Isogloss
Open Journal of Romance Linguistics

Received: 25-02-2024
Accepted: 23-08-2024
Published: 18-02-2025

Prosody and focus recognition in Spanish: A fresh look at the Focus Prominence Rule

Christoph Gabriel
Johannes-Gutenberg University Mainz, Germany
christoph.gabriel@uni-mainz.de

Steffen Heidinger
University of Graz, Austria
steffen.heidinger@uni-graz.at

How to cite: Gabriel, Christoph & Steffen Heidinger. 2025. Prosody and focus recognition in Spanish: a fresh look at the Focus Prominence Rule. In *Interfaces in Spanish and beyond*, eds. Katrin Schmitz & Tim Diaubalick. *Special issue of Isogloss. Open Journal of Romance Linguistics* 11(2)/4, 1-33.
DOI: https://doi.org/10.5565/rev/isogloss.404

Abstract

At the interface between information structure and prosody, discourse-pragmatic categories are mapped onto prosodic structures and vice versa. The Focus Prominence Rule (FPR), which stipulates that the nuclear stress must fall within the focus domain, is considered one of the cross-linguistically most stable principles governing this mapping. However, FPR violations have received only little attention. To determine whether they occur during production or perception, we combine (i) the results of an earlier perception experiment on Argentinean Spanish showing that in about 30% of the stimuli the focus was not duly recognized and (ii) a detailed prosodic analysis of 30 lexically identical [S]-$_{F}$V-dO-iO stimuli used in that study. The following parameters are considered: first, voice quality in terms of the degree of post-focal devoicing; second, the alignment of the focal pitch accent's high target (H) and the following low target (L), the scaling of H, and the steepness of the subsequent fall; third, the duration of the stressed syllable of the focused XP in relation to the following stressed syllable. A Relative Weight Analysis shows that duration and alignment account for 80% of variability in accuracy rate, speaking in favor of the assumption that FPR violations mainly occur in production.

Keywords: Focus Prominence Rule, prosody, focus, Spanish, production vs. perception.

1. Introduction

At the information structure-prosody interface, discourse pragmatic categories are mapped onto different prosodic structures and vice versa. The Focus Prominence Rule (FPR), which was originally proposed by Jackendoff (1972) and requires that the nuclear stress be realized within the focus domain, is considered one of the most stable principles governing this mapping across languages. Despite the large body of research on the information structure-prosody interface, violations of the FPR have received only little attention (but see, e.g., Breen et al. 2010; Feldhausen & Vanrell 2015; Calhoun et al. 2018; El Zarka & Hödl 2021). This is surprising since the study of such violations broadens our understanding of interface phenomena in several respects: It helps us understand whether the vulnerability of interfaces (see, e.g., Sorace 2012) not only refers to second language (L2) learning and language contact but also extends to monolingual language use and whether the respective mappings at the interfaces are symmetrical.

In this paper, we concentrate on this latter aspect and address the question of whether FPR violations occur primarily during production or perception. To this end, we combine the results from a perception experiment on Argentinean Spanish (Gabriel & Heidinger 2022) with a detailed prosodic analysis of 30 productions of the declarative clause *María le da el diario a su hermano* 'Mary gives the newspaper to her brother' used as stimuli in that earlier study. The prosodic parameters considered comprise both F0 alignment and scaling as well as duration and voice quality.

Using a Relative Weight Analysis (Johnson 2000), we identify several prosodic predictors for correct focus recognition, the strongest of which being the alignment of the high tone in relation to the end of the stressed syllable. This suggests that violations of the FPR typically occur during production and not during perception: Stimuli with a clear nuclear stress on the subject are indeed interpreted as sentences with a focal subject. Given that it is also conceivable that violations would occur during perception (namely by interpreting such stimuli as sentences with a non-focal subject), our study allows to discriminate between these two options, thereby increasing our understanding of violations of the FPR and of interface vulnerability more generally.

The paper is structured as follows: In Section 2 we provide the relevant theoretical background and present relevant existing studies on the FPR and violations thereof. In Section 3 we briefly report on the perception experiment (Gabriel & Heidinger 2022), before presenting the prosodic analysis performed on a sample of the stimuli used in that experiment (Section 4). The paper closes with some concluding remarks (Section 5).

2. Background

2.1. The focus prominence rule

2.1.1. General characterization

The FPR states that in a prosodically well-formed utterance the nuclear stress falls within the focus domain (see (1b) vs. (1b')).[1]

(1) a. What did John buy?
b. John bought [a BIKE]$_F$.
b'. #JOHN bought [a bike]$_F$.

The FPR has been assumed irrespective of the focus type and should hold in the case of both information focus (see (1) and contrastive focus (see (2)).

(2) a. John bought a house, right?
b. No, John bought [a BIKE]$_F$.
b'. #No, JOHN bought [a bike]$_F$.

There are several formulations, which capture the FPR's idea that in a prosodically well-formed sentence the main prosodic prominence must be located within the focus domain. An early formulation can be found in Jackendoff (1972), who states that the main prominence must be within the focus and that stress rules determine the element of the focus the main prominence will be assigned to.

> If a phrase P is chosen as the focus of a sentence S, the highest stress in S will be on the syllable of P that is assigned highest stress by the regular stress rules (Jackendoff 1972: 237).

A shorter version of Jackendoff's formulation is given in Truckenbrodt (1995), who leaves aside the question of which element of the focus the main prominence will be assigned to.

> If a phrase P is chosen as the focus of a sentence S, the highest stress in S will be within P (Truckenbrodt 1995: 152).

The following quotes from Zubizarreta (1998; 2016) show that single authors may provide slightly different formulations of the FPR, highlighting different aspects of the relation between prosodic prominence and focus.

> Focus prominence rule
> Given two sister categories C_i (marked [+F]) and C_j (marked [-F]), C_i is more prominent that C_j (Zubizarreta 1998: 21).

[1] Throughout this paper nuclear stress is indicated by capitalizing the metrically strong syllable of the respective word, unless in abbreviations where it is signaled by boldface (e.g., **dO**).

> Focus prosody correspondence principle
> The F-marked constituent of a phrase must contain the rhythmically most prominent word of that phrase (Zubizarreta 1998: 38).

> The focused constituent must contain the rhythmically most prominent word, i.e. the word that bears the Nuclear Stress (NS) (Zubizarreta 2016: 166).

Within optimality theory (Prince & Smolensky 2004), the FPR has a corresponding constraint, which again can be formulated in different ways. Gabriel's (2010) formulation states that the focused phrase bears nuclear stress. Samek-Lodovici's (2005) constraint in addition captures the idea that the focus must be most prominent within its domain (typically a sentence).

> STRESSFOC: $[XP]_F$ bears nuclear stress (Gabriel 2010: 203).

> STRESS-FOCUS: For any XP_f and YP in the focus domain of XP_f, XP_f is prosodically more prominent than YP (Samek-Lodovici 2005: 696).

Note that in the above formulations prosodic prominence is referred to as *stress*, *prosodically more prominent* or *rhythmically most prominent word*. In the following we will stick to the term *nuclear stress*.

The FPR has been proposed to be active in a wide range of languages, and languages where the FPR is not active (see below) seem especially noteworthy.[2] Spanish is among the many languages for which it is assumed that the FPR mediates the relation between focus and nuclear stress. In the minimal pair in (3), (3b) is judged pragmatically infelicitous by the authors, because the nuclear stress is outside the focus.[3]

(3) a. Llegó tarde $[PePÍN]_F$. (Bosque & Gutiérrez-Rexach 2009: 681)
b. #$[$Llegó tarde$]_F$ PePÍN. (Bosque & Gutiérrez-Rexach 2009: 682)
arrived late Pepín
'Pepín arrived late.'

[2] This does not imply that the FPR is active in most languages. It might well be that the languages for which the realization of focus and the focus-prosody interface have been studied are languages where the FPR is active.

[3] Utterances violating the FPR are not consistently labeled in the literature. They are sometimes marked as ungrammatical (indicated by the asterisk: *; e.g., Olarrea 2012: 606, Bosque & Gutiérrez-Rexach 2009: 681) or as contextually/pragmatically infelicitous (indicated by the pound sign: #; e.g., Bosque & Gutiérrez-Rexach 2009: 682 in (3b)). Using the pound sign, as we do in this paper, highlights that the respective utterance – with its linear ordering and position of nuclear stress – is perfectly acceptable in other contexts. For example, the sentence in (1) would be fine in the context of a wh-question targeting the subject. Note, however, that in an OT analysis the respective constraint (e.g., STRESSFOC) would be part of the grammar, and its violation would thus target a grammatical constraint, and violations may therefore lead to ungrammaticality. As concerns the relation between acceptability and contextual felicity, Hoot (2016: 354) interprets the rather high scores of utterances violating the FPR (see Section 2.2) as indicating a difference between acceptability and contextual felicity (where infelicitous utterances still score in the mid-range of the acceptability scale, i.e., infelicity does not imply unacceptability).

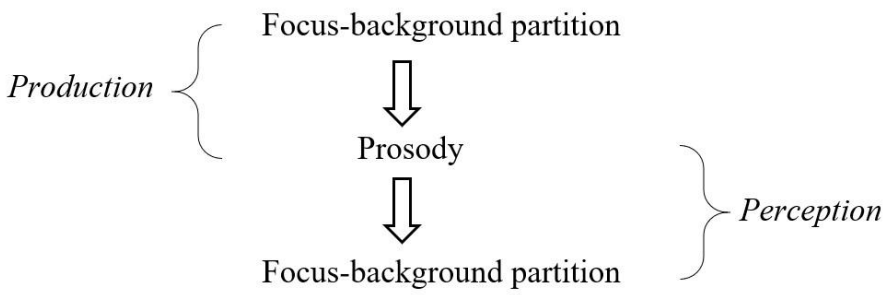
Examples for the claim that the FPR holds for Spanish, which we have already seen above, are Zubizarreta's (1998) formulation of the FPR, which she extends to Spanish, Gabriel's (2010) STRESSFOC constraint, or Bosque & Gutiérrez-Rexach's (2009: 681f.) minimal pair. Furthermore, Olarrea (2012: 606) states that in Spanish "the highest syntactic node marked as focus must dominate the constituent that contains the Nuclear Stress […]". Feldhausen & Vanrell (2015: 124–125) and Heidinger (2015: 118) propose constraint rankings for Spanish in which STRESSFOCUS is dominated by no other conflicting constraint, suggesting a close relation between focus and prosodic prominence in Spanish.

Beyond Spanish, we do not have to look far for other languages for which the validity of the FPR has also been postulated. Samek-Lodovici (2016: 217) assumes that in English and Italian, STRESSFOCUS dominates all other constraints it conflicts with (STAY, RIGHTSTRESS). By contrast, he also discusses languages in which STRESSFOCUS plays a lesser role in the sense that optimal candidates, i.e., grammatical sentences, violate STRESSFOCUS. For instance, in the Bantu language Northern Soto, there is no requirement "to highlight the focus prosodically" (Zerbian 2006: 165). Another example of a language which is said to lack any prosodic manifestation of focus is Wolof, spoken in Senegal, Gambia, and southern Morocco, which is characterized by a "lack of any specific intonation for utterances containing a focus" (Rialland & Robert 2001: 897). Instead, focus marking is achieved by morphological means: "One of the characteristic features of Wolof grammar is the expression of information structure in the verbal morphology" (Rialland & Robert 2001: 895). Büring (2009: 205) argues – based on a cross-linguistic survey of focus realization – that a prominence-based theory of focus is on the right track, but that prominence is not always expressed prosodically. Cases reported in the literature where the FPR is violated are introduced in Section 2.2.

