Production, perception and online processing of prominence in the post-focal domain

Inaugural dissertation
to complete the doctorate from the Faculty of Arts and Humanities
of the University of Cologne
in the subject Linguistics

presented by

Caterina Ventura

born on the 16th of October 1992

in San Daniele del Friuli (Udine), Italy

Cologne, September 2020

This dissertation was accepted by the Faulty of Arts and Humanities of the University of Cologne in 2020. It was supported by the CRC “Prominence in Language” (SFB 1252 – Project A01), founded by the Deutsche Forschungsgemeinschaft.

First supervisor: Prof. Dr. Petra B. Schumacher

Second supervisor: Prof. Dr. Martine Grice

Third supervisor: apl. Prof. Dr. Stefan Baumann

Chair of the doctoral committee: apl. Prof. Dr. Doris Mücke

Date of the defense of the disseration (Defensio): 3 December 2020

Acknowledgements

The two people I would like to thank the most are my two supervisors, Petra B. Schumacher and Martine Grice. Their constant support, their advice and the time they have invested in me have advanced my development immensely. They have both contributed to my research significantly and inspired me deeply. Petra has always been able to encourage me and to point me in the right direction whenever I was overwhelmed by doubts. Martine has always provided me with new stimuli, new questions and new perspectives on how to deal with problems. I would have been lost without their guidance. Another important guide to me has been Stefan Baumann, whose profoundly intelligent comments and whose attention to details have grately helped improving my manuscripts.

I would like to express the deepest gratitude to Michelina Savino, who has guided me throughout my years of research, not only by constantly contributing to deepen my understanding of Italian intonation, but also by advising me wisely about choices in life. We share not only a mutual interest for the field of study, but also the same view on many other non-related topics. I am also very grateful to her for being a generous host, welcoming me with open arms to her home in Bari, guiding me in exploring the city and preparing for me the most delicious meals. She also introduced me to Mario Refice, who I am very glad to have had the chance to meet and whose profoundly insightful questions helped me in the understanding of my topic and in improving my confidence. His help was fundamental in organising the installation of the temporary EEG lab in Bari.

I am also particularly tankful to Brita Rietdorf, not only for the great help in carrying out the EEG experiment, but also for all the laughs and the good time we had together in Bari. I thank all the members of the department of Education, Psychology and Communication at the University of Bari. In particular, I would like to thank Davide Rivolta for promoting the participation in my experiment among his students. My gratitude also goes to Annalisa Palmisano and Mariagrazia Violante, who helped me gather participants and setup experimental sessions and who were always able to find new and interesting topics to discuss during the long hours together. I also thank all the consultants that participated in the tiresome experiments, for their commitment and their feedback.

I am profoundly grateful to all the colleagues who have helped me during these years. First of all, I would like to express my gratitude to Aviad Albert, whose genius ideas, guidance and strong passion have always inspired me and dramatically improved my work. In addition, I would like to express my gratitude to Mark Philipp for taking the time to teach me everything about an EEG lab, Ingmar Brilmayer for being a constant support with data analysis and Mark Ellison for all the inspiring discussions about my work. I would also like to thank Diana Dimitrova, for her work and her ideas have been crucial for my dissertation. I am also grateful to Christine Röhr, for helping me navigate through life in the lab and for kindly supporting me in times of need. Finally, I thank Anne Gerlach, who not only has assisted me in my work but has also become a true friend.

I am very thankful to many other people in the *Institute of Phonetics*. They have all taught me a lot and influenced my work in some way: Alicia Janz, Doris Mücke, Francesco Cangemi, Hariett Hanekamp, Simon Roessig, Simona Sbranna, Theo Klinker, I have learned many important lessons from all of you and I am grateful for all the support you have given me. The people of the *Department of German Language and Literature I* also deserve my gratitude: Florian Bogner, Melanie Fuchs, Clare Patterson, Barbara Tomaszewicz-Özakin, Paul Compensis, Christina Rath, Lena Straßburger, Maria Spychalska, Celine Cuma, Magdalena Repp, Filiz Oezden, your constant support and feedback have helped me enormously.

Beside the people that helped me in my academic life, my immense gratitude goes to those who supported me in my personal life. I would like to thank my mother and my father for always being there when I needed them and for supporting me during all the years that brought me to this achievement. I also thank my two "little" sisters who have always helped me face myself and my fears. I am fortunate enough to have incredible friends in my life, I would not be the person that I am today without them and I could not imagine the last years without them. Among those my deepest gratitude goes to Luca, without whom I could not have made it and who always inspieres me to have a deeper understaniding of the world and helps me redefining my views of life. Many thanks go to Marco, Elisa, Andrea, Maria Rosaria, Marduk, Jens, Julian, Giulia, Caterina and Claudia, the best people I could have hoped to have beside me.

Contents

List of Figures

List of Tables

General introduction

In a communicative situation, interlocutors usually ‘package’ information into more informative and less informative parts, i.e., they informationally structure the discourse. In many languages, prosody contributes to the packaging of information by highlighting new or unpredictable information and attenuating shared or expected information. Therefore, prosodic features convey cues about the communicative intentions of the speakers and influence the attentional states (i.e., the focus of attention) of the listeners.

Prosodic highlighting can be attained through accentuation, which is reflected in different acoustic cues, such as modulation of F0 (increase in its dynamic movement), local increase of duration, of spectral properties and of overall energy. The additional effort afforded to produce the enhancement of phonetic features, makes the element which features these characteristics *prosodically prominent*. Generally, the more effort is afforded, the more an element is prosodically prominent. Prosodic prominence refers to the property of an element to *stand out* from its *environment* by virtue of its acoustic characteristics. In this definition, the environment corresponds to a set of elements of equal type and rank, among which one is singled out (e.g., syllables; cf. von Heusinger & Schumacher, 2019; see also Himmelmann & Primus, 2015).

