in the calculation.

No. of arguments	Verb	A1	A2	A3
1	319 (43)	179 (16)		
2	330 (53)	190 (17)	175 (19)	
3	316 (34)	198 (25)	194 (28)	183 (22)

an argument to another (for instance, between A1 and A2), but not in the other transition (for instance, between A2 and A3).

Table 10 shows the average F0 maximum values in each case. The clear postfocal F0 drop after the verb turns up again. Comparing the individual arguments across sentences, the difference in scaling of the arguments increased as a function of the number of arguments in a sentence. In other words, the aggregated means for A1 increased by about 11 Hz from one-argument sentences to two-argument sentences, and by about 8 Hz from two-argument sentences to three-argument sentences. For A2 the increase from two-argument sentences to three-argument sentences was about 19 Hz. A within-sentence comparison of the arguments showed that the scaling of later arguments was lower relative to the previous ones. Note that the data analysed in this section include all data points where a local downward step larger than 7 Hz occurred independent of the previous or following downward step. Thus, the averaged mean frequencies in Table 10 do not represent the averages per sentence, that is, the difference between A2 and A3 in three-argument sentences is, on average, 4 Hz only.

Figure 8 shows the F0 maximum of the data exhibiting a clear downstep according to our definition. Each point stands for the averaged F0 maximum of each argument. As in Figure 6, the first argument's F0 value was higher when followed by more arguments. The main difference between Figure 6 and Figure 8 lies in the steepness of downstep on the final argument, which is larger in Figure 8 than in Figure 6 where all F0 maximum data points were considered.

We fitted a linear mixed-effects model for F0 maxima on the first and second argument with 'number of arguments' as the fixed factor and 'speaker' and 'item' as random factors, applying random slopes also for speaker. The results are presented in Table 11. In the cross-sentence comparison, a significant difference for F0 maximum on the first argument as a function of the number of arguments could be established: in the comparison of one- versus two-argument sentences (first contrast in Table 11), in the comparison of one- versus three-argument sentences (second contrast in Table 11) and in the comparison of two- versus three-argument sentences (third contrast in Table 11). The comparison of F0 maximum on the second argument as a function of the number of arguments between two- and three-argument sentences also yielded significance (fourth contrast in Table 11).

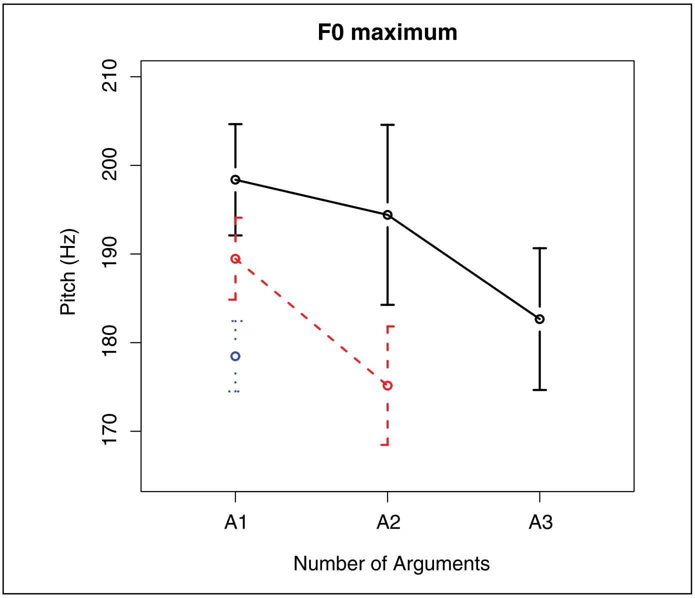

Figure 8. F0 maximum in Hertz on first (A1), second (A2) and third arguments (A3) for one-argument (dotted line), two-argument (dashed line) and three-argument sentences (solid line); data include values that have a difference of at least 7 Hz between two adjacent arguments (= downstep).

Table 11. Treatment contrast of linear mixed-effects model for F0 maximum on first (A1) and second (A2) argument as a function of number of arguments; clear downstep cases. * = significant at the $p < 0.05$ level.