2.1.2. The focus-prosody interface and the two sides of the FPR

As illustrated in Figure 1, the focus-prosody interface involves two mappings: During the production of a sentence, a given focus-background partition is mapped onto a prosodic form in such a way that the nuclear stress falls within the focus (see (4)); during the perception (or: interpretation) of a sentence, a given prosody is mapped onto a focus-background partition in such a way that the nuclear stress falls within the focus (see (5)).

Figure 1. Mapping between focus and prosody in production and perception.



(4) FPR during production
John + bought + [a bike]$_F$ *produced as* John bought [a BIKE]$_F$

(5) FPR during perception
JOHN bought a bike *interpreted as* [JOHN]$_F$ bought a bike.

Given that the position of the nuclear stress is relevant for the focus-background partition assigned to a given sentence, we do not only expect it to be realized within the focus domain but also that it is perceivable as such by the hearer and that it guides the hearer's focus interpretation. After all, one would expect that speakers choose among different possible sentence forms in such a way that they facilitate the hearer's focus interpretation of the sentence. This idea is captured by the notion *information packaging* (Chafe 1976; Vallduví 1990; Vallduví & Engdahl 1996), as described in the following representative quote.[4]

> *Information packaging* […] is a structuring of sentences by syntactic, prosodic, or morphological means that arises from the need to meet the communicative demands of a particular context or discourse. In particular, information packaging indicates how information conveyed by linguistic means fits into the (hearer's mental model of the) context or discourse (Vallduví & Engdahl 1996: 460).

Consequently, we expect the packaging done by the speaker to be noticeable for the hearer. More precisely, if focus shapes prosody for the purpose of information packaging, then the nuclear stress should not only be *produced* within the focus domain by the speaker but also be *perceived* within the focus domain by the hearer. In other words: the prosodic form is only helpful for focus interpretation if the nuclear stress is prominent enough to be perceived as such. The current representation of the nuclear stress in the literature, and we do not exclude ourselves here, where the position of the nuclear stress is unambiguously indicated by graphic means, is of course a simplification of the typically gradual nature of differences in prosodic prominence (see, e.g., Feldhausen et al. 2011 for a perception-based approach to the gradualness of focal pitch accents in Argentinean Spanish, and also Roessig 2021 on categorical vs. continuous aspects of prosodic focus marking).

It should be added that adhering to the FPR does not necessarily result in an unambiguous signaling of the focus of the sentence. For instance, a sentence with an SVO order and the nuclear stress placed on the final constituent, as in (6), is ambiguous

[4] However, one must be cautious since formal reflexes of focus-background partition should not be generally interpreted as focus marking devices (Matić & Wedgwood 2013: 129). Consider, for example, the transitive sentence in (i), where the subject *el gato* appears in sentence-final position. Zubizarreta (1999) notes that this word order can only be interpreted as *el gato* being the narrow focus of the sentence. However, this does not mean that the reason for putting the subject in final position is to create a sentence form that unambiguously signals focus-background partition.

(i) Se comió un ratón el gato. (Zubizarreta 1999: 4232)
REFL ate a mouse the cat
'The cat ate a mouse.'

The reason for the focused subject's final position might be rather the alignment between focus and the unmarked sentence-final position of nuclear stress. The unambiguity of (i) might thus rather be a consequence and not the purpose of the deviation from the unmarked SVO order.

between various focus-background partitions (sentence focus, VP focus, object focus), as shown by its felicity in the context of different wh-questions.

(6) John bought a BIKE.
  a. What happened?
  b. What did John do?
  c. What did John buy?

In this case, the position of the nuclear stress indicates that the stressed constituent is part of the focus, but it does not indicate the extension of the focus domain: focus might project from the prosodically highlighted constituent to other parts of the sentence (see Höhle 1982: 99 and Selkirk 1984 on focus projection).[5] Even in such cases the FPR has an important function in the hearer's determination of the focus-background partition, as it rules out several focus-background partitions: The FPR excludes interpretations of (6) as a narrow focus on an element that does not carry the nuclear stress (e.g., $[\text{John}]_F$ or $[\text{bought}]_F$).

2.2. Violations of the FPR

In Section 2.1.2, we have argued that the FPR makes a prediction about the relation between focus and prosody in both production and perception, i.e., (i) the mapping of a focus-background partition onto a certain prosodic form and (ii) the mapping of a prosodic form onto a certain focus-background partition. Consequently, violations of the FPR may occur during both production and perception.

A violation of the FPR during production occurs whenever the prosody (as produced by speaker) does not correspond to the focus-background partition (as determined by the context) in that the nuclear stress lies outside the focus (see (7)).

(7) Violation of the FPR during production
  John + bought + $[\text{a bike}]_F$ *produced as* JOHN bought $[\text{a bike}]_F$

A FPR violation during perception (or: interpretation) occurs whenever the focus-background partition assumed by the hearer does not correspond to the prosody of the utterance in that the nuclear stress lies outside of the assumed focus (see (8)).

(8) Violation of the FPR during perception
  JOHN bought a bike *interpreted as* JOHN bought $[\text{a bike}]_F$

In the following, we will discuss several studies suggesting that violations of both types are attested in the existing experimental literature on the focus-prosody interface. Calhoun et al. (2018: 18) conducted a production experiment using a picture description task. They report for focused subjects in intransitives in Venezuelan Spanish that three types of syntactic-prosodic structures were produced: subject-verb with the nuclear stress on the verb ($[\text{S}]_F$-**V**), subject-verb with the nuclear stress on the subject ($[\textbf{S}]_F$-V), and verb-subject with the nuclear stress on the subject (V-$[\textbf{S}]_F$). As for

[5] Note, however, that the left edge of the focus domain may be signaled by an intermediate high phrase boundary (H-) as in Sp. *Mariana miraba* H- $[\textit{la luna}]_F$ 'Mariana was watching the moon' (Hualde 2005: 261).

the nuclear stress, the authors first determined its position auditorily and then corroborated their judgments with several acoustic measures showing that the syllables classified as bearing nuclear stress exhibited greater F0 maxima and longer durations as compared to the metrically strong syllable of the other lexical item.[6]

Table 1 shows the relative frequencies of the three structures separately for unergative and unaccusative verbs. In most of the cases analyzed, the participants produced structures with the focused subject in initial position. It is striking that the most frequently produced structure is the one that violates the FPR, namely [S]$_F$-**V** (see Calhoun et al. 2018: 18). In this structure the nuclear stress is in sentence final position and lies outside the focused subject. The data thus shows a surprisingly high frequency of violation of the FPR (56% and 42% of the cases, respectively).

Table 1. Information focus on subject in intransitives (relative frequency; boldface indicates position of nuclear stress).

	[S]$_F$-**V**	[**S**]$_F$-V	V-[**S**]$_F$	
unergative	56	32	12	100 (N = 81)
unaccusative	42	33	25	100 (N = 76)

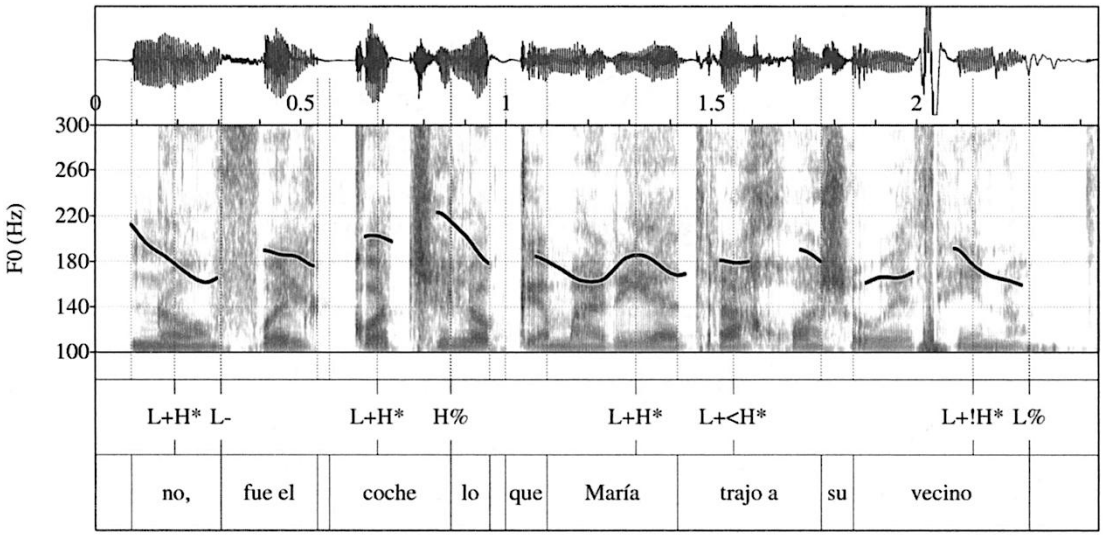
Source: Calhoun et al. (2018: 18; modified)

Another relevant study is provided by Feldhausen & Vanrell (2015), who present experimental production data on Spanish clefts where the nuclear stress does not fall within the focus. The example they put forward is a simple cleft sentence where the clefted constituent is the contrastive focus (*el coche* 'the car'). According to their prosodic analysis, the nuclear accent does not fall on the focused constituent, but on the sentence-final XP *a su vecino* 'to her neighbor' (see (9) and Figure 2).

(9) Feldhausen & Vanrell 2015: 48; adapted
No, fue [el coche]$_F$ lo que María trajo a su veCIno.
no it.was the car that Mary brought to her neighbor
'No, it was the car what Mary brought to her neighbor.'

[6] Surprisingly, they also found later peak alignment for nuclear syllables as compared to stressed syllables bearing non-nuclear stress. Scaling was not taken into account.

Figure 2. Simple cleft sentence, where the nuclear accent (L+!H*) lies outside the focus domain.



Source: Feldhausen & Vanrell (2015: 48)

It should be kept in mind, however, that the cases of a mismatch between focus and nuclear stress reported by Feldhausen &Vanrell (2015) concern cleft sentences, which, according to their analysis, consist of two intonational phrases (IP), but not simple (i.e., non-cleft) sentences.[7]

Turning to perception, El Zarka & Hödl (2021) conducted two forced-choice perception experiments on Egyptian Arabic and asked listeners to decide which of two auditory stimuli is a more suitable response to a given question. In the first experiment, which is the one reported here, auditory stimuli that were produced either in a VP focus or a subject focus context were presented as optional answers to questions targeting a VP focus or a subject focus.[8] Hence, in the first condition, a VP focus question was presented together with two lexically and syntactically identical answers, which only differed with respect to the context in which they were originally produced; in the second condition, a subject focus question was presented together with two lexically and syntactically identical responses, which only differed with respect to the context in which they were originally produced.

(10) Condition 1
Subject focus question (e.g., *Who switched on the lights?*)
- o Answer produced to subject focus question
- o Answer produced to VP focus question

[7] The L+H* pitch accent realized on focused item *coche* can thus be interpreted as the nuclear stress of the first IP, which would be in perfect accordance with the FPR. The absence of deaccentuation in the post-focal part of the cleft second IP *lo que María trajo a su vecino* is less surprising when we consider that post-focal deaccentuation is not even compulsory when the focal and the post-focal part of the clause are only separated by a lower-level boundary (ip) and thus phrased within the same IP (Beckman et al. 2002: 17–18).

[8] El Zarka & Hödl (2021: 59) refer to the VP focus question as "topic-comment question" and to the subject focus question as "focus background question".