Interestingly, there is an inverse relationship between prosodic prominence and information status (related to the degree to which information is already given in the discourse), in that, all things being equal, prominent information in discourse (i.e. given information, defined as being already active in the discourse) is prosodically attenuated, while information that has not recently been mentioned, being thus less prominent in discourse (i.e. new information, defined as inactive in the discourse) is compatible with prosodic highlighting (see Baumann & Riester, 2012). In addition, different patterns of highlighting and attenuation usually distinguish different focal structures: to signal a word as the focus (i.e., most informative part) of the utterance, the subsequent part (post-focal) normally needs to be realised as attenuated (see e.g., Cruttenden, 2006 on deaccentuation and reaccentuation; Terken & Hirschberg, 1994 on deaccentuation, Xu, 2011 on post-focal compression, among others).

However, the association between prosodic form as described by discrete intonation categories, and the information structure is not one to one but can be better described by a probabilistic relation (Avesani & Vayra, 2005; Roettger, Mahrt & Cole, 2019) and by tendencies in the distribution of continuous phonetic parameters (Cangemi & Grice, 2016; Grice, Ritter, Niemann & Roettger, 2017). Moreover, across languages the distribution of highlighted and attenuated elements can be different within utterances, as can the probabilistic association between information structure, i.e., the formal packaging of information within a sentence, or information status, i.e., the degree in which information is already given in the context, and prosodic marking.

For example, Italian reportedly shows a probabilistically looser relation between degrees of givenness and prosodic marking in comparison to the relation shown by West-Germanic languages. Indeed, studies on Italian have reported lower percentages of co-occurrence of given material with the absence of accents (Avesani, 1997; Avesani & Vayra, 2005), which led authors to conclude that speakers of Italian may not prosodically attenuate entities that are repeated in the same discourse segment (i.e., given entities). In addition, Avesani, Bocci, Vayra and Zappoli (2015) and Swerts, Krahmer & Avesani (2002) report the presence of accents in the given post-focal elements of noun phrases (NPs): the NP *triangolo rosso* (lit. triangle red) is reported to be realised in Italian always with an accent occurring on the adjective (*rosso*), irrespective of the information status of the noun and the adjective (Swerts et al., 2002). These post-focal accents are not found in the productions of native speakers of German and Dutch analysed in these two experiments, and generally are not attested for West-Germanic languages.

In sentence-length utterances the above-described pattern observed in NPs is not attested: the post-focal material is realised with a flat and low F0. Nonetheless, this pattern has been reported to be different in sentence-length interrogatives in some varieties of Italian (e.g., those spoken in Bari, Naples, Palermo). In these varieties given, post-focal items are realised with an accent. This accent has, however, reduced cues to prominence and serves to signal sentence modality (the speech act of request; Cangemi & D’Imperio, 2013; Grice, 1995; Grice, D’Imperio, Savino & Avesani, 2005, among others). An example of a stylised contour of an interrogative utterance with the subject (*Marco*) in narrow focus (i.e. the following part of the utterance in post-focal position) is provided in (1). In the example a compressed rise-fall accent is displayed after the focal accent occurring on the subject (post-focal accent on the word *Colonia*, Cologne).

(1) MARCO viene a Colonia?
Is MARCO coming to Cologne?

Accents that feature F0 movement are attested in the post-focal position only for questions in varieties that have a rising-falling intonation pattern in interrogatives. Nonetheless, Bocci and Avesani (2011) argue that in the central variety of Italian spoken in Tuscany, for statements, where the F0 in the post-focal position is flat and low, there is still the possibility of finding accents in the post-focal region, accents which do not feature a wide-ranging movement in the F0 but are characterised by spectral emphasis and increased duration.

The prosodic marking of elements has been shown to play a role in the perceptual domain, with the acoustic features related to prominence being strongly perceived by listeners and identified as marking highlighted/important elements in the utterances. However, perceptual prominence derives not only from the signal, but is also expectation-driven (Bishop, 2012; Cole, Mo & Baek, 2010; Cole, Mo & Hasegawa-Johnson, 2010). Recently it has been pointed out that the distribution of accents and the mapping between accents and information structure/information status can influence the perception of prominence. The hypothesis is that expectations are derived from the probability of acoustic correlates of prominence to appear in a specific context (see Roettger et al., 2019). In addition, given that the attentional resources of the brain are limited, when listening to speech in a natural context attention is not equally allocated to all parts of the message, but it is attracted towards the prosodically highlighted (i.e., prosodically prominent) ones, and it is shifted away from the attenuated parts (Li & Ren, 2012).

Despite the amount of literature on prominence perception, there is still no clear picture of how expectation-driven and signal-driven inferences interact. Moreover, how and whether listeners' language-specific expectations in the perception of phrasal prosodic prominence are transferred from the native language (L1) to the L2, has not yet been explored, and a clear account is still unavailable. In addition, questions still arise both on the interplay of acoustic features in signalling more fine-grained cues to prominence, such as the ones present in compressed accents or in accents that do not feature F0 movement (Bocci & Avesani, 2011), and on their role in the perceptual domain.

The overarching goal of the present thesis is to gain a broader understanding of prosody as a prominence-lending and attention-orienting device. In order to reach this goal, this research sets out to further investigate both the production of different degrees of prominence, connected to a specific position in the utterance and in the prosodic structure, and the influence of bottom-up (signal-driven) and top-down (expectation-driven) inferences in the perception of prominence.