Contrasts	*SE*	*t*	Sign.
A1: one argument–two arguments	0.01088	5.07	*
A1: one argument–three arguments	0.02078	5.13	*
A1: two arguments–three arguments	0.01591	3.1	*[7]
A2: two arguments–three arguments	0.01483	8.46	*

Similar to the cross-sentence comparison of all data points (previous section), this result shows that the F0 scaling of initial and later postfocal arguments differed as a function of the number of postfocal arguments. In other words, the more arguments follow, the higher the scaling of the initial arguments to realize subsequent downstep on following arguments.

In the results for all points in the previous section, it could be shown that downstep always applied between the verb and the first argument, and it could also be shown that, taking all data together, all intra-sentence maxima values were significantly downstepped relative to each other. Obviously, this result is strengthened when considering only the values with downward step of at least 7 Hz. We fitted linear mixed-effects models, again with ‘speaker’ and ‘item’ as random factors, applying random slopes for speakers, and number of arguments as the fixed factor. Significance was amplified in all relevant comparisons: between the first and the second argument in two-argument sentences (first contrast in Table 12), between the first and second arguments (second contrast in Table 12), between the first and third arguments (third contrast in

Table 12. Treatment contrasts of linear mixed-effects models for F0 maximum in two-argument (2Arg) and three-argument sentences (3Arg) comparing the individual argument's scaling within a sentence. * = significant at the $p < 0.05$ level.

Contrasts	SE	t	Sign.
2Arg: A1–A2	0.00831	−10.22	*
3Arg: A1–A2	0.01520	−3.17	*[8]
3Arg: A1–A3	0.01484	−6.74	*
3Arg: A2–A3	0.01677	−5.9	*

Table 13. Number of data points per argument in which the difference from the previous constituent is at least 7 Hz (upstep).

No. of arguments	No. of upstepped cases (%)		
	A1	A2	A3
2		3 (5%)	
3	2 (3%)	8 (12%)	10 (15%)

Table 12), as well as between the second and third arguments (fourth contrast in Table 12) in three-argument sentences.

To sum up this section, the data with a downward step of at least 7 Hz showed a similar picture to the all-point data. The cross-sentence comparison showed an increase in scaling as a function of the number of arguments in the sentence. The within-sentence comparison revealed downstep of the individual arguments.

4.3 Analysis of the remaining data

Turning now to the remaining data, 40% of the two-argument sentences and a little more than 50% of the three-argument sentences had at least one transition from one to the next argument with no clear downstep. This set of cases is divided into two groups for further discussion: a smaller group in which upstep occurred, and a larger group of 'flat' realizations[9] in which the difference in F0 maximum between two consecutive arguments was less than 7 Hz.

We turn first to the smaller group with upstep. The threshold for grouping the data as 'upstep' was set, similar to that of downstep, as an increase of at least 7 Hz in the transition between two arguments. In the three-argument sentences, 3% of the transitions between the verb and the first argument showed upstep. Leaving this 3% aside, which resulted from mistakes in the interpretation of the contexts (see above), we focus on the transitions between arguments showing upstep. In the two-argument sentences, it was of course always the shift from the first to the second argument that could be upstepped, but in the three-argument sentences, it could be the shift between the first and the second, or between the second and the third argument.

In Table 13, it can be seen that 5% of the transitions from the first to the second argument in sentences with two arguments showed upstep; in sentences with three arguments, 12% of the transitions between the first and the second argument, and 15% between the second and the third argument showed upstep, altogether 23 transitions. Note that the frequency of upsteps increased with sentence length, that is, the longer the postfocal area the more likely an upstep. Because of the small number of cases of upstep, no statistical analysis could be calculated for these data.

Table 14. Number of data points per argument in which the difference from the previous argument is between −6 and +6 Hz (i.e., no downstep and no upstep).

No. of arguments	No downstep, no upstep		
	A1	A2	A3
2		19 (35%)	
3		27 (41%)	24 (36%)

Table 15. F0 range–average F0 fall and F0 rise in Hertz from or to F0 peak within the corresponding domain for remaining data.