(11) Condition 2
VP focus question (e.g., *What did Samira do?*)
- o Answer produced to subject focus question
- o Answer produced to VP focus question

The authors measured the proportion of correct responses, i.e., the frequency with which the focus-background partition of the answer corresponds to the focus-background partition of the question. The main result of the experiment is that the proportion of correct focus recognition amounts to 55.9% and is therefore only slightly above chance level (i.e., 50%) (El Zarka & Hödl 2021: 59).[9]

As for their prosodic properties, the auditory stimuli are described as "especially clear examples" (El Zarka & Hödl 2021: 59) of the distinct prosodic realizations of VP focus and subject focus. Hence, we can assume that the two realizations involve different positions of the nuclear stress: on the subject constituent in the case of subject focus, and within the VP in the case of VP focus. The low rate of correct focus recognition in the experiment must therefore be interpreted as violations of the FPR during perception. Stimuli with a nuclear stress outside the subject constituent were frequently interpreted as answers to subject focus questions, and stimuli with a nuclear stress outside the VP were frequently interpreted as answers to VP focus questions.

In addition to the above-mentioned studies where the attested violations of the FPR can be clearly attributed to either production (Calhoun et al. 2018; Feldhausen & Vanrell 2015) or perception (El Zarka & Hödl 2021), there are also studies reporting on FPR violations that cannot be clearly localized at the level of production or perception.

Breen et al. (2010) report on a combined production and perception experiment on the mapping between prosody and focus in (American) English. Two participants (one speaker, one listener) sat at two different computers in the same room. The speaker had to produce answers to questions for the listener, and the listener then had to indicate which question the speaker was answering (Breen et al. 2010: 1066). The experiment included seven experimental conditions resulting from the crossing of two focus types (contrastive vs. non-contrastive) and four focused constituents, i.e., whole sentence (only non-contrastive), subject, verb, and direct object (Breen et al. 2010: 1053). The speakers' productions were analyzed with respect to several prosodic parameters, and four of them (mean F0, maximal F0, duration and intensity) differentiated subject focus, verb focus, and object focus from each other in terms of prosodic form (Breen et al. 2010: 1067, 1069). As to the perception part of the results, listeners' overall accuracy was 55% (Breen et al. 2010: 1073), i.e., in 55% of the responses the

[9] In Roettger et al. (2019), listeners were presented with one question prompting a specific focus structure (e.g., narrow subject focus) and two answers that were either produced in the context of this question (narrow subject focus) or in the context of another question (e.g., broad focus). The task was to choose between the two answers, and, simplifying somewhat, the authors measured how often the focus prompted by the question matched the focus of the chosen answer. Although their stimulus material includes utterances produced under narrow focus on the subject and narrow focus on the object (Roettger et al. 2019: 844), in their presentation of the results no distinction between these two types of narrow focus is made. As a consequence, we cannot compare how often an utterance with a nuclear stress on constituent A is interpreted such that A is not part of the focus (which would constitute a violation of the FPR during perception).

listeners chose the same question as the one provided to the speaker as a preceding context. Given that sentence focus items were excluded from this analysis, the chance level of correctly identifying the focused constituent (subject, verb, or object) was 0.33 (Breen et al. 2010: 1073). The fact that the accuracy rate is above chance level is sufficient for the authors to conclude that "listeners were highly successful" in determining the focus location (Breen et al. 2010: 1078). With respect to the predictions of the FPR, however, the results show that in a large number of cases the focus is not correctly recognized and the FPR is violated. Although the authors do not elaborate on whether the lack of accuracy is due to production or perception, they give one important hint: Items where the focus was correctly recognized and those where the focus was not recognized have very similar prosodic properties (Breen et al. 2010: 1067). This suggests that the failure to correctly recognize the focused constituent results from a violation of the FPR during perception. But unlike in El Zarka & Hödl (2021) the stimuli for which the focused constituent had to be indicated were not *a priori* controlled for with respect to their prosodic well-formedness or prototypicality. Hence, it cannot be decided with certainty whether the violations of the FPR reported in Breen et al. (2010) occur during production or perception.

Regarding focus recognition in German, Krüger (2009) reports on a combined production and perception experiment, where auditory stimuli were created in the production experiment and then used in the perception experiment: How often does the contextual question assumed by the listener correspond to the actual question the speaker was presented with in production? In the perception part, the listeners heard S-AUX-O-V sentences and were asked to choose between four different questions as possible preceding contexts (Krüger 2009: 66). These questions elicit four different focus structures: sentence focus, information focus on the object, contrastive focus on the object, and contrastive focus on the subject (Krüger 2009: 43). The accuracy rates in focus choice strongly depend on the focus structure of the stimulus: they are highest for contrastive focus on the subject (99.1%), followed by contrastive focus on the object (60.5%), sentence focus (57.8%), and, finally, information focus on the object (49%) (Krüger 2009: 71). As concerns the FPR, these results show that violations of the FPR rarely occur in the case of contrastive focus on the subject. The remaining accuracy rates can, however, not be interpreted with respect to the FPR, since it is unclear whether incorrect focus choices for stimuli with information focus on the object are due to interpretations of the subject as focus (= violation of the FPR) or interpretations as contrastive object focus (= no violation of the FPR).

Another study reporting violations of the FPR which cannot be clearly attributed to one of the two mappings is Hoot (2016). In this study, acceptability judgements from Mexican raters were collected for different combinations of word orders and stress patterns in the context of wh-questions targeting either the subject or the object as focus. Besides stimuli where focus and nuclear stress coincide (following the FPR), Hoot also tested the acceptability of mismatches, i.e., stimuli violating the FPR. The results in Table 2 and 3 show that these stimuli receive relatively high scores (3.29 and 3.26 on a 5-point Likert scale with 5 as the best score) (see Hoot 2012; 2017 for similar results). Additionally, in the case of the focused subject, stimuli with a mismatch receive higher scores than stimuli where the focused subject ends up in sentence final position via p-movement.

Table 2. Acceptability scores for focused subjects (information focus).

Structure	Mean rating	
V-dO-[**S**]$_F$	2.83 (SD = 1.0)	
[**S**]$_F$-V-dO	4.43 (SD = 0.7)	
[S]$_F$-V-**dO**	3.29 (SD = 1.0)	Violation of FPR!

Source: Hoot (2016: 352; adapted)

Table 3. Acceptability scores for focused direct objects (information focus).

Structure	Mean rating	
S-V-iO-[**dO**]$_F$	4.08 (SD = 0.6)	
S-V-[**dO**]$_F$-iO	4.23 (SD = 0.6)	
S-V-[dO]$_F$-**iO**	3.26 (SD = 0.9)	Violation of FPR!

Source: Hoot (2016: 353; adapted)

The results presented in Table 2 and 3 do not show violations of the FPR during production, since the stimuli with a mismatch are not spontaneous data, but material produced for the sake of being used as experimental stimuli. At the same time, the results also do not show violations during perception, because the experimental task does not involve the assignment of a focus-background partition to a stimulus with a certain prosodic form. What these results do indicate, however, is the vulnerable status of the FPR. If the FPR were invulnerable and STRESSFOCUS a highly ranked constraint (and thus had a high violation cost), we would expect a greater impact on the acceptability of the respective sentences. Interestingly, Hoot (2016: 354) does not interpret the rather high scores of mismatches as hints towards the vulnerability of the FPR; instead, he interprets the scores as indicating a difference between acceptability and contextual felicity (where infelicitous stimuli still score in the mid-range of the scale).

2.3. Interim summary and the contribution of our study

It is widely accepted that the FPR mediates the relation between focus and prosody. However, there are hints in the existing literature that violations of the FPR do occur and that these violations may happen during either production or perception. In the following we will present an experiment on focus recognition and FPR violations in Spanish (Gabriel & Heidinger 2022) and a novel post-hoc study to determine whether the observed violations occur during perception or production.

Our study is similar to the ones by Breen et al. (2010) and Krüger (2009) in that we use relatively unfiltered semi-spontaneous production data (i.e., not just prototypical items). One difference to Breen et al. (2010), however, is that production and perception do not take place in one run but are separated in time (even widely). The key difference compared to both Breen et al. (2010) and Krüger (2009) is that for some of our stimuli, we examined the relationship between prosodic properties and accuracy in focus recognition and can thus determine whether violations of the FPR in our data are more likely to occur during production or during perception.

This also sets us apart from El Zarka & Hödl (2021), who only tested prosodically "good" stimuli during perception (which is why the violations they observed can only result from perception). Conversely, we also go beyond Calhoun et al. (2018) as their data contain violations of the FPR during production, but no perceptual data are available for these utterances with respect to focus recognition.

3. A perception experiment testing focus recognition

3.1. Method and data sources

In Gabriel & Heidinger (2022), we conducted a forced-choice perception experiment with auditory stimuli (77 different stimuli in total) in which participants listened to a target sentence as an audio file and were then asked to indicate which of two wh-questions (presented in written form) was more suitable as the preceding context of the sentence they had heard. One wh-question targeted the subject as a narrow focus, the other one targeted the direct object. The auditory stimuli stem from a previous production experiment with speakers of Argentinean Spanish and were produced either as an answer to a question targeting the subject or a question targeting the direct object as a narrow focus. As the FPR predicts that the nuclear stress is perceived within the focus, we took the participants' choice of the respective question as indication of their focus interpretation.

This gives us four combinations of the focus in the stimulus and the perceived focus in the perception experiment (see Table 4). If the FPR were followed by the speakers in the production experiment and by the listeners in the perception experiment, then two foci should match (= ✓*FPR*). Consequently, we counted every instance where the perceived focus does not correspond to the focus of the stimulus as a violation of the FPR (= **FPR*).

Table 4. Obedience and violation of the FPR

Interpretation / Stimulus	$[\text{subject}]_F$	$[\text{direct object}]_F$
$[\text{subject}]_F$	✓FPR	*FPR
$[\text{direct object}]_F$	*FPR	✓FPR

In addition to the focused constituent the stimuli also varied with respect to the focus type, i.e. whether the stimulus was produced in the context prompting a contrastive focus or an information focus. Furthermore, we recruited participants from two different varieties of Spanish: native speakers of European vs. native speakers of Argentinean Spanish.[10] Our concern, however, was not diatopic variation in the area of

[10] Participants were asked where they had spent most of their life. Argentineans could choose between provinces; most of them selected Río Negro (65%) followed by Neuquén (28%). Note in this context that the varieties spoken in the northern Patagonian provinces do not substantially differ from Buenos Aires Spanish at the intonational level, i.e., they exhibit the same tonal inventory (see Prieto & Roseano 2009-2010). However, the prosody of other Argentinean varieties, e.g., those of Córdoba and Tucumán, crucially differs from the one of the variety spoken in the capital (Terán & Ortega-Llebaria 2017; Gabriel 2021). Spaniards

focus marking and focus recognition *per se*, since this would require testing a larger number of varieties. Instead, European Spanish served as a comparative variety that allows to test whether listeners of the variety in which the stimuli were produced recognize the focus better than listeners who speak a variety with different prosodic features. In our post-hoc study, which aims to determine whether violations of the FPR occur during perception or production, we will again refer to the two varieties (Section 4).

The above considerations result in a design with three factors (independent variables), with two levels each: *focus type* and *focused constituent* as within-subjects factors, and *variety spoken by participants* as a between-subjects factor. The factors are given in (12), and the experimental conditions of the stimuli in (13).