In order to answer these questions, the present thesis comprises three experiments which address more specific research questions:

(1) How are the multiple and continuous acoustic cues to prominence distributed in utterances with different information structures, and how do they interact? Can acoustic cues to prominence other than F0 movement contribute to signal prominence in the post-focal region?

(2) Does the distribution of prominence-lending features and their probabilistic mapping with the information structure and information status in a language affect the perception of prominence in the post-focal domain?

(3) Does the presence of post-focal cues to prominence reorient attention to the post-focal region?

The investigation of these questions will be carried out using the particular case of Italian and will concentrate on the post-focal domain in this language. Indeed, varieties of Italian constitute an interesting case to further deepen the understanding (i) of how and whether post-focal phrasal prominence can be expressed without F0 movement, (ii) of the interplay of expectation and stimulus-driven inferences, (iii) of the role of the probabilistic distribution of prominence on its perception, and (iv) of the influence of fine-grained cues in perception. The reasons why Italian is suited to address these questions consist in the presence of post-focal pitch accents in some varieties, in the reported presence of prominence without F0 movement, and in the reportedly weaker correlation between information status and prosodic form. Moreover, another characteristic of Italian which is interesting in this respect is its reportedly higher tendency, in comparison with West-Germanic languages (in particular with English), to manipulate word order to place new (prominent) information at the end of the phrase. An example is the possibility in Italian to express the sentence "Giovanni is coming" either as *Giovanni viene*, or *viene Giovanni*, changing word order to signal that the last element in the utterance is the new one, while English employs the same word order for both conditions. In addition, research on Italian is interesting in itself because there is to date only scant research concerning the perception of prosodic prominence in this language.

The present thesis deals with one northern variety of Italian, spoken in Udine, and one southern, spoken in Bari. These two varieties differ in the distribution of features related to prominence in the post-focal position, the former not presenting modulation of the F0, the latter presenting it in one sentence modality (interrogative). The question addresses whether the greater paradigmatic choice in the degree of prominence shown by Bari Italian in the post-focal position creates differences in the prominence perception between the two varieties. A further question is how prominence in stimuli from both varieties is perceived by German learners of Italian. The intent of this latter investigation is to explore how the top-down inferences stemming from the L1 and the reduced knowledge of the phonological features of Italian (especially regional varieties) available to learners, would influence the perception of prominence, in particular in the post-focal domain. The variety spoken in Bari is chosen to further explore whether the presence of more fine-grained cues to prominence in post-focal position (such as compressed accents showing F0 movement) can affect the orienting of attention and consequently the depth of semantic processing.

The present work comprises the following experiments: (i) a production study (Chapter 5), which serves the twofold purpose of investigating the prominence marking of words in different focal structures and of establishing the basis for the following perception experiment (Chapter 6); (ii) the aforementioned perception experiment (Chapter 6), consisting of a prominence rating task; (iii) an online perception experiment (Chapter 7). More precisely, the experiment in Chapter 5 collects production data on the variety of Italian spoken in Udine featuring sentences realised with three different focal structures: (i) broad focus, (ii) narrow contrastive focus on the object, and (iii) the object occurring in post-focal position (i.e., after the focus), with narrow contrastive focus on the preceding verb. Different prominence relations are predicted between the verb and the object within utterances and between the object occurring in the various focal structures. These differences are predicted to be realised through the use of more or less pronounced prominence-lending parameters. The presence/absence of F0 dynamics, enhanced energy and enhanced duration are the parameters taken to indicate the prominence degree of words.

The design of the production experiment in Chapter 5 is further used to collect a smaller sample of data from the Bari variety, which is presented in Chapter 6, and which serves as perceptual stimuli for the main part of this experiment. The perception experiment in Chapter 6 aims to assess whether the prominence relations found in the production experiment on the variety of Udine and in the smaller sample of stimuli analysed for the variety of Bari, can be identified by listeners. Native speakers of the two varieties of Italian (Bari and Udine) and native speakers of German (learners of Italian) are asked to rate the level of prominence of words with different levels of prominence occurring in utterances realised in the production experiments. This latter experiment is in particular concerned with how the post-focal domain is perceived by both native speakers of the two different varieties, which present a different probabilistic association between post-focal position and presence of cues to prominence. The focus is on the role of top-down inferences on prominence perception. In addition, the experiment investigates how the prominence patterns found in these two varieties of Italian are perceived by learners, thus further contributing to the understanding of the role of bottom-up and top-down inferences. The different distribution of accents typical of the Bari variety are expected to have an impact on the prominence perceived by the group of native speakers of this variety: the higher probability of finding prominence in the post-focal position should increase the expectations of finding prominence in this position. This is predicted to result in the perception of a higher degree of prominence for the group of speakers from Bari compared to the group of native speakers of the Udine variety and to the group of learners.

The experiment in Chapter 7 is an event-related potential (ERP) experiment, which aims at defining the contribution of fine-grained cues to prominence in the (re-)orienting of attention from a very prominent stimulus to a less prominent one. In order to do so, this experiment compares the processing of statements and questions realised in the variety of Italian spoken in Bari, in which the post-focal region of statements is realised with a flat and low F0, whereas the post-focal region of questions presents a compressed accent with a rising-falling F0. The higher degree of F0 movement in the post-focal position of questions is expected to attract a higher allocation of attentional resources compared to the same position in statements.

Outline

This thesis is divided into three parts: a first part consisting of three chapters, providing the theoretical background, a second part, comprising the experimental chapters, and a third part containing a summary of the experimental results, as well as a general discussion and a general conclusion.