Arguments	F0 rise		F0 fall	
	A2	A3	A2	A3
2	9.1	–	14.1	–
3	11.2	9.9	13.4	19.9

In the remaining 'flat' cases, neither downstep nor upstep could be measured. This group of transitions made up about 25% of all realizations; see Table 14. Since the within-sentence comparison did not reveal any downstep relation in the remaining data, we calculated the F0 range for the individual arguments to assess whether the F0 course is really 'flat' as suggested by the within-sentence comparison of scaling or whether a significant amount of F0 movement exists on the arguments to indicate postfocal prominences. Table 15 displays the average F0 rise and F0 fall for the second and third arguments in two- and three-argument sentences, respectively. The F0 rise amounted to about 10 Hz on average, while the F0 fall amounted to an average of 13–19 Hz. Roughly speaking, these data resembled the overall means of F0 range reported in the section on all data above (cf. Tables 1 and 2). We observe a considerable change in the amount of F0 range, which leads to the assumption of the presence of postfocal prominences. Hence, although the F0 maxima in these data were not in a downstep relation, the underlying prominences were realized by means of a change in F0 range, which supports the hypothesis on the presence of postfocal pitch accents.

5 Discussion and conclusion

The preceding sections have reported on a production experiment on the phonetic correlates of postfocal givenness in German. The sentences used consisted of a narrowly focused participle followed by one, two or three given arguments; see some examples in (12)–(14). Eleven female native speakers of Standard German uttered 17 sentences each, which were analysed for their acoustic properties with special attention to F0 scaling. The aim of the experiment was to assess two mutually exclusive hypotheses on postfocal prosodic phrasing and accentuation, as formulated in Section 2. The first hypothesis (H1), formulated in (10), predicted deaccenting and dephrasing of the postfocal, that is, postnuclear, material in a German sentence with an early pitch accent. More specifically, H1 claimed that the postnuclear material, regardless of how long, is included in the Φ-phrase of the last pitch-accented word and that Φ-phrases in the postnuclear region are deleted. If postnuclear phrasing is as in (8), repeated here as (17), one prediction is that no pitch accents are realized on the postfocal constituents (P1a). This is shown schematically in Figure 9. The postfocal

Figure 9. Flat realization of postnuclear arguments including postfocal declination; small dots represent non-nuclear tones, while the thick dot represents the nuclear accent (Option 1).

constituents have no pitch accents, although other correlates of lexical stress, such as intensity and duration, may be preserved. No pitch movement whatsoever should be present in the postfocal part of the sentence. The second prediction is that F0 on the postfocal constituents shows declination but no downstep (P1b).

	(x			)	ι-phrase
	(x			)	Φ-phrase
	(x)	(x)	(x)	(x)	ω-word
(17)	(VORgestellt$_F$	hat der Hummer	den Reiher	dem Hammel)$_Φ$	

According to H2, formulated in (11), pitch accents and Φ-phrases are preserved in the postfocal region, even though the register used there is extremely compressed as compared to the focused part of the sentence. The phrasing is as in (9), repeated in (18). Syntax-motivated prosodic phrasing is not suppressed, and remnants of pitch accents in a sequence of postfocal constituents may be maintained even if the register is much smaller than in an all-new sentence (P2a). In addition to declination, the postfocal constituents are in a downstep relation to each other (P2b), as shown in Figure 10.[10] The final low point of a declarative sentence is consistently reached at the end of the sentence.

	(x			)	ι-phrase
	(x)	(x)	(x)	(x)	Φ-phrase
	(x)	(x)	(x)	(x)	ω-word
(18)	(VORgestellt$_F$)$_Φ$	(hat der HUMmer)$_Φ$	(den REIher)$_Φ$	(dem HAMmel)$_Φ$	

The detailed analysis of the phonetic realization of the postfocal given arguments clearly spoke in favour of Option 2. In two- and three-argument sentences, 51% of the sequences of given arguments presented downstep and only 37% were flat (not considering the downstep between the verb and first argument, which was present in nearly all data). Upstep, another effect of phrasing, was a marginal realization concerning only 5% of the two-argument sentences and 13.5% of the three-argument sentences. Adding up the data with downstep and those with upstep, we found that 63% of the data had F0 correlates of phrasing in the postfocal region (again excluding the difference between the verb and the first argument). The correlates of phrasing were the scaling relationship between the Φ-phrases, which showed that each argument of the verb was still phrased individually.