(12) a. Focused constituent (Subject vs. direct object)
b. Focus type (Contrastive vs. information focus)
c. Variety spoken by participants (European vs. Argentinean Spanish)

(13) Condition 1: $[S]_{CF}$-V-dO
Condition 2: S-V-$[dO]_{CF}$
Condition 3: $[S]_{IF}$-V-dO
Condition 4: S-V-$[dO]_{IF}$

Note that in some stimuli for the conditions with subject focus (i.e., Conditions 1 and 3) the direct object is followed by some other postverbal constituent (locative adjunct or indirect object). Crucially, the stimuli used in the post-hoc analysis reported in Section 4 are of this type and the direct object is followed by an indirect object. For ease of exposition, we will use S-V-dO as a unified representation in Section 3.

The dependent variable was the participants' choice between two wh-questions, based on which of the questions s/he considered more suitable as the preceding context of the sentence. Based on this choice, we calculated the accuracy, i.e., the rate of correspondence between the focus in the respective stimulus and the focus as indicated by the participants' choice (✓FPR in Table 4).

A total of 90 persons participated in the experiment: 40 monolingual native speakers of Argentinean Spanish and 50 monolingual native speakers of European Spanish. The participants were unfamiliar with the experiment's purpose and the underlying concepts. The experiment was presented in a web-based environment using the experimental software Limesurvey. It was self-paced, but participants were instructed to listen to the stimuli not more than three times.

3.2. Results

In the perception experiment, we collected a total of 990 judgments (on which of the two questions was a more suitable preceding context) for each of the four conditions. This amounts to a total of 3960 judgments for the subsequent analysis. First, we examined how often the focus of the stimulus (as controlled in the production experiment) and the focus according to the selected question coincide. Across all 90 partici-

could choose between autonomous communities; most of them selected Andalucía (48%) followed by Comunidad de Madrid (20%).

pants, the average accuracy amounts to 70% (SD = 11.2). Thus, on average the perceived focus corresponds to the focus of the stimulus in 70% of the answers (the accuracy is thus clearly above the chance-level of 0.5). This means in turn that in 30% of the answers the two foci do not correspond, which we count as violations of the FPR (see Section 2.2). Table 5 shows the frequency with which stimuli with subject focus were perceived as subject focus and how often stimuli with object focus were perceived as object focus (shaded cells indicate correct focus recognition).

Table 5. Confusion matrix.

Interpretation / Stimulus	$[\text{subject}]_F$	$[\text{direct object}]_F$	
$[\text{subject}]_F$	78%	22%	100%
$[\text{direct object}]_F$	37%	63%	100%

Second, we observed that the accuracy varies to a considerable degree once the factors *focused constituent*, *focus type*, and *variety* are considered (see Table 6).

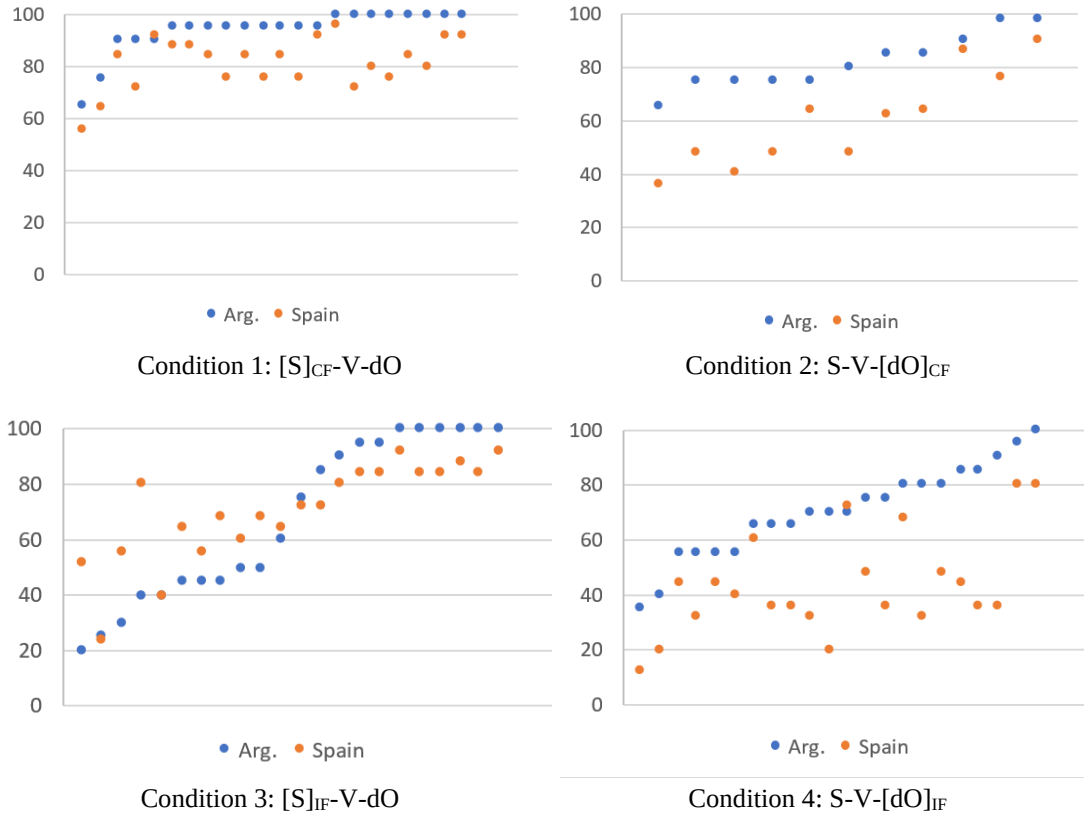
Table 6. Accuracy of focus choice in % (perception experiment).

Conditions	Argentina	Spain
1 $[S]_{CF}$-V-dO	94	81
2 S-V-$[dO]_{CF}$	82	60
3 $[S]_{IF}$-V-dO	68	70
4 S-V-$[dO]_{IF}$	70	43
Average	78	64

Accuracy is significantly higher for stimuli with contrastive focus than for stimuli with information focus, it is also significantly higher for stimuli with subject focus than for stimuli with object focus, and for listeners from Argentina than for listeners from Spain (see Gabriel & Heidinger 2022 for the inferential statistics and a detailed discussion of these results including the differences between contrastive and information focus, between subject and object focus, and between the two groups of listeners).

Finally, we observed considerable variation in accuracy among the stimuli of one and the same condition. Figure 3 gives the accuracy rate of each stimulus used in the experiment; each panel represents one condition and accuracy is shown separately for listeners from Argentina and Spain. The variation in accuracy is smaller for stimuli with contrastive focus than for stimuli with information focus.

Figure 3. Accuracy (%) per stimulus for four conditions.



Condition 1: $[S]_{CF}$-V-dO

Condition 2: S-V-$[dO]_{CF}$

Condition 3: $[S]_{IF}$-V-dO

Condition 4: S-V-$[dO]_{IF}$

The obvious question is whether the results from the perception experiment may provide new insights into the locus of the violations, i.e., whether they take place during production or perception. With respect to this question, the variation in accuracy among the stimuli of one and the same condition plays an important role. Since the syntactic shape of the stimuli is fixed as S-V-dO (with an additional postverbal constituent in some stimuli for subject focus), prosody is the obvious suspect for the variation in accuracy shown in Figure 3. Crucially, the variation in accuracy allows us to determine whether accuracy depends on the stimuli's prosodic properties in that prosodically well-formed or prototypical stimuli receive higher accuracy scores than non-prototypical stimuli. If prosody predicts accuracy, we may assume that the violations occur already during production. If prosody does not predict accuracy, we may assume that the violations occur during perception. To determine which of these two scenarios holds true, we conducted a post-hoc analysis of the prosodic properties of a subset of our stimuli and connected the stimuli's prosodic properties with their accuracy. This analysis is presented in Section 4.

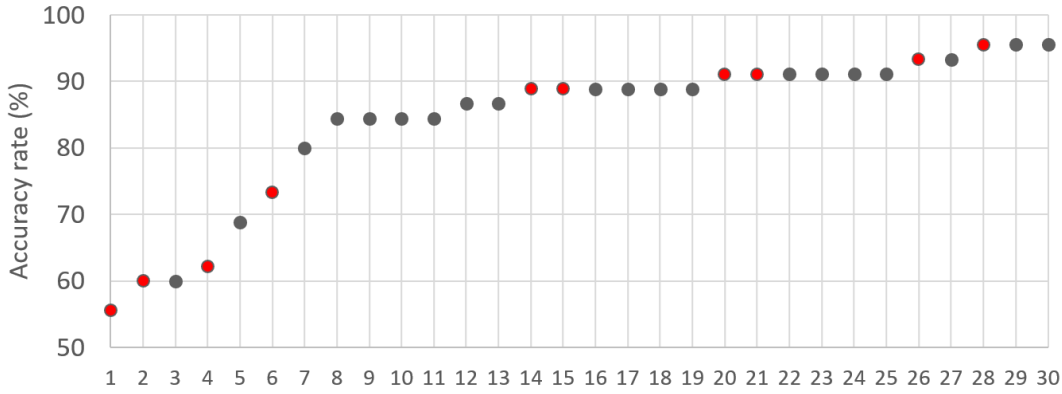
4. Prosody as a factor for focus recognition

In the following, the reader is provided with the relevant information on the prosodic post-hoc analysis performed on the stimuli judged by listeners from Spain and Argentina in the perception experiment described in Section 3.

4.1. Prosody and accuracy

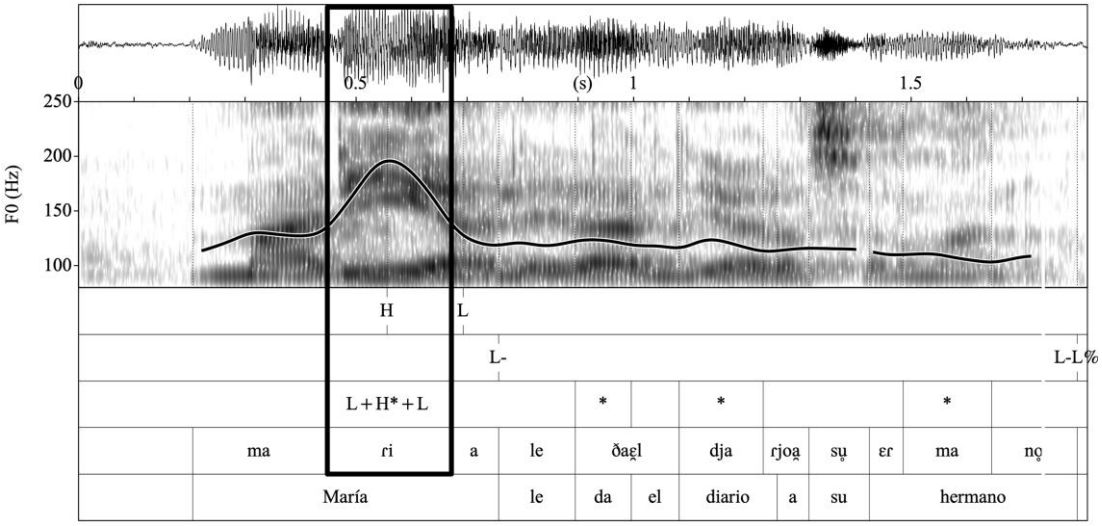
We performed a prosodic post-hoc analysis on 30 syntactically and lexically identical stimuli of the form *María le da el diario a su hermano* ([S]$_F$-V-dO-iO). All 30 stimuli were evaluated in the previously described perception experiment with regard to the correct recognition of the subject focus. Figure 4 shows the variation in accuracy rate of these 30 stimuli.