The theoretical background (Chapters 1, 2 and 3), discusses the most relevant aspects of prosody and information structure, and the relation between them. Moreover, it addresses the influence of prominence on the perceptual domain and on processing. Chapter 1 focuses on delineating the concept of prosody, concentrating in particular on one of its main functions, highlighting. The first section (1.2) aims to provide a picture of the basic functions and features of prosody in intonation languages as well as of the Tones and Breaks Indices (ToBI) framework, the most widespread framework used for describing and categorising intonation events, such as pitch accents and boundary tones. Further, the use of continuous parameters to implement the categorical analysis is discussed. In particular, a recently flourishing approach to model intonation, comprising the use of periodic energy, is described. The second section of the chapter (1.3) deepens the concept of prosodic prominence, with a focus on its acoustic correlates and on prominence relations within utterances. Chapter 2 deals with the notions of information structure and information status and their interrelation with prosody. The first part (2.2) is devoted to delineate the basic concepts of information status and the more comprehensive notion of information structure. The second section (2.3) involves the prosodic marking of information structure and information status, providing a comparison between German and Italian, particularly focusing on the prosodic realisation of given post-focal elements. Chapter 3 presents the current understanding of prominence as an attention orienting device and of the role of acoustic features and their contribution in prominence perception. The first section (3.2) is concerned with the signal-based perception of prominence. The second section (3.3) explains the probabilistic mapping between prosody and information structure and information status, and its consequences on processing. The third section (3.4) deals with the expectation-driven inferences in the perception and in the processing of prominence. The fourth section (3.5) provides a fundamental picture on the effects of prominence in the online processing. These effects are further deepened in the following section (3.6), which presents an overview on the interplay between attention and prosodic prominence. Finally, the last section (3.7) presents the current understanding of the influence of language-specific differences on prominence perception and reports on the current understanding of prominence perception in Italian, a topic that has received little attention to date. Chapter 4 provides a summary of the theoretical background and an introduction to the experiments reported in the following chapters.

The experimental part consists of three chapters, which present the abovementioned experiments: (i) the production experiment involving the variety of Italian spoken in Udine (Chapter 5); (ii) the production experiment on a smaller sample of speakers of the variety of Italian spoken in Bari (Chapter 6); (iii) the prominence rating task experiment (Chapter 6), conducted on the stimuli collected from the aforementioned two varieties of Italian, and involving three groups of listeners with different native speaker backgrounds (native speakers of two different varieties of Italian and native speakers of German); (iv) the ERP experiment (Chapter 7), involving the variety of Italian spoken in Bari and its native speakers.

To conclude, Chapter 8 presents a summary and a general discussion and conclusion, referring to the research questions previously presented.

Chapter 1

Prosody and prosodic prominence

1.1 Introduction

The aim of this chapter is to outline the concepts concerning prosody and its highlighting function and, in particular, to provide a motivation for the measures used to analyse the acoustic characteristics of the recordings acquired in the production experiments (Chapter 5 and part of Chapter 6) and of the acoustic stimuli used in the perceptual experiments (part of Chapter 6 and Chapter 7). Although this thesis is primarily concerned with two varieties of Italian, the one spoken in Bari and the one spoken in Udine, references to research on the prosodic system of Italian in general will be provided throughout this chapter.

This chapter consists of sections and subsections describing (i) prosody, (ii) the categories used to describe intonation contours, (iii) the continuous phonetic characteristics that determine the definition of the categories, (iv) the interplay between categorical and continuous aspects, and (v) prominence relations within prosodic structure. Firstly, section 1.2 presents the Autosegmental-Metrical (henceforth, AM) theory and the transcription system within this framework, Tone and Breaks Indices (henceforth, ToBI), which is currently the most widespread system for analysing intonation contours. In the AM tradition, F0 contours are analysed as sequences of tones, which can be either culminative (serving the highlighting function of intonation) or delimitative (serving the phrasing function of intonation). The present chapter is particularly concerned with the former class of tones, referred to as pitch accents. This class is composed of different *pitch accent types*, which are cued by a bundle of phonetic properties, the nature of which contribute to the definition of pitch accent categories. Different pitch accent types are considered to serve different communicative functions. For different languages a different inventory of pitch accent types serving different functions can be established. Since the present thesis is concerned with two varieties of Italian, the inventory of pitch accents common to varieties of this language is described in subsection 1.2.3. Secondly, section 1.2 presents the issue concerning the relation between the categorical classification of the ToBI system and the continuous phonetic properties characterising the categories identified in the system. Generally, the issue in this relation has been underlined by a few studies emphasising the need of a broader consideration of continuous aspects not only to implement and motivate the distinction and the boundaries between categories, but also to allow a deeper understanding of the way in which continuous multidimensional cues interact to convey a certain pragmatic function in a given context (Cangemi, El Zarka, Wehrle, Baumann & Grice, 2016; Cangemi & Girce, 2016; Cangemi, Krüger & Grice, 2015; Cole & Shattuck-Hufnagel, 2016; Grice, Ritter, Niemann & Roettger, 2017). Along these lines, a recent proposal to describe F0 contours with continuous parameters connected to periodic energy is then described (Albert, Cangemi & Grice, 2018; Cangemi, Albert & Grice, 2019). Periodic energy is defined as an intensity measure which only reflects the periodic components of the signal (characteristic of vowels and of syllable nuclei) and is related to perceived pitch. Parameters connected to this measure will be used in the experimental chapter in order to provide continuous measures to describe F0 contours.