Postfocal prosodic prominences have been analysed as phrase accents (Grice et al. 2000). Also in case of SOF in German, Baumann et al. (2010) proposed to analyse acoustic cues of prominences induced by SOF as phrase accents. Contrary to an analysis of phrase accents, the

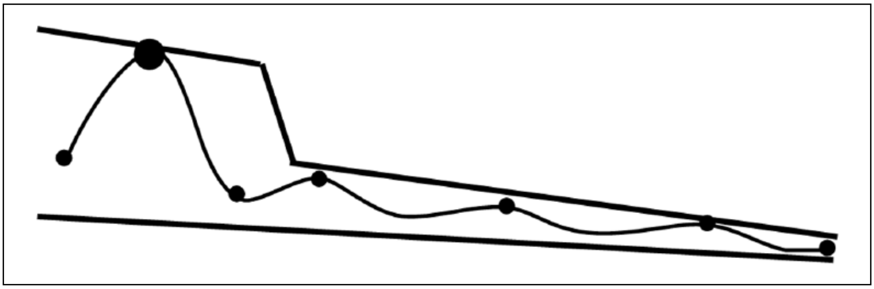

Figure 10. Downstep in the postfocal region; small dots represent non-nuclear tones, while the thick dot represents the nuclear accent (Option 2).

assumption argued for here is that the observed postfocal prosodic prominences originate from prosodic phrasing and constitute pitch accents associated with heads of Φ-phrase. Hence, the type of tonal event, pitch accent or secondarily associated tone of phrase edges differs compared to Grice et al. (2000). Another important difference between our data and that of Grice et al. (2000) concerns the number of postfocal tonal events. While the phrase accent analysis assumes a single tonal event between the nuclear pitch accent and the boundary tone, as assumed by the original phrase tone analysis proposed by Bruce (1977), which Grice et al. lean on, the number of postfocal prominent elements in our data varies from one to three, dependent on the number of postfocal Φ-phrases.

The issue of postfocal prosodic prominence recently became a matter of debate concerning the annotation of German intonation. In a consensual annotation system of German intonation (DIMA, *Deutsche Intonation–Modellierung und Annotation* 'German intonation–modelling and annotation'), Kügler et al. (2015) proposed to label prosodic prominence separately from and independently of a tonal annotation. The prominence layer distinguishes three levels of perceived prominence, weak, strong and extra strong prominence (cf. the Kiel intonation model, Kohler, 1991). Strong prominence is typically caused by syllables that are associated with a pitch accent, while weak prominence is typically caused by metrical strength or tonal events (Kügler et al., 2015, p. 2). The prosodic prominences found in the postfocal domain in the present study resemble that of weak prominence according to the DIMA annotation system. Thus, the presence of pitch accents as postfocal prominences as argued here, although in a compressed register, can be annotated in an annotation system that takes both a tonal and a prominence layer into account.

Further results of the experiment can be summed up as follows. Firstly, the scaling of F0 depended on the number of postfocal arguments; the more postfocal arguments, the higher the difference in tone scaling. Secondly, even in the cases in which no downward step larger than 7 Hz could be detected, pitch accents measured in terms of F0 range were still detectable. Thirdly, the end of the sentence was realized with the lowest F0 value of the sentence, independent of sentence length.

The main effect of an early narrow focus on F0 is pitch compression of the postfocal regions. The compression of pitch range does not cancel the phrasing in the postfocal given constituents, and pitch accents are still realized. Note, however, that these pitch accents cannot be compared to fully fledged pitch accents because of the compressed pitch register. As a consequence of these properties, we propose the following principle.

(19) Inalterability of syntax-driven prosodic phrasing
Prosodic phrasing created by syntax is not affected by information structure.

The absence of down- or upstepping of pitch accents, observed in 37% of the data (69 out of 187 sentences), can be explained in this model. We propose that in these cases, the F0 range used in the postfocal part of the sentence is totally compressed. However, it must be emphasized that this variant was not frequent. In nearly all three-argument sentences, at least one of the two transitions presented either downstep or upstep. Only 11 out of the 66 three-argument sentences had no downstep and no upstep whatsoever.