Figure 4. Accuracy rate of 30 stimuli used in post-hoc analysis (red = information focus; grey = contrastive focus).



Given the otherwise identical external form of the stimuli, the considerable differences (between 56 and 96%) in accuracy rate of the subject focus are essential when it comes to determine the locus of FPR violations: If the different accuracy rates correlate with the prosodic shape of the stimuli, it can be assumed that low accuracy rates are caused by violations of the FPR during production. If, by contrast, there are no respective correlations, it can be assumed that the participants misinterpreted prosodically well-formed stimuli and that the FPR violations occur during perception. Given that in Argentinean Spanish clause-initial focused subjects are typically marked through a tritonal pitch accent consisting of an F0 rise and fall within the temporal limits of the metrically strong syllable, i.e., L+H*+L (Gabriel et al. 2010; see also Toledo 1989 and Lang-Rigal 2011), a canonical realization of a stimulus with a subject XP focused *in situ* should include both a high (H) and a low (L) target aligned with the stressed syllable of the focused XP. The post-focal material should be produced as a low F0 plateau, i.e., with post-focal deaccentuation. An example is given in Figure 5. Note that a low intermediate phrasal boundary (L-) is set at the right edge of the focus domain.

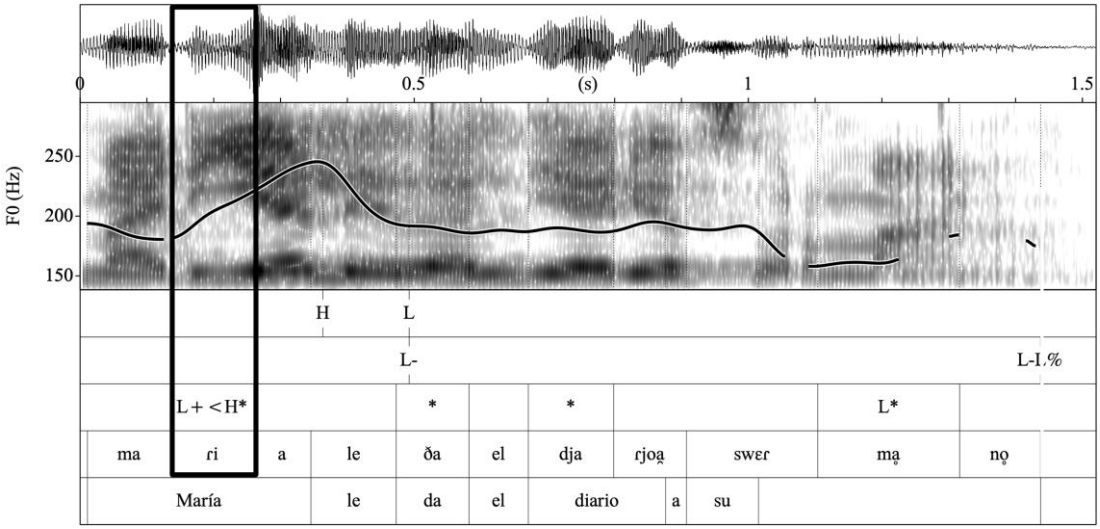
Figure 5. Canonical realization of a stimulus produced under the subject-focus condition with a focal pitch accent L+H*+L. Tiers (from top to bottom): (i) positions of the H and L targets; (ii) boundary tones; (iii) pitch accents; (iv) phonetic transcription (syllables); (v) orthographic transcription (words). Here and in the following figures, metrically strong syllables lacking noteworthy pitch movements are annotated using the star symbol '*' without any tonal specification.



As expected, the stimulus depicted in Figure 5 attained a high accuracy rate in the perception experiment: 91%. The remaining 9% of erroneous interpretations as object focus should be interpreted as cases of mismatches caused by violations of the FPR in perception. By contrast, a non-canonical realization with a delayed peak (L+<H*) as the one represented in Figure 6 is only correctly recognized as subject focus in 60% of the cases. We interpret the remaining 40% cases as mismatches caused by violations of the FPR in production. As can be seen in Figure 6, the alignment properties of both the H and the L target crucially differ from those of the stimulus depicted in Figure 5. [11] In addition, the low target (L) does not mark the end of the focused word *María* but is considerably shifted to the right. Further, the post-focal stretch comprises an L* pitch accent, i.e., it is not entirely deaccented.

[11] Figures 5–10 were produced using Praat (Boersma & Hayes 2021).

Figure 6. Non-canonical realization of a stimulus produced under the subject-focus condition with a focal pitch accent L+<H*. Tiers (from top to bottom): (i) positions of the H and L targets; (ii) boundary tones; (iii) pitch accents; (iv) phonetic transcription (syllables); (v) orthographic transcription (words).

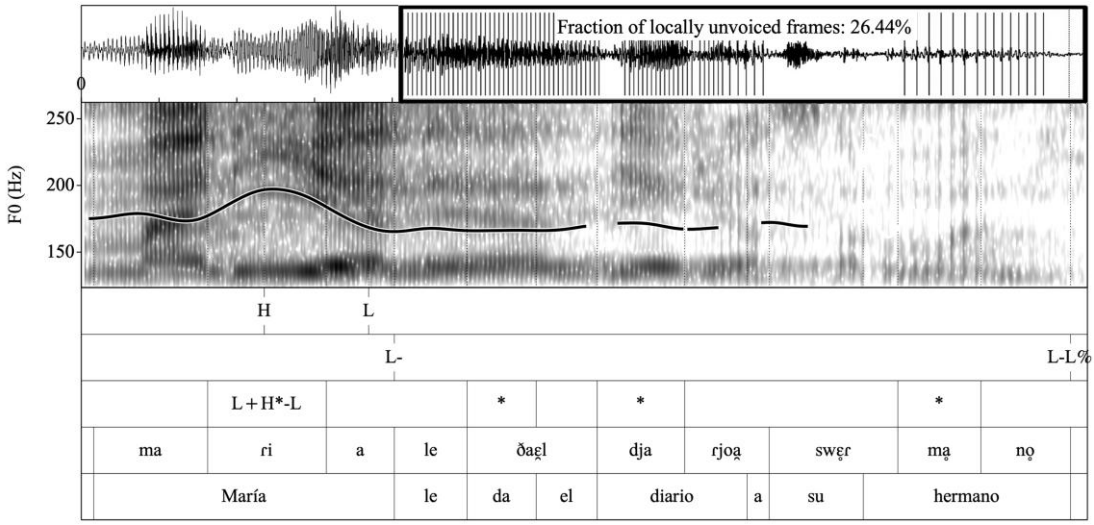


In the remainder of this section, we first introduce the methodology of the prosodic analysis performed on the data (Section 4.2), before presenting the results (Section 4.3) and discussing them in the context of our research question, i.e., the question of the locus of FPR violations, and the current literature (Section 4.4).

4.2. Parameters of the prosodic analysis

Based on studies showing that changes in voicing and voice quality can signal information-structural categories such as focus (see Sluijter & van Heuven 1996; Yanushevskaya et al. 2016), we investigated the degree of (de)voicing of the post-focal domain as a possible indication of the information-structural category *background* as part of an exploratory analysis. In this vein, we followed the standard assumption that the degree of voicing is higher in more prominent parts of speech as in e.g., focused constituents, which, in turn, means that non-focal material should tend to be devoiced. This criterion was operationalized using the Praat function Voice Report (Fraction of locally unvoiced frames). An example is given in Figure 7.

Figure 7. Post-focal devoicing at a rate of 26.44% (Fraction of locally unvoiced frames). The highlighted portion of the utterance, which corresponds to the post-focal stretch of speech, shows missing glottal pulses as a correlate of devoicing.



The remaining prosodic cues of the focus-background articulation we considered, namely F0 alignment and scaling as well as duration, were inspired by Vanrell et al. (2013). Regarding the F0-related parameters, we took into account, first, the alignment of the focal pitch peak (H*) and the following L target as either part of the tritonal L+H*+L pitch accent or the intermediate boundary tone L- (14–16), second, the scaling of the focal pitch accent (17), and, third, the steepness of the F0 descent following the focal high target (18). As concerns alignment, the following three parameters are considered:

(14) AlignH-σ]: alignment of the high target with respect to the stressed syllable σ, i.e., position of H* expressed as the percentage (%) of the duration of σ

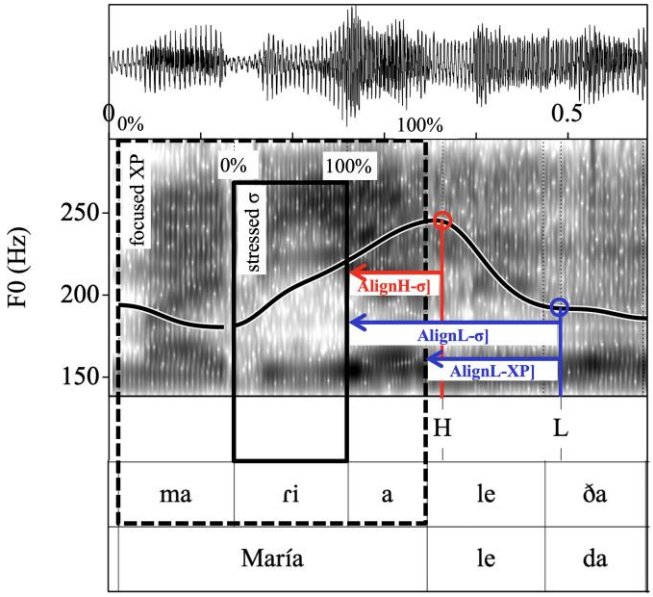
(15) AlignL-XP]: alignment of the low target with respect to the focused constituent XP, i.e., position of L (trailing tone of L+H*+L or L-) expressed as the percentage (%) of the duration of XP

(16) AlignL-σ]: alignment of the low target with respect to the stressed syllable σ, i.e., position of L (trailing tone of L+H*+L or L-) expressed as the percentage (%) of the duration of σ

For all three alignment-related parameters given in (14–16), the lower the percentage value, the better the focus should be recognized. Note that the percentage can be below or above 100%, depending on the position of the respective target in relation to the stressed syllable or focused constituent. For instance, with a tritonal L+H*+L pitch accent the peak is obligatorily located within the temporal limits of the metrically strong syllable, which means that the value for (14) AlignH-σ] is below 100%. With a rising L+H* pitch accent, whose peak is located around the end of the stressed syllable, the value is around 100%, and, finally, in case of a delayed pitch peak (L+<H*) the

value is necessarily greater than 100%. In Figure 8, the three alignment-related parameters (14–16) are exemplified by means of a non-canonical realization of a clause-initial focused subject XP with a delayed pitch accent.

Figure 8. Alignment-related parameters illustrated with the example of an L+<H* pitch accent: AlignH-σ], AlignL-XP], and AlignL-σ].



Note that the phonological analysis of the low target highlighted in Figure 8 in terms of ToBi labels is not relevant at this point. Following the analysis of Argentinean Spanish proposed by Gabriel et al. (2010), this L tone either can be part of the tritonal pitch accent L+H*+L or correspond to a low intermediate boundary tone L- indicating the end of the focused constituent. In European Spanish, which lacks tritonal pitch accents and signals focus through early-peak alignment of the bitonal rising pitch accent (i.e., L+H*; see Face 2002a; 2002b), this low target can only be interpreted as an L- boundary tone. However, we are only concerned here with the phonetic surface of the F0 contour.