Section 1.3 delineates the concept of prosodic prominence and helps in fostering its understanding. This section starts from a preliminary and general definition of prosodic prominence, which implies the occurrence of an element which has structural and acoustic characteristics that make it stand out compared to neighbouring elements. The section continues in deepening the understanding of the different degrees of prominence and the interplay of the bundle of acoustic features characterising it. Subsection 1.3.3 is concerned with the acoustic measure of *Periodic Energy Mass* (PEM), which relates to periodic energy and which is an effective measure to estimate the degree of prominence of an element. Calculated as the area under the periodic energy curve, PEM indeed permits an account for both the parameters of periodic energy and duration. Finally, section 1.3 further addresses the conceptualisation of prominence as either binary (elements categorised as prominent or non-prominent) or gradient (elements presenting gradual changes in prominence), relating to studies that underline the need of categorical and continuous aspects to be considered as part of a single system (Grice et al., 2017; Roessig & Mücke, 2019; Roessig, Mücke & Grice, 2019).

1.2 Prosody

1.2.1 Fundamental features and functions of prosody

Spoken language is not possible without prosody. Thus, it is present in every utterance in every language. Prosodic phenomena vary in response to the discourse context and provide communicatively relevant meaning to the utterance. This meaning can be both linguistic and paralinguistic (e.g., Grice & Baumann, 2007). Functions that are linguistic comprise the expression of speech acts, sentence modality (e.g., statements and questions), disambiguation between syntactic structures, regulation of turn taking, implicatures and information structure. By contrast, paralinguistic functions involve emotional states (e.g., sadness, excitement, uncertainty) or attitude (e.g., hostility, friendliness). In addition, prosody also encodes extralinguistic aspects of the speaker such as gender, age, dialect, profession and many others. The distribution and interaction of the prosodic attributes in the flow of speech is language- and speaker-specific.

The prosodic system of a language involves the division of speech into smaller units (phrasing, including silences), speech melody (F0 contour), highlighting at word level (stress), highlighting at utterance level (accent), variation in speech rate and rhythm in the marking of prominence relations (e.g., Crystal, 1969, 1970; Grice & Baumann, 2007; Ladd, 2008). Central in the prosody of a language is the component called *intonation.* This term has been defined in a narrow sense as the modulation of the F0 over the domain of the utterance (cf. Röhr, 2016; see also Nolan, 2006) and as the association of the tune (tones in the Autosegmental-Metrical model, see below) with the prosodic structure. However, this narrow definition of intonation is not sufficient in explaining its functions. In its broader sense, intonation is treated as a complex interaction of different phonetic features that particularly concern pitch patterns in relation to timing and loudness (cf. Crystal, 1969:78). In fact, the interaction between intonation and stress, where stress is the component to which the rhythm function (timing) is devoted, is very close in many languages (Nolan, 2006). Grice and Baumann (2007; see also Grice, 2006) also indicate that, at least in intonation languages, intonation solves both the function of highlighting and phrasing (see 1.2.2). Thus, the distinction between prosody and intonation can be seen as rather artificial and the terms are often used interchangeably (Grice, 2006). The notion of intonation in a broader account can therefore be used as the equivalent of prosody.

As the linguistic structure of spoken language, prosody concerns phenomena such as: (i) phrasing, namely the division of the utterance in smaller meaningful units (chunks) which are pragmatically and semantically coherent, and (ii) highlighting at word level (stress) and at utterance level (accent), which is interconnected with the marking of prominence relations. Phonetically, prosody is expressed through the following acoustic correlates: (i) fundamental frequency, henceforth F0, (ii) intensity, (iii) spectral quality and (iv) duration of segments. These acoustic properties have their perceptual correlates, which are (i) perceived pitch, (ii) loudness, (iii) vowel quality and (iv) length, respectively. The acoustic manifestation of the F0 arises when during speech production, the pulmonic air stream can induce a quasi-periodic vibration of the vocal folds. The F0 correlates with the frequency of the oscillations, namely the pattern of opening and closing of the glottis. Variations in the F0 are due to changes in the vibration rate of the vocal folds: the higher the frequency of the vibrations, the higher is the fundamental frequency of a sound[1].

Table 1 gives an overview and a summary of the physiological, acoustic and perceptual correlates of the relevant phonetic features that are connected to prosody and intonation.

ARTICULATION	**ACOUSTICS**	**PERCEPTION**
quasi-periodic vibrations of vocal folds	fundamental frequency (F0) measure: Hertz (Hz)	pitch perceived scale: high-low
articulatory effort, subglottal air pressure	intensity measure: decibel (db)	loudness perceived scale: loud-soft
duration and phasing of speech gestures	duration of segments measure: millisecond (ms)	length perceived scale: long-short
vocal tract configuration, articulatory precision	spectral quality measure: formant values in (Hz)	vowel quality perceived scale: full-reduced

Table 1. Correlates of the phonetic parameters involved in prosody at the respective levels of description: articulatory, acoustic and perceptual. Adapted from Baumann and Grice (2007:2), see also Baumann (2006:12).

[1] A second cause of higher or lower pitch is due to the size of the larynges. In general, female larynges are smaller in the front to back dimension than the ones of males, and children have overall smaller larynges than adults (Gussenhoven, 2004). This causes the F0 range to be between 90 and 220 Hz for males, between 180 and 450 Hz for females and between 300 and 700 Hz for children (Neppert, 1998).

The way in which speech is structured can be described within the conceptual framework of AM model (Pierrehumbert 1980; see Ladd 2008 for an overview) and the transcription system within this framework, ToBI (Beckman & Hirschberg, 1994; Beckman & Ayers, 1997). The ToBI system has been developed for the analysis of (Mainstream American) English intonation, but several ToBI systems for other languages have been established (see below and see e.g., Jun, 2005 for an overview on prosodic typology and systems for the description of different languages). The present dissertation refers to the AM terminology and conceptual framework. The reasons for this choice are twofold. First, the AM framework has proven to be flexible in order not only to adapt to typologically different languages, but also to be interpreted by different researchers consistently, enabling them to compare results among various studies. These characteristics have made this transcription system the most commonly used one to describe intonation. Second, the AM model describes common properties of the intonational and the prosodic system taken in their narrow sense and helps in their unified explanation. However, the aim of the current thesis is not to give a detailed description of the languages investigated within this theoretical framework nor to discuss in depth the validity of this prosodic transcription. The AM model is here utilised in order to allow a common understanding of the phenomenon that will be encountered in the experimental part. The AM framework will thus be only briefly explained.