The results are compatible with an approach to the relationship between phrasing and pitch accents in which both effects are driven by independent modules, namely syntax in the case of phrasing, and information structure (i.e., the focus-given pattern of the sentence) in the case of the scaling of pitch accents, as proposed by Féry and Kügler (2008). The formation of Φ-phrases occurs at the syntax–prosody interface, but the scaling of the pitch accents is driven by the information structure of the constituents. In the default case, when the sentence has no narrow focus, and no given part, the accent scaling obeys a default pattern with downstep of the constituents, but if a constituent has an early narrow focus and the remainder of the sentence is given, the relationship between pitch accents is changed. In this case, PFC takes place.

This leads us to the asymmetry between prefocal and PFC addressed in Section 2. The fact that prefocal accents are realized in a slightly compressed pitch register and postfocal accents in a strongly compressed pitch register cannot be explained by information structure alone, nor by syntactic structure alone. It has been proposed a number of times that focus (nuclear pitch accent) in German tends to be aligned with the right-hand edge of the intonation phrase. This has been proposed in different models by phonologists working in different frameworks, such as von Stechow and Uhmann (1986), Truckenbrodt (1995) and Féry (2013), among others. German has a preference for syntactic structures in which this tendency is fulfilled, but if right-alignment of focus is violated, as in our sentences, the nuclear pitch accent is the last pitch accent of the intonation phrase. In this case, the pitch accent is right-aligned in its domain: there is no following pitch accent at the level of the ι-phrase. The result is compression of the postfocal part in order to enhance the finality effect of the nuclear pitch accent. The prefocal material, by contrast, being located before the nuclear accent, does not need to be compressed completely; slight compression of prefocal pitch register, as observed by Féry & Kügler (2008), is due to the given information status of prefocal constituents. This explains the asymmetry between pre- and postfocal given material in a language like German (see also Ladd, 2008; Wagner, 2005, for English).

Notice that the interpretation of the grid marks in the prefocal region, as in (20), and in the postfocal region of a sentence, as in (18), is the same in both cases. Grid structures express relative strengths of accents. The highest grid mark is the strongest, and the lower ones are weaker. How high the F0 is depends on the pitch register used in both cases.

```
      (                                     x               )           ι-phrase
      (     x       ) (         x     ) ((      x     )     )           Φ-phrase
      (     x       ) (         x     ) (       x     ) ( x )           ω-word
(20)  (der HUMmer)Φ (hat den REIher)Φ (dem HAMmel vorgestellt)Φ
```

To conclude, the metrical representation in (18) and (20) is only partial. However, it is in line with the intention of most authors using grid structures to express relative prominence: F0 range is not defined locally, but only globally. In order to show the difference between pre- and postfocal register, we need an independent phonological representation: either something like Figures 9 and 10 or some other representation. We propose using Figures 9 and 10 for the moment and leave it up to future research to determine how to account for register compression in a better way.

Acknowledgements

We wish to thank Bernadett Smolibocki and Sophia Jähnigen for helping with the recordings and labelling. The work was presented at the DIMA-III workshop at ZAS, Berlin, in May 2013, and we thank all participants for comments and discussion.

Funding

This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft) (grant number SFB 632 'Information structure', projects D5 and T2).

Notes

1. Information structural terminology, such as focus and givenness, is based on Krifka (2008).
2. The principle FocusProminence (Büring, 2010; Truckenbrodt, 1995) is redundant in a theory assuming ALIGN-FOCUS-R and DESTRESSGIVEN (Féry, 2011; Féry & Samek-Lodovici, 2006), an independently needed principle regulating the absence of stress in the postfocal domain.
3. A third option, suggested to us by Sun-Ah Jun, would be to regroup all given arguments in one Intonation Phrase separated from the focused participle as in (i), thus resembling a right-dislocated structure.