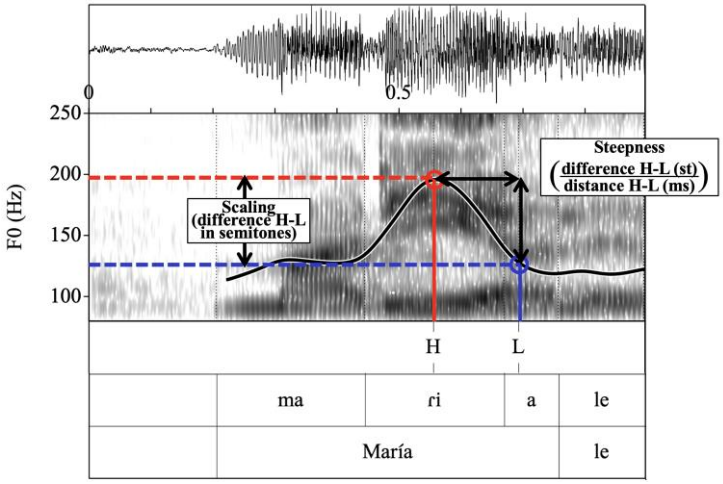
The scaling of the focal pitch accent and the steepness of the fall are operationalized as described in (17) and (18), respectively.

(17) Scaling: pitch difference between the high (H) and subsequent low (L) target (in semitones)

(18) Steepness: relation between the pitch difference between H and L (in semitones) and the distance between H and L (in milliseconds), i.e., the quotient resulting from dividing the pitch difference H–L (st) by the distance H–L (ms)

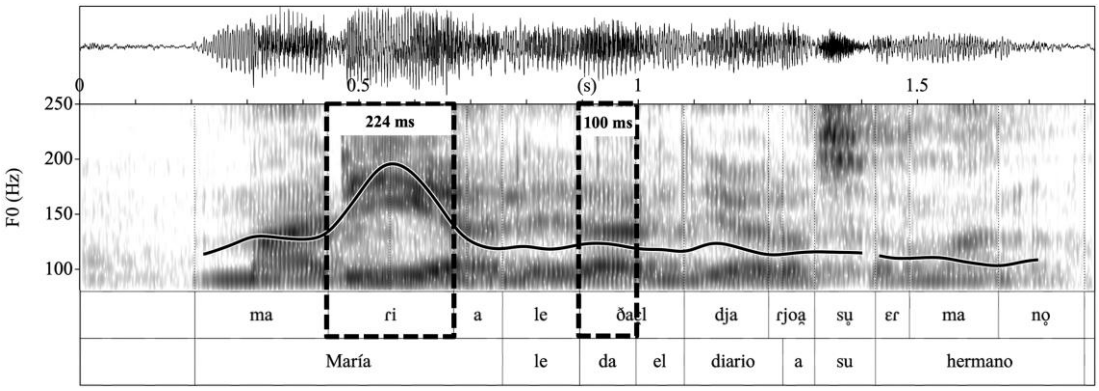
It holds for both scaling and steepness that the higher the corresponding value, the better the focus should be recognized. In Figure 9, these two parameters are illustrated with the example of a canonical realization of a clause-initial subject-focus XP marked through a L+H*+L pitch accent.

Figure 9. Scaling and steepness of the fall illustrated with the example of an L+H* +L pitch accent.



Turning to the durational properties of the stimuli used in the perception experiment, we considered the duration of the nuclear syllable of the focused constituent in relation to the first stressed syllable of the post-focal domain (i.e., the quotient resulting from the division of the two durations). This is exemplified in Figure 10. We expect that a longer duration of the stressed syllable of the focused constituent as compared to the duration of the following stressed syllable results in a better recognition of the focus. In other words, the greater value of the quotient resulting from the division of the two durations, the better the focus should be recognized.

Figure 10. Duration of the nuclear syllable of the focused XP (here: 224 ms) in relation to the first stressed syllable of the post-focal stretch (here: 100 ms).



In addition to the measures performed on the 30 stimuli according to the parameters outlined so far, we annotated the pitch accents and boundary tones in accordance with the labeling system “Tone and break indices” for Spanish (Sp_ToBI, following Hualde & Prieto 2015 as well as Gabriel et al. 2010 regarding the labeling conventions for Argentinean Spanish).

Note that these parameters involve both prominence and alignment in the sense of Féry (2013). While scaling, steepness, and duration are related to prominence (see Féry 2013: 685), AlignL-XP] concerns the alignment of L (i.e., the trailing tone of L+H*+L or a boundary tone L-) with the right edge of the focused constituent.

4.3. Results

In this section, we connect the accuracy rates from the perception experiment with the results of both the phonetic measurements introduced in Section 4.2 and the ToBI annotation (the Appendix provides for each of the 30 stimuli the phonetic measurements, the ToBI annotation and the accuracy rate). Starting with the pitch accents realized on the focused constituent [*María*]$_F$, it is noticeable that the stimuli produced with a canonical tritonal L+H*+L pitch accent attained the highest mean accuracy rate, while those of the stimuli produced with one of the less canonical bitonal pitch accents (L+H*, L+<H*) were lower. In Table 7, below, the absolute frequency and the mean accuracy rates are given for each of the three pitch accents.[12]

Table 7. Absolute frequency and mean accuracy rate (%) for the pitch accents L+<H*, L+H* and L+H*+L.

Pitch accent	L+<H*	L+H*	L+H*+L
n	4	12	14
Spain (%)	65.0	76.0	87.1
Argentina (%)	66.3	88.8	97.1
Spain + Argentina (%)	65.6	81.7	91.6

Since Sp_TobI labels are only abstract representations of the alignment and scaling properties of the F0 excursion related to the temporal limits of the stressed syllable, it is not surprising that the expectations formulated in Section 4.1. regarding the individual prosodic parameters are reflected in the results. For instance, as illustrated in Table 9 in the Appendix, the AlignH-σ] values for the "best-rated" group of stimuli, i.e., those marked through the canonical L+H*+L pitch accent, range between 29 and 66% (mean = 49%), i.e., they are all considerably below 100%, meaning that the pitch peak and the subsequent fall occur within the temporal limits of the stressed syllable. The Sp_ToBI labels are consequently not included as factors in the statistical analysis, since they are mirrored in the results obtained from the measurements performed on the data with respect to the F0-related parameters introduced in Section 4.2.

To determine which of the F0-related parameters have the greatest effect on focus detection and how duration and voice quality relate to this, we conducted in a first step a multiple linear regression analysis (with fixed factors) including all seven prosodic parameters. Although the analysis identifies several parameters as significant factors for focus recognition, it also shows substantial signs of multicollinearity, as indicated by extremely high VIF values (VIF = variance inflation factor) for certain predictors in the model, i.e., Scaling: 6.689366; Steepness: 6.833440; AlignL-σ]: 39.224422; AlignL-XP]: 49.975922 (as tested by the "vif" function from the car package in R; Fox & Weisberg 2019). VIFs above 5 warrant some caution while values above 10 are indicative of serious collinearity. Given this collinearity of predictors, the requirements for a multiple linear regression analysis are not met. We therefore proceed with a relative weight analysis (Johnson 2000) as an alternative. The relative weight is the proportionate contribution of each predictor to R2, i.e., the amount of

[12] Since the number of participants in the perception experiment was not the same for Spain and Argentina, the accuracy for Spain + Argentina does not necessarily amount to the mean of the accuracy for Spain and the accuracy for Argentina.

observed variability that is accounted for by the predictors. Additionally, the analysis of significance of relative weights was calculated (Tonidandel et al. 2009; we used the web application Tonidandel & LeBreton (n. d.) and did back up checks using R for relative weights and R2). Separate analyses were conducted for three data sets: participants from (i) Spain and Argentina, (ii) Spain only, (iii) Argentina only.

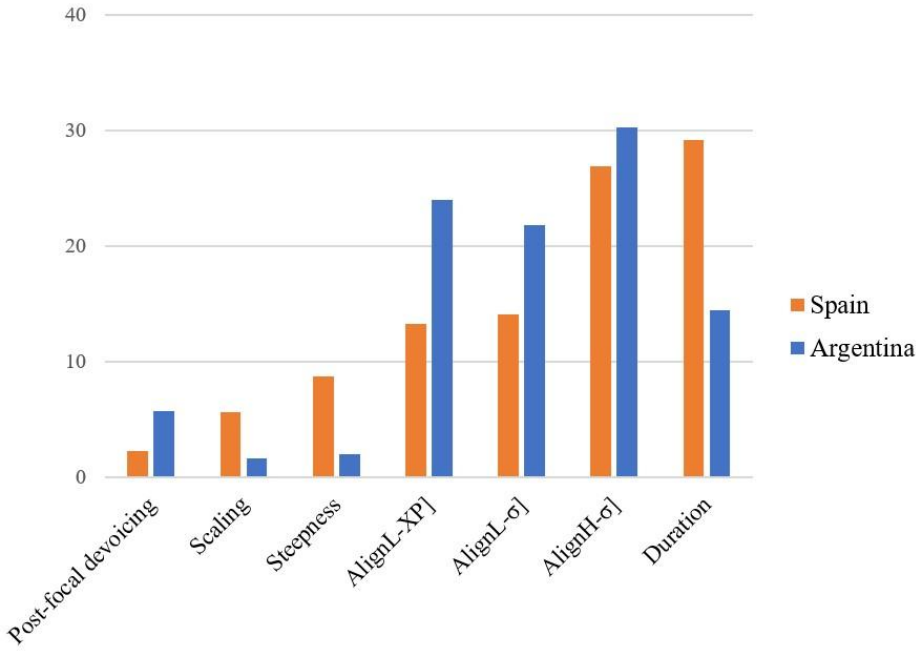
The prosodic parameters account for about 80% of the variability of accuracy (R2 = 0.8075), which indicates that prosody has a strong impact on accuracy. Table 8 shows the relative weights of the seven prosodic predictors calculated on the basis of the whole data set (i). As shown by the relative weights in Table 8, the alignment-related parameters AlignL-σ], AlignL-XP], and AlignH-σ] as well as the duration-related parameter Duration have the strongest impact on accuracy. Hence, both prominence and alignment seem to play a role in focus recognition (although the impact of the prominence related parameters Scaling and Steepness is minimal).

Table 8. Relative weights of prosodic predictors for accuracy (participants from Argentina and Spain; significant predictors in bold face).

Variables	Raw Relative Weight	Rescaled Relative Weight
Scaling	0.0259	3.21
Steepness	0.031	3.83
Post-focal devoicing	0.0325	4.03
AlignL-σ]	0.1518	18.79
AlignL-XP]	0.1561	19.33
Duration	0.168	20.81
AlignH-σ]	0.2422	30

In Figure 11, the rescaled relative weights of the prosodic parameters are given as calculated on the basis of data set (ii) (participants from Spain only) and of data set (iii) (participants from Argentina only).

Figure 11. Rescaled relative weights of prosodic predictors for accuracy.