The aspects of relevance in the use and description of the AM model for the purposes of this thesis are the general functioning of the prosodic description of Italian (and its varieties), the discussion around the post-nuclear (or post-focal) region and its prominence status. Moreover, the understanding of the AM framework will facilitate the discussion of the existent literature on the prosodic marking of information structure (in Chapter 2) and the existent literature on the processing of prosody and prominence (in Chapter 3). Crucially, this thesis will also describe and adopt a different method of analysing prosodic data – i.e. measures related to periodic energy – and to some extent provide a brief overview of the critical issues within the AM model. The reader with interest in more detailed discussion of the AM model, its advantages and its limits, is referred to Bird & Ladd (1991) and Martin (2012).

In the AM model, prosodic structure and intonation are defined in a narrow sense (Ladd, 2008). The term prosodic structure refers to the system that groups utterances into smaller units and assigns relative prominence to elements within these units. This representation describes the metrical aspect, proposed by Liberman and Prince (1977). The term intonation indicates the system that associates tonal events with this metrical structure and represents the autosegmental aspect, primarily based on the work of Goldsmith (1976) and Leben (1971). As mentioned before (see the discussion on the notion of intonation in a broader account), this distinction between the two systems is interrelated within the AM framework (see below).

The AM model assumes that the level of description of segments is separated from the one of the tonal events. However, a more complete and accepted way of describing speech would be that elements within the segmental description are incorporated into the prosodic structure and both prosodic structure and intonation are phonetically manifested at the segmental level, in the sense that the consonants and vowels are affected by the position in the prosodic hierarchy and tones that are associated with the position (see Grice, Savino & Roettger, 2018). Therefore, AM formalisation needs to be seen in light of the results of various studies that show how prosodic context is a significant source of phonetic variation not only for the so-called suprasegmental properties of speech (e.g., pitch, loudness, spectral modulation, and duration) but also for segmental properties (e.g., vowel formant patterns and consonant voicing; see Gussenhoven & Rietveld, 1988; Sluijter & van Heuven, 1996; Turk & Sawusch, 1996; Cole, Kim, Choi, & Hasegawa-Johnson, 2007; Kim & Cho, 2013; Kochanski, Grabe, Coleman, & Rosner, 2005, and many others; see also Roettger & Grice, 2019 for an overview of the findings on the interaction between the tune and the segmental composition of words). To better explain the consequences of the prosodic structure (and of intonation) on the segmental level, the following example can be taken into consideration: one word can be made prominent (can be highlighted, see below) by a rise in pitch that associates with the stressed syllable of the word in question. This association with pitch movement is accompanied by longer, louder, and more clearly articulated segments in comparison to the same segments occurring in non-prominent positions. A similar effect can be observed at the beginning and at the end of chunks that the prosodic structure entails. Examples to clarify these concepts are provided in 1.2.2.

As described in more detail in 1.2.2, F0 movements reflecting linguistic functions are annotated in the AM model using variants of the ToBI transcription system, which are relevant for the language analysed (ToBI for Mainstream American English; GToBI for German, see Grice, Baumann & Benzmüller, 2005; ToBIt for Italian, see Avesani 1995 and Grice et al., 2005; and other acronyms for other languages; the ToBIt transcription will be adopted in the present thesis). ToBI transcription is not intended to reflect listeners' conscious knowledge, since during communication listeners do not explicitly identify types and location of prosodic boundaries and prominences (cf. Mo, 2011:7). Nonetheless, both these elements and their phonetic implementation are perceived by listeners (see 3.2) and serve communicative purposes of semantic coherence (grouping by prosodic boundaries) and of highlighting the word of the phrase that carries the important information (prominence).

1.2.2 Prosodic structure and Tones and Break Indices (ToBI) system

Prosodic structure has been studied by a large body of research which has described its hierarchical nature (see below). In the AM model utterances can be broken down in multiple levels of phrasing. Although some levels of the structure are common to a large number of languages, there is not one prosodic structure common to all languages, since the relevant phrases differ from language to language. This thesis is concerned with Italian and, to a lesser extent with German, which are intonation languages (or *stress accent languages*). Moreover, they do not have tonal contrasts that are determined in the lexicon. In these languages, changes in pitch (i.e., tonal movements) do not alter the lexical meaning of individual words, but only the meaning of sentences and phrases as a whole. Moreover, the lexical minimal pairs are distinguished by the place of stress (e.g., the contrast between PERmit, noun, and perMIT,verb or in Italian between ANcora, "anchor", and anCOra, "yet"), which is not (necessarily) associated to tones, as will be made clearer in subsection 1.3.1.