   (i) Nein. (VORgestellt$_F$)$_\iota$ (hat der Hummer den Reiher dem HAMmel)$_\iota$

   We do not consider this option here because the participants never realized any kind of pause after the focused verb.
4. We acknowledge the concerns of a reviewer that the experimental sentences are rather infrequent in spoken German. The important point is that the sentences are perfectly grammatical in German, and our participants did not report any difficulties while producing these sentences. Furthermore, a corpus query in the DWDS corpus (www.dwds.de) revealed instances of participle initial sentences with following constituents. The reason why we used such slightly unnatural sentences was the need to control different potentially disturbing factors, such as animacy, gender, syllabic and foot structure of the target words, thus restricting the flexibility of the experimental material. Early focus with following full noun phrases (NPs) can be observed in spoken speech. Figure 11 illustrates this with an example from a map task dialogue of speakers of Upper Saxon German, which was analysed by Kügler (2007). The realization of postfocal constituents resembles that of our data.

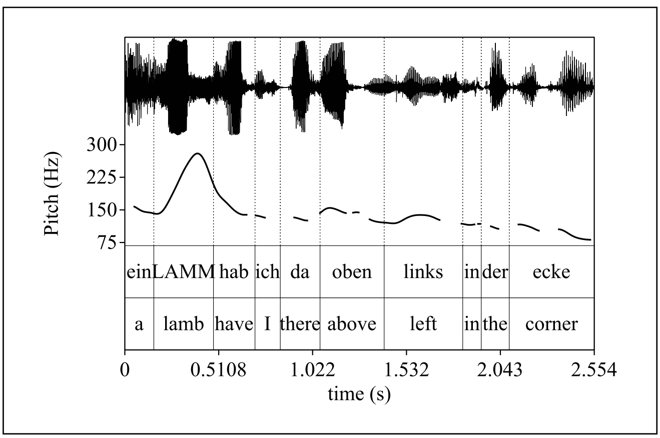

Figure 11. Illustration of a sentence initial focus with following full NPs, taken from a map task corpus of Upper Saxon German (Kügler, 2007); the sentence is 'I have a lamb up there in the left corner.'

5. A reviewer is concerned about microprosodic influences of the target words. We believe that we have reduced the effects of microprosody. To balance the well-known effect of vowel height on F0 (Lehiste & Peterson, 1961; Whalen & Levitt, 1995) we used high- (/u/) and low-ending vowels (/a/) and the closing diphthong and (/ai̯/). The effects of intrinsic F0 are further discussed in Section 4. In addition, each of the three target words occurred in all positions of the sentence, that is, in each argument. Any effect of onset consonants was reduced to a minimum since the onset consonants /h, r/ were intervocalically voiced [dɛɐ̯ ɦaml̩] (der Hammel) [dɛɐ̯ ɦʊmɐ] (der Hummer) [dɛɐ̯ ʁaɪ̯jɐ] (der Reiher). With this manipulation, any F0 raising due to an initial voiceless fricative was reduced: voiced fricatives are known not to raise F0 at vowel onsets (Hanson, 2009; Möbius et al., 1987; Whalen & Levitt, 1995).
6. Note that in the case of a model that applies random slopes for speakers and item the comparison within three-argument sentences becomes non-significant for the second and fourth contrasts in Table 8.
7. The contrast for the first argument between two- and three-argument sentences is not significant when applying random slopes for speakers and items in the model.
8. The contrast for three-argument sentences between the first and second arguments becomes not significant if applying random slopes for speakers and items in the model.
9. In flat realizations of a sequence of two pitch accents, or plateau-shaped accents, the second one sounds higher in pitch and more prominent (Knight, 2008). We thank the reviewer who pointed to this strategy of expressing prominence, which can be seen in the last postfocal accent in Figure 6. If speakers use a pitch plateau as a strategy to express prominence, it well serves our point that we are dealing with postfocal prominences even in cases where there are no pitch range effects on postfocal constituents. What we do not know, however, is whether plateau-shaped accents also have this perceptual effect in a compressed pitch register. Further research in the perception of plateau-shaped accents in different pitch registers would shed light on this issue.
10. In the experiment, Φ-phrases are equivalent to prosodic words, and as a reviewer observes, it could be that the phonetic correlates are assigned at the level of the prosodic words. While pitch accents may signal prosodic words, differences in pitch scaling do not. We propose that they are correlates of Φ-phrases.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2013). *lme4: Linear mixed-effects models using Eigen and S4*. [R package version 1.0-4]. Retrieved from http://CRAN.R-project.org/package=lme4/

Baumann, S., & Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics*, *38*, 1636–1657.