Looking at the impact of prosodic parameters on accuracy separately for Spanish and Argentinean participants, two aspects are particularly striking. While the Spanish participants seem to rely more strongly on the factor Duration (i.e., on the length of the nuclear syllable of the focused XP compared to that of the following metrically strong syllable), the alignment-based parameters AlignH-σ], AlignL-XP], and AlignL-σ] are more relevant for the detection of the focus-background partition by the Argentinean participants. The fact that Spaniards rely more on the duration of stressed syllables than participants from Argentina do when determining the focus-background partition of a given stimulus is probably due to the fact that in Argentinean Spanish stressed syllables are significantly longer than unstressed syllables, whereas in European Spanish there is no significant difference in duration between stressed and unstressed syllables (Estebas-Vilaplana 2010; Gabriel & Kireva 2014).[13] In other words, a significantly lengthened syllable probably sounds more salient to Spanish ears than to Argentinian ears, which is why this cue is presumably used more for focus recognition. Regarding the question of why the alignment-based parameters are used more for focus detection by Argentineans than by Spaniards, it can be said that the former are more sensitive to the absence of the early alignment of the focus accent typical of Argentinean Spanish. Turning to the parameters that turned out not to be strong predictors of focus recognition, it is noticeable that Spanish listeners seem to rely more of the steepness of the fall H-L and on the scaling of the focal pitch accent than Argentinean listeners do. The greater sensitivity of Spanish listeners to the sharp pitch drop after the focal high tone might be explained by the fact that it is more salient to the ears of Spanish listeners, since Argentinean listeners, whose native tone repertoire includes the tritonal pitch accent L+H*+L with its obligatory fall after the pitch peak within the limits of the stressed syllable, are more accustomed to such tonal movements. However, it is surprising that Argentineans react more sensitively to scaling than Spaniards, as the former have a scaling-based contrast in their tonal system with the functional opposition L+H*+L (used to signal contrastiveness and emphasis) vs. L* (used as a phrase-final nuclear accents in unmarked contexts; see Feldhausen et al. 2011). This requires further research; the same holds true for the question of why Argentines react slightly more sensitively to post-focal devoicing.

4.4. Discussion

The FPR is often described as a robust mapping mechanism between prosody and information structure. Nevertheless, violations of the FPR are not infrequent (Krüger 2009; Breen et al. 2010; Calhoun et al. 2018; El Zarka & Hödl 2021; Gabriel & Heidinger 2022). Given that violations of the FPR may occur during production or perception/interpretation, the question emerges at which mapping FPR violations primarily occur. The prosodic analysis of 30 syntactically and lexically identical stimuli has shown that the accuracy rate of a given stimulus strongly depends on the prosodic

[13] A similar result was reported for Lecce Italian by Vanrell et al. (2013: 215), who attribute the increased duration of the tonic syllable to the complexity of the (rising-falling) F0 movement, which, in turn, resembles the Argentinean one. The role of the stressed syllable's increased duration under certain conditions in Peninsular Spanish such as *verum focus* was pointed out by Escandell-Vidal (2011). Durational differences between stressed accented, stressed unaccented, and unstressed syllables in Castilian Spanish are reported in Prieto et al. (2012).

properties of that stimulus (see Section 4.3). Since prosody predicts accuracy, we must conclude that the violations of the FPR occur already during production. If prosody had no impact on accuracy, we would have to assume that the violations occurred during perception.

The conclusion that in our data violations of the FPR happen primarily during production is in line with recent studies on the prosody-focus interface which also detect such violations during production. Calhoun et al. (2018) and Feldhausen & Vanrell (2015) presented data from production experiments where the nuclear accent does not fall on the focused constituent. In comparison to these two studies the present paper adds the value of combing a prosodic analysis of production data with data on focus recognition. El Zarka & Hödl (2021), by contrast, have identified violations of FPR primarily during perception. The authors report that auditory stimuli which are "especially clear examples" (El Zarka & Hödl 2021: 59) show low rate of correct focus recognition. Although such cases cannot be excluded for our data set, they must be rather infrequent.

The result that focus recognition mainly depends on the prosodic properties of the stimuli (and that violations of the FPR thus tend to occur during production) raises further questions about the distinction between prominence vs. alignment as two types of focus realization (Féry 2013). According to Féry (2013: 685), a prominent constituent is realized with acoustic correlates (duration, pitch, intensity) that increase its perceptibility, which links prominence directly to focus recognition. No such increase in perceptibility is mentioned for aligned constituents. It would therefore be interesting to scrutinize the role that prominence and alignment play in focus recognition. In the perception experiment (Gabriel & Heidinger 2022), we counted as violations of the FPR all cases where the focus of the stimulus and the perceived focus do not coincide. Since this way of determining FPR violations does not consider the prosodic properties of the stimuli, it is not sensitive to the distinction between prominence and alignment. In the post-hoc analysis, however, we did determine the impact of several prosodic parameters on focus recognition. While most of them are related to prosodic prominence (in Féry's terms), e.g., duration, the alignment of the L tone with the focused constituent (AlignL-XP]) is most probably linked to the alignment between the focused constituent and a boundary tone. Crucially, the results reported in Section 4.3 show that this alignment related prosodic parameter has a considerable impact on focus recognition – and it has a stronger impact than some prominence related parameters (e.g., scaling, steepness).

On a more general level the study of violations of the FPR broadens our understanding of interface phenomena and helps us understand whether the vulnerability of interfaces (see, e.g., Sorace 2012) not only refers to second language (L2) learning and language contact but also extends to monolingual language use and whether the respective mappings at the interfaces are symmetrical. The results from our study suggest that in Spanish, the interface between prosody and focus indeed shows some vulnerability in that violations of the FPR are not infrequent. Moreover, our data suggests for Spanish that the two mappings at this interface are not equally vulnerable, as violations primarily occur during production. In light of the distinction between prominence and alignment in focus realization (Féry 2013), we should add that our view of interface vulnerability is rather broad. It includes cases where prosodic means for focus realization are used, but do not suffice for correct focus detection. If the focused constituent and a prosodic boundary are aligned but this alignment does not signal

focus, we would still consider this a case of interface vulnerability since information is lost at the prosody-focus interface.

Further, the outcomes of the present study are also relevant for the ongoing discussion on focus marking. Previous studies on the predictive power of sentence form in Spanish suggest that sentence forms often remain ambiguous with respect to focus-background partition (see Heidinger 2016 and 2018, based on data collected in production experiments). Dufter & Gabriel (2016: 422) also point out that sentences without further context often remain ambiguous regarding the information-structural reading they convey. In the same vein, Zimmermann & Onea (2011: 1658) state that ambiguity with respect to focus-background partition is a frequent and even expected property of sentence form. According to these authors, it is expected due to the nature of the category focus, which is a cognitive category that operates at the level of information structure and not a genuine linguistic category. The main result of the present study, i.e., that focus is not always prosodically signaled in an unequivocal way, is fully in line with the above characterizations of focus marking. Evidence pointing in the same direction have been reported in Gili-Fivela (2009) and Borràs-Comes et al. (2014) with respect to the discrimination between sentences with and without contrastive focus based on prosodic cues.

5. Conclusions

Starting from the observation that the focus prominence rule is frequently violated in Spanish (Gabriel & Heidinger 2022), we addressed the question of whether these violations occur primarily during production or perception. To answer this question, we conducted a post-hoc prosodic analysis of 30 stimuli and combined the results from this analysis with the stimuli's accuracy rate, i.e., the rate of correct focus recognition.

Generally speaking, the prosodic properties of the stimuli determine the stimuli's accuracy rate to a large extent. This suggests that violations of the FPR occur primarily during production: Prosodically "bad", i.e., non-canonical stimuli exhibit lower accuracy scores than prosodically well-formed ones. As an additional result, the post-hoc prosodic analysis showed that different prosodic parameters contribute to focus recognition to a different degree (e.g., F0 alignment is more important than post-focal devoicing) and that speakers of different varieties of Spanish rely on different prosodic cues in focus recognition.

Several issues for future research arise from the present study. As concerns focus recognition, it would be desirable to include stimuli produced by native speakers of European Spanish and then compare accuracy rates of listeners from Spain with those obtained from Argentineans. More generally, the current research paradigm could be easily expanded by including more diatopic varieties. As to the link between prosodic properties and accuracy, it would be interesting to see which prosodic parameters are relevant for focus recognition when the focus is located in sentence-final position (recall that our post-hoc analysis only included sentence-initial foci). Including sentence-final foci would then also allow to scrutinize the role of prominence vs. alignment in focus realization and recognition. Finally, taking into consideration the reaction time during perception may highlight differences between stimuli which are prosodically distinct but show similar accuracy.

In view of the numerous violations of the FPR at the production level, it is legitimate to ask whether it needs to be modified or even discarded altogether. Ultimately, the decision depends on how the concept of the rule is defined. If it is understood as an irrevocable law that applies without exception, such a rule cannot be upheld if numerous prosodic productions fail to comply with it to varying degrees. However, if a rule is understood more in the sense of an optimality-theoretic constraint, which can be violated depending on its position in the constraint hierarchy (Prince & Smolensky 2004; see Section 2.1.1), the FPR can easily be maintained as a valid guideline for the prosodic realization of focus constructions.

Acknowledgments

We are grateful to three anonymous reviewers and to the editors of the special issue, Katrin Schmitz and Tim Diaubalick, whose most helpful comments and suggestions contributed significantly to improving the quality of the paper. Many thanks go to Petra Hödl for her support in the statistical analysis of our data. All remaining errors and shortcomings are our responsibility.

References

Beckman, Mary, Díaz-Campos, Manuel, McGory, Julia Tevis, & Morgan, Terrell, A. 2002. Intonation across Spanish, in the Tones and Break Indices framework. *Probus* 14: 9-36. DOI: 10.1515/prbs.2002.008

Boersma, Paul, & Hayes, David. 2021. *Praat. Doing phonetics by computer* 6.1.47. https://www.fon.hum.uva.nl/praat/ (accessed: July 24, 2024).

Borràs-Comes, Joan, Vanrell, Maria del Mar, & Prieto, Pilar. 2014. The role of pitch range in establishing intonational contrasts. *Journal of the International Phonetic Association* 44(1): 1-20. DOI: 10.1017/S0025100313000303

Bosque, Ignacio, & Gutiérrez-Rexach, Javier. 2009. *Fundamentos de sintaxis formal.* Madrid: Akal.

Breen, Mara, Fedorenko, Evelina, Wagner, Michael, & Gibson, Edward. 2010. Acoustic correlates of information structure. *Language and Cognitive Processes* 25: 1044-1098. DOI: 10.1080/01690965.2010.504378

Büring, Daniel. 2009. Towards a typology of focus realization. In M. Zimmermann, & C. Féry (eds). *Information structure*, 177-205. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199570959.003.0008

Calhoun, Sasha, La Cruz, Erwin, & Olssen, Anna. 2018. The interplay of information structure, semantics, prosody, and word ordering in Spanish intransitives. *Laboratory Phonology* 9(1): 8, 1-30. DOI: 10.5334/labphon.65

Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (ed). *Subject and topic*, 25-55. New York: Academic Press.

Dufter, Andreas, & Gabriel, Christoph. 2016. Information structure, prosody, and word order. In S. Fischer, & C. Gabriel (eds). *Manual of grammatical interfaces in Romance*, 419-455. Berlin: De Gruyter. DOI: 10.1515/9783110311860-017

El Zarka, Dina, & Hödl, Petra. 2021. A study on the perception of prosodic cues to focus by Egyptian listeners. Some make use of them, but most of them don't. *Speech Communication* 132: 55-69. DOI: 10.1016/j.specom.2021.05.002

Escandell-Vidal, Victoria. 2011. Verum focus y prosodia: cuando la duración (sí que) importa. *Oralia* 14: 181-201. DOI: 10.25115/oralia.v14i.8186

Estebas-Vilaplana, Eva. 2010. The role of duration in intonational modelling: a comparative study of Peninsular and Argentinean Spanish. *Revista española de lingüística aplicada* 23: 153-174.