The configuration of a prosodic hierarchy is not just language-related, but also theory-related. In fact, there are several different proposals concerning the number of prosodic domains (e.g., Jun, 1998; Ladd, 2008; Liberman & Prince, 1977; Nespor & Vogel, 1986; Pierrehumbert, 1980). Not all prosodic units have been attested cross-linguistically and are dependent on the followed theoretical framework. Figure 1 shows a simplified prosodic structure, reflecting the parts that are widely accepted in the field (cf. Keating, Cho, Fougeron & Hsu, 2003). In the prosodic structure, small prosodic domains such as syllables and words are hierarchically organised into bigger entities up to the level of the utterance. The unit of the utterance (*υ*) consists of one or more Intonational Phrases (*ι* or IP), which in turn comprise one or more smaller phrases (XP in Figure 1). The XP contains one or more words (*ω*), which are organised into metrical feet (Σ), in turn organised into one or more syllables (*σ*). For the smaller phrases, which are the intermediate level between the IP and the word, there is no common structure and name; rather they are based on the language under investigation and on the theoretical preference of the investigator. Depending on the analysis, the constituents occupying the level of the XP have different names and a different structure: phonological phrase (*φ* or PP; e.g. Gussenhoven 2004, for English), intermediate phrase (ip, e.g., Beckman & Pierrehumbert, 1986; Pierrehumbert & Beckman, 1988; for English; Grice, Baumann & Benzmüllller, 2005, for German) or accentual phrase (the latter attested only in some languages, e.g., see Jun & Fougeron, 2002 for French; Jun, 1998 for Korean; Chong & German, 2017 for Singapore English).

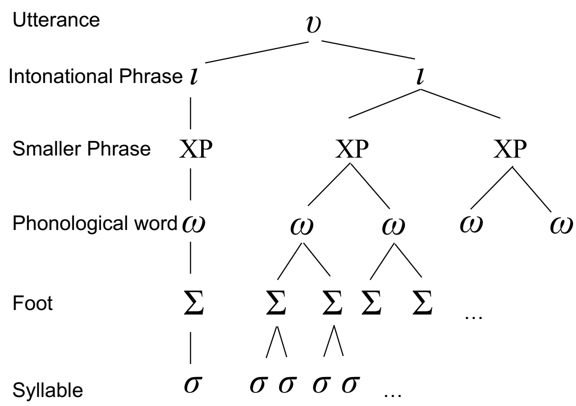


Figure 1. Adapted from Keating et al., (2003).

To reflect the prosodic structure, speakers shape speech through the modulation of multiple acoustic parameters in terms of pitch, timing (the relative length of units of speech sounds), loudness, strength of articulation and silence. Therefore, prosody is a major source of phonetic variation: elements at the edges of prosodic units and elements in strong positions, to which prominence is assigned, are phonetically distinct from the same elements occurring in other prosodic contexts (see 1.3). Consider, for example, the following sentences:

(1) a. Mary LOVES chocolate.

b. Mary loves chocolate.

(2) a. Mary loves chocolate.

b. Mary loves chocolate sweets.

The word *loves* in (1a.) can be made prominent by a rise in pitch (the prominence is here represented orthographically by capital letters and by the rising movement of the stylised pitch contour; see 1.3 for an in-depth explanation of the concept of prominence), which co-occurs with longer, louder, and more clearly articulated segments in comparison to the ones of the same word in (1b.), which is not made especially prominent compared to the preceding and following words. Similarly, the finality of the word chocolate in (2a.) in comparison to (2b.) where it is not final, can be signalled by a fall in pitch (represented by the stylised pitch contour, a decrease in loudness and a lengthening of the final segments. This lengthening is caused by a slowing down of the articulators at the edges of domains, reflected in the signal by final and domain-initial lengthening. The higher the domain in the prosodic structure the greater is the degree of lengthening (e.g., Wightman, Shattuck-Hufnagel, Ostendorf & Price, 1992; Keating, Cho, Fougeron & Hsu, 2003). For instance, in English a /t/ sound at the beginning of a given domain is realised with greater aspiration and greater contact area between tongue and palate (i.e., it is strengthened), compared with at the beginning of a domain at a lower level of the prosodic hierarchy (cf. Grice 2006; see also Keating, Cho, Fougeron & Hsu, 2003). An example is given in (3-5), where the different manifestations of /t/ occur in different prosodic positions (the ones that are of interest are underlined in the examples): IP initial (3), word initial but IP internal (4) and word internal (second t of Tomtom). The /t/ in (3) would usually be realised as stronger than the one in (4), which in turn would be stronger than the one in (5).

(3) <u>T</u>om is here.

(4) I like <u>T</u>om better than John.

(5) Can I borrow your Tom<u>t</u>om navigator system?

Moreover, at lower levels of the prosodic hierarchy the processes of assimilation, dependent on connected speech, are more frequent than at higher levels. Furthermore, a /t/ sound at the beginning of a stressed syllable is pronounced with greater strength compared to the one at the beginning of an unstressed syllable. In addition, if the stressed syllable bears a pitch accent, the /t/ sound at the beginning of this syllable is stronger than in a syllable not bearing the accent (cf. Grice, 2006:780). This relation is exemplified in (6) and (7), where the initial /t/ of Tom in the answer in (6) is stronger than the initial /t/ of Tom in the answer in (7). Capital letters indicate words where the pitch accent occurs. Examples (6) and (7) are adapted from Grice (2006).

(6) Q: Do you love Tobias?
A: I love TOM.
(7) Q: Do you like Tom?
A: I LOVE Tom.

A depiction of the prosodic structure in the form of a tree as intended by Gussenhoven (2002) is provided in Figure 2. Referring to this figure allows a better comprehension of the prosodic structure and the relation with the significant aspects of the F0 annotation. In the figure, labels provided to indicate intonational events belong to the ToBI system. Tonal events include both the movements of pitch that co-occur with strong elements of the prosodic constituents and the ones that co-occur with edges. The former are called pitch accents and the latter edge tones.

The Intonational Phrase (henceforth IP; note the capital letters) and the intermediate phrase[2] (henceforth ip; note the lowercase letters) are marked and defined by the distributions of tones: they have tones occurring at one or both of their edges (cf. Grice, 2006). The tones marking the IP are symbolised with a percent (%) sign after the tone, whereas the tones marking the ip are symbolised with a minus (-) sign after the tone. The Intonational Phrase (henceforth IP) is defined by the distribution of H% or L% *boundary tones* at its edges, where the final boundary tone is seen as obligatory (Beckman & Pierrehumbert, 1986). The intermediate phrase also has an obligatory final edge tone (H- or L-) called *phrase accent*. This leads to the sequence of two edge tones at the end of an Intonational Phrase. For instance, in Figure 2, there is a high edge tone at the right edge

[2] Here I choose to refer to the intermediate phrase rather than to the phonological phrase purely for convenience reasons.

of the initial ip (H-) and a low edge tone at the right edge of the final ip (L-), which is followed by the right edge of the final IP (L%). These tones signal the boundaries of the constituents. Pitch accents occurring with the words /tuː/, /kʊks/, /spɔːɪl/ and /brɒθ/, signal the communicatively most important positions (see 2.3). The tonal targets (both pitch accents and edge tones) are formally represented either by a single tone (monotonal target) or by a combination of tones (bitonal target), expressed in terms of H(igh) or L(ow). As far as pitch accents are concerned, they consist of minimally one tone, which is signalled by a "star" (e.g., H*, L*). Optionally the starred tone can be joined with an unstarred tone and become bitonal (signalled by +). The starred tone is considered the one with the main association to the lexically stressed syllable of the word that bears the accent (see Figure 3).

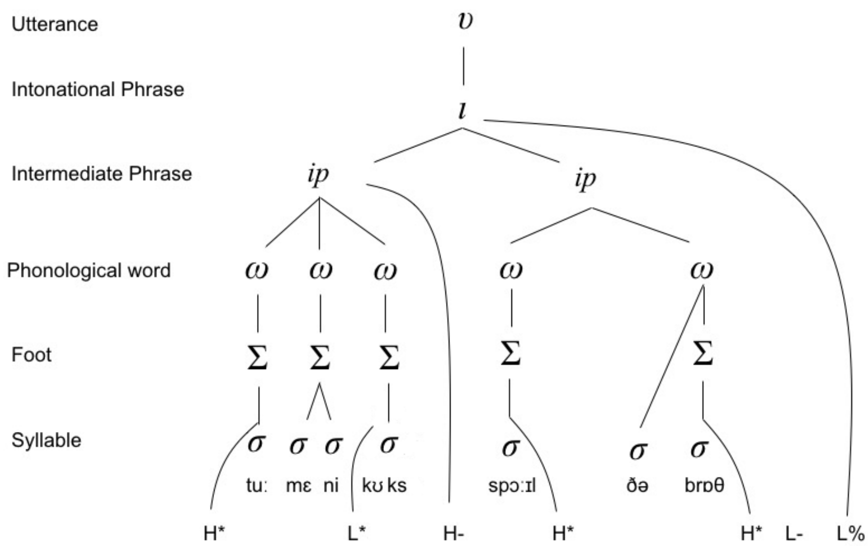


Figure 2. Example of a prosodic structure for the English sentence "Too many cooks spoil the broth", adapted from Gussenhoven (2002:271) cited in Grice (2006). At the bottom: tonal structure as proposed by Beckman & Pierrehumbert (1986).

In the case of bitonal accents, a widely accepted procedure is to relate to the phonetic alignment with the segmental material. Figure 3 shows two different pitch accent types for English. The pitch of both types is rising close to the stressed syllable (grey box), but the difference between the two is in their phonetic alignment. The schematised contour of the L*+H pitch accent has a rise starting within the stressed syllable and reaching its maximum in the following syllable. The schematised contour for the L+H* accent starts its rise in the syllable preceding the stressed syllable and reaches the pitch maximum in the following syllable (the stressed syllable). This difference in the alignment is explicated by means of the starred tone, indicating which of the two tonal targets is reached within the stressed syllable[3].

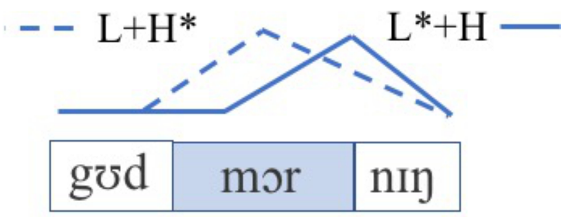


Figure 3. Schematic pitch contours of bitonal pitch accents. Adapted from Grice (2006:278).

The abstract phonological notation provides useful means to refer to intonation contours that are different in respect to the phonetic realisation but functionally equivalent. The notation of H and L does not reflect absolute values of F0: high tones are situated roughly in the top three quarters of the speaker's pitch range and low tones are situated in the bottom quarter (cf. Grice, Baumann & Benzmüller, 2005:13). Therefore, the concepts of *dowstep* and *upstep* have been applied to describe phonologically relevant differences in *global* and *local scaling* of intonation contours. Global scaling involves the raising or the lowering of the contour across a phrase, while local scaling involves the raising or lowering from one tone to the next. The term downstep entails that a high tone is lowered in relation to preceding high tones. The use of the term upstep indicates that a high tone is higher than preceding high tones.

1.2.3 (Varieties of) Italian Tones and Break Indices

The ToBI system of transcription developed within the AM framework (see Beckman & Ayers, 1997; Beckman & Hirschberg, 1994) for (Mainstream American) English has become a standard for developing a transcription system for many other languages. In this thesis, the system developed for Italian will be referred to as a tool for describing the corpus of the two varieties that will be collected in the experiments in Chapter 5 and Chapter 6. This transcription will be adopted in order to refer to