Baumann, S., Mücke, D., & Becker, J. (2010). Expression of Second Occurrence Focus in German. *Linguistische Berichte*, *221*, 61–78.

Beaver, D., Clark, B. Z., Flemming, E., Jaeger, T. F., & Wolters, M. (2007). When semantics meets phonetics: Acoustical studies of second-occurrence focus. *Language*, *82*, 245–276.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, *3*, 255–309.

Bierwisch, M. (1966). Regeln für die Intonation deutscher Sätze. *Studia Grammatica*, *7*, 99–201.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer [Computer program Version 5.3.53]. Retrieved from http://www.praat.org/

Bruce, G. (1977). *Swedish word accent in sentence perspective*. Lund, Sweden: Travaux de l'Institut de Linguistique de Lund 12, Gleerup.

Bruce, G., & Gårding, E. (1978). A prosodic typology for Swedish dialects. In E. Gårding, G. Bruce, & R. Bannert (Eds.), *Nordic prosody* (pp. 219–228). Lund, Sweden: Gleerup.

Büring, D. (2010). Towards a typology of focus realization. In M. Zimmermann & C. Féry (Eds.), *Information structure. Theoretical, typological, and experimental perspectives* (pp. 177–205). Oxford, UK: Oxford University Press.

Cinque, G. (1993). A null theory of phrase and compound stress. *Linguistic Inquiry*, *24*, 239–297.
Fanselow, G., & Lenertová, D. (2011). Left peripheral focus: mismatches between syntax and information structure. *Natural Language and Linguistic Theory*, *29*, 169–209.
Féry, C. (2011). German sentence accents and embedded prosodic phrases. *Lingua*, *12*, 1906–1922.
Féry, C. (2013). Focus as alignment. *Natural Language and Linguistic Theory*, *31*, 683–734.
Féry, C. (2016). *Intonation and prosodic structure*. Cambridge, UK: Cambridge University Press.
Féry, C., & Ishihara, S. (2009). The Phonology of second occurrence focus. *Journal of Linguistics*, *45*, 285–313.
Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, *36*, 680–703.
Féry, C., & Samek-Lodovici, V. (2006). Focus projection and prosodic prominence in nested foci. *Language*, *82*, 131–150.
Grabe, E. (1998). *Comparative intonational phonology: English and German*. (Doctoral dissertation), Universiteit Nijmegen, NL (MPI Series in Psycholinguistics, 7).
Grice, M., Ladd, D. R., & Arvaniti, A. (2000). On the place of phrase accents in intonational phonology. *Phonology*, *17*, 143–185.
Gussenhoven, C. (1992). Sentence accents and argument structure. In I. Roca (Ed.), *Thematic structure. Its role in grammar* (pp. 79–106). Berlin, Germany: Foris.
Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *The Journal of the Acoustical Society of America*, *125*, 425–441.
Hyman, L. M. (2006). Word-Prosodic Typology. *Phonology*, *23*, 225–257.
Ishihara, S. (2011). Japanese focus prosody revisited: Freeing focus from prosodic phrasing. *Lingua*, *121*, 1870–1889.
Jacobs, J. (1993). Integration. In. M. Reis (Ed.), *Wortstellung und Informationsstruktur. Linguistische Arbeiten* (pp. 63–116). Tübingen, Germany: Niemeyer.
Jun, S-A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology*, *15*, 189–226.
Knight, R.-A. (2008). The shape of nuclear falls and their effect on the perception of pitch and prominence: Peaks vs. plateaux. *Language and Speech*, *51*, 223–244.
Kollmeier, B., Brand, T., & Meyer, B. (2008). Perception of speech and sound. In J. Benesty, M. Mohan Sondhi, & Y. Huang (Eds.), *Springer handbook of speech processing* (pp. 61–82). Berlin, Germany: Springer.
Kratzer, A., & Selkirk, E. (2007). Phase theory and prosodic spellout: The case of verbs. *The Linguistic Review*, *24*, 93–135.
Krifka, M. (1984). Fokus, Topik, syntaktische Struktur und semantische Interpretation [Manuscript]. Retrieved from http://amor.cms.hu-berlin.de/~h2816i3x/Publications/Krifka%201984%20Fokus.PDF
Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, *55*, 243–276.
Kohler, K. J. (1991). A model of German intonation. In K. J. Kohler (Ed.), *Arbeitsberichte (AIPUK): Vol. 25. Studies in German Intonation* (pp. 295–360). Kiel, Germany: Institut für Phonetik und digitale Sprachverarbeitung.
Kügler, F. (2007). *The intonational phonology of Swabian and Upper Saxon*. Tübingen, Germany: Niemeyer.
Kügler, F., Smolibocki, B., Arnold, D., Braun, B., Baumann, S., Grice, M., …Wagner, P. (2015). DIMA–Annotation guidelines for German intonation. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. Paper number 317: 1–5. Retrieved from http://www.icphs2015.info/pdfs/Papers/ICPHS0317.pdf
Ladd, D. R. (1980). *The structure of intonational meaning: Evidence from English*. Bloomington, IN: Indiana University Press.
Ladd, D. R. (2008). *Intonational phonology*. Cambridge, UK: Cambridge University Press.
Laniran, Y. O., & Clements, G. N. (2003). Downstep and high raising: interacting factors in Yoruba tone production. *Journal of Phonetics*, *31*, 203–250.
Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *The Journal of the Acoustical Society of America*, *33*, 419–425.

Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. T. Oehrle (Eds.), *Language sound structure* (pp. 157–233). Cambridge, MA: Massachusetts Institute of Technology Press.

Möbius, B., Zimmermann, A., & Hess, W. (1987). Microprosodic fundamental frequency variations in German. In *Proceedings of the 11th International Congress of Phonetic Sciences* (Vol. 1, pp. 146–149). *Tallin*.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht, The Netherlands: Foris.

O'Connor, J. D., & Arnold, G. F. (1961). Intonation of colloquial English. London, UK: Longman.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. (Doctoral dissertation), Cambridge, MA: Massachusetts Institute of Technology Press.

R Core Team (2013). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Retrieved from http://www.R-project.org/

Rochemont, M. (1986). *Focus in generative grammar*. Amsterdam, The Netherlands: John Benjamins Publishing Company.

Samek-Lodovici, V. (2005). Prosody-syntax interaction in the expression of focus. *Natural Language and Linguistic Theory*, *23*, 687–755.

Selkirk, E. O. (1984). *Phonology and syntax. The relation between sound and structure*. Cambridge, MA: Massachusetts Institute of Technology Press.

Selkirk, E. O. (1995). Sentence prosody: Intonation, stress and phrasing. In J. Goldsmith (Ed.), *Handbook of phonological theory* (pp. 550–569). Chichester, UK: Blackwell Publishing Ltd.

Selkirk, E. (2011). The syntax-phonology interface. In J. Goldsmith, J. Riggle, & A. Yu (Eds.), *The handbook of phonological theory*, 2nd edition (pp. 435–484). Chichester, UK: Blackwell Publishing Ltd.

von Stechow, A., & Uhmann, S. (1986). Some remarks on focus projection. In W. Abraham & S. de Meij (Eds.), *Topic, focus and configurationality* (pp. 295–320). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Truckenbrodt, H. (1995). *Phonological phrases: Their relation to syntax, focus and prominence*. (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge.

Truckenbrodt, H. (2002). Upstep and embedded register levels. *Phonology*, *19*, 77–120.

Truckenbrodt, H. (2006). Phrasal stress. In K. Brown (Ed.), *The encyclopedia of languages and linguistics*, *2nd Edition*, *Vol. 9* (pp. 572–579). Oxford, UK: Elsevier.

Truckenbrodt, H. (2007). Upstep on edge tones and on nuclear accents. In C. Gussenhoven & T. Riad (Eds.), *Tones and tunes: Studies in word and sentence prosody*. *Vol 2* (pp. 349–386). Berlin, Germany: Mouton de Gruyter.

Wagner, M. (2005). *Prosody and recursion*. (Doctoral dissertation). Massachusetts Institute of Technology, Cambridge.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, *23*, 349–366.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*, 159–197.

Zubizaretta, M. L. (1998). *Prosody, focus and word order*. Cambridge, MA: Massachusetts Institute of Technology Press.