Face, Timothy. 2002a. Focus and early peak alignment in Spanish intonation. *Probus* 13: 223-246. DOI: 10.1515/prbs.2001.004

Face, Timothy. L. 2002b. *Intonational marking of contrastive focus in Madrid Spanish.* Munich: Lincom Europa.

Feldhausen, Ingo, Pešková, Andrea, Kireva, Elena, & Gabriel, Christoph. 2011. Categorical perception of *Porteño* nuclear accents. In W.-S. Lee, & E. Zee (eds). *Proceedings of the 17th International Congress of Phonetic Sciences 2011*, 116-119. Hong Kong, China.
https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/SpecialSession/Session4/Feldhausen/Feldhausen.pdf

Feldhausen, Ingo, & Vanrell, Maria del Mar. 2015. Oraciones hendidas y marcación del foco estrecho en español. Una aproximación desde la Teoría de la Optimidad Estocástica. *Revista internacional de lingüística iberoamericana* 26: 39-60. DOI: 10.31819/rili-2015-132604

Féry, Caroline. 2013. Focus as prosodic alignment. *Natural Language and Linguistic Theory* 31: 683-734. DOI: 10.1007/s11049-013-9195-7

Fox, John, & Weisberg, Sanford. 2019. *An R companion to applied regression.* Third edition. Thousand Oaks, CA: Sage.

Gabriel, Christoph. 2010. On focus, prosody, and word order in Argentinean Spanish: A minimalist OT account. *Revista virtual de estudos da linguagem* 4: 183-222. http://www.revel.inf.br/files/artigos/revel_special_4_on_focus_prosody_and_word_order.pdf (accessed: July 24, 2024).

Gabriel, Christoph. 2021. La prosodia de las variedades americanas. In Eckkrammer, E. M. (ed). *Manual del español en América*, 465-487. Berlin: De Gruyter. DOI: 10.1515/9783110334845-028

Gabriel, Christoph, Feldhausen, Ingo, Pešková, Andrea, Colantoni, Laura, Lee, Su-Ar, Arana, Valeria, & Labastía, Leopoldo. 2010. Argentinian Spanish intonation. In P. Prieto, & P. Roseano (eds). *Transcription of intonation of the Spanish language*, 285-317. München: Lincom.

Gabriel, Christoph, & Kireva, Elena. 2014. Prosodic transfer in learner and contact varieties. Speech rhythm and intonation of Buenos Aires Spanish and L2 Castilian Spanish produced by Italian native speakers. *Studies in Second Language Acquisition* 36: 257-281. DOI: 10.1017/S0272263113000740

Gabriel, Christoph, & Heidinger, Steffen. 2022. The focus prominence rule in Spanish from a perception perspective. *Borealis. An International Journal of Hispanic Linguistics* 11, 141–172. DOI: 10.7557/1.11.1.6483

Gili-Fivela, Barbara. 2009. From production to perception and back. An analysis of two pitch accents. In S. Fuchs, H. Loevenbruck, D. Pape, & P. Perrier (eds). *Some aspects of speech and the brain*, 363-405. Frankfurt: Lang.

Heidinger, Steffen. 2015. Optionality and preferences in Spanish postverbal constituent order: An OT account without basic constituent order. *Lingua* 162: 102-127. DOI: 10.1016/j.lingua.2015.05.003

Heidinger, Steffen. 2016. El orden de los constituyentes posverbales y la expresión del foco informativo en español. In A. Ledgeway, M. Cennamo, & G. Mensching (eds), *Actes du XXVIIe Congrès international de linguistique et de philologie romanes (Nancy, 15-20 juillet 2013). Section 4: Syntaxe*, 217-232. Nancy: ATILF.

Heidinger, Steffen. 2018. *Sekundäre Prädikation und Informationsstruktur: Fokus und Informationsstatus bei spanischen Depiktiven.* Frankfurt: Lang. DOI: 10.3726/b14723

Höhle, Tilman N. 1982. Explikationen für “normale Betonung” und “normale Wortstellung”. In W. Abraham (ed), *Satzglieder im Deutschen. Vorschläge zur syntaktischen, semantischen und pragmatischen Fundierung*, 75-154. Tübingen: Narr.

Hoot, Bradley. 2012. *Presentational focus in heritage and monolingual Spanish.* PhD thesis, University of Illinois at Chicago: Chicago. https://eric.ed.gov/?id=ED552262 (accessed: July 24, 2024).

Hoot, Bradley. 2016. Narrow presentational focus in Mexican Spanish. Experimental evidence. *Probus* 28: 335-365. DOI: 10.1515/probus-2014-0004

Hoot, Bradley. 2017. Narrow presentational focus in heritage Spanish and the syntax-discourse interface. *Linguistic Approaches to Bilingualism* 7: 63–95. DOI: 10.1075/lab.14021.hoo

Hualde, José Ignacio. 2005. *The sounds of Spanish.* Cambridge: Cambridge University Press.

Hualde, José Ignacio, & Prieto, Pilar. 2015. Intonational variation in Spanish. European and American varieties. In S. Frota, & P. Prieto (eds), *Intonation in Romance*, 350-391. Oxford: Oxford University Press. DOI:10.1093/acprof:oso/9780199685332.003.0010

Jackendoff, Ray. 1972. *Semantic interpretation in Generative Grammar*. Cambridge, MA: MIT Press.

Johnson, JW. 2000. A heuristic method for estimating the relative weight of predictor variables in Multiple Regression. *Multivariate behavioral research* 35: 1-19. DOI: 10.1207/S15327906MBR3501_1

Krüger, Martina. 2009. *Produktion und Perzeption von Fokus im Deutschen*. MA thesis, Köln: Universität zu Köln.

Lang-Rigal, Jenifer. 2011. Perception of narrow focus prosody in Buenos Aires Spanish. In S. M. Alvord (ed). *Selected proceedings of the 5th Conference on Laboratory Approaches to Romance Phonology*, 118-126. Somerville, MA: Cascadilla.

Matić, Dejan, & Wedgwood, Daniel. 2013. The meanings of focus. The significance of an interpretation-based category in cross-linguistic analysis. *Journal of Linguistics* 49: 127-163. DOI:10.1017/S0022226712000345

Olarrea, Antxon. 2012. Word order and information structure. In J. I. Hualde, A. Olarrea, & E. O'Rourke (eds). *The handbook of Hispanic linguistics*, 603-627. Malden, MA: Wiley-Blackwell. DOI: 10.1002/9781118228098.ch28

Prieto, Pilar, Vanrell, Maria del Mar, Astruc, Llüísa, Payne, Elinor, Post, Brechtje. 2012. Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication* 54: 681-702. DOI: 10.1016/j.specom.2011.12.001

Prieto, Pilar, & Roseano, Paolo (eds). 2009-2010. *Atlas interactivo de la entonación del español*. Barcelona: Grup d'estudis de prosòdia. https://prosodia.upf.edu/atlasentonacion/ (accessed: July 24, 2024).

Prince, Alan, & Paul Smolensky. 2004. *Optimality Theory. Constraint interaction in Generative Grammar.* Malden, MA: Blackwell. DOI: 10.1002/9780470759400

Rialland, Anne, & Robert, Stéphane. 2001. The intonational system of Wolof. *Linguistics* 39: 893-939. DOI: 10.1515/ling.2001.038

Roessig, Simon. 2021. *Categoriality and continuity in prosodic prominence.* Berlin: Language Science Press. DOI: 10.5281/zenodo.4121875

Roettger, Timo B., Mahrt, Tim, & Cole, Jennifer. 2019. Mapping prosody onto meaning – the case of information structure in American English. *Language, Cognition and Neuroscience* 34: 841-860. DOI: 10.1080/23273798.2019.1587482

Samek-Lodovici, Vieri. 2005. Prosody-syntax interaction in the expression of focus. *Natural Language and Linguistic Theory* 23: 687-755. DOI: 10.1007/s11049-004-2874-7

Samek-Lodovici, Vieri. 2016. Constraint conflict and information structure. In C. Féry, & S. Ishihara (eds), *The Oxford Handbook of Information Structure*, 203-223. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199642670.013.27

Selkirk, Elizabeth O. 1984. *Phonology and syntax. The relation between sound and structure.* Cambridge, MA: MIT Press.

Sluijter, Agaath M. C., & van Heuven, Vincent J. 1996. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100: 2471-2485. DOI: 10.1121/1.417955

Sorace, Antonella. 2012. Pinning down the concept of interface in bilingual development. A reply to peer commentaries. *Linguistic Approaches to Bilingualism* 2: 209-216. DOI: 10.1075/lab.2.2.04sor

Toledo, Guillermo Andrés. 1989. Señales prosódicas del foco. *Revista argentina de lingüística* 5: 205-230.

Tonidandel, Scott, & LeBreton, James M. no date. Relative importance analysis: programs for calculating relative weights in multiple, multivariate and logistic regression (web application). https://rwa-web.shinyapps.io/multipleregression/ (accessed: July 24, 2024).

Tonidandel, Scott, Lebreton, James M., & Johnson, Jeff W. 2009. Determining the statistical significance of relative weights. *Psychological methods* 14: 387-399. DOI: 10.1037/a0017735

Vallduví, Enric. 1990. *The informational component*. PhD thesis, Philadelphia: University of Pennsylvania. https://core.ac.uk/download/pdf/76379509.pdf (accessed: July 24, 2024).

Vallduví, Enric, & Engdahl, Elisabet. 1996. The linguistic realization of information packaging. *Linguistics* 34: 459-519. DOI: 10.1515/ling.1996.34.3.459

Vanrell, Maria del Mar, Stella, Antonio, Gili-Fivela, Barbara, & Prieto, Pilar. 2013. Prosodic manifestations of the Effort Code in Catalan, Italian and Spanish contrastive focus. *Journal of the International Phonetic Association* 43: 195-220. DOI: 10.1017/S0025100313000066

Terán, Virginia, & Ortega-Llebaria, Marta. 2017. A description of Tucumán Spanish intonation in Argentina. *Open Linguistics* 3, 456-490. DOI: 0.1515/opli-2017-0023

Truckenbrodt, Hubert. 1995. *Phonological phrases. Their relation to syntax, focus, and prominence.* PhD thesis, Cambridge, MA: MIT. http://www.ai.mit.edu/projects/dm/theses/truckenbrodt95.pdf (accessed: July 24, 2024).

Yanushevskaya, Irena, Ní Chasaide, Ailbhe, & Gobl, Christer. 2016. The interaction of long-term voice quality with the realisation of focus. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (eds), *Proceedings of Speech Prosody 2016*, 390-394. Boston. DOI: 10.21437/SpeechProsody.2016-191

Zerbian, Sabine. 2006. *Expression of information structure in the Bantu language Northern Sotho*. Berlin: Zentrum für Allgemeine Sprachwissenschaft, Sprach-typologie und Universalienforschung. DOI: 10.21248/zaspil.45.2006.331

Zimmermann, Malte, & Onea, Edgar. 2011. Focus marking and focus interpretation. *Lingua* 121: 1651-1670. DOI: 10.1016/j.lingua.2011.06.002

Zubizarreta, María Luisa. 1998. *Prosody, focus, and word order.* Cambridge, MA: MIT Press.

Zubizarreta, María Luisa. 1999. Las funciones informativas. Tema y foco. In I. Bosque, & V. Demonte (eds). *Gramática descriptiva de la lengua española. Tomo 3*, 4215-4244. Madrid: Espasa.

Zubizarreta, María Luisa. 2016. Nuclear stress and information structure. In C. Féry, & S. Ishihara (eds), *The Oxford handbook of information structure*, 165-184. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199642670.013.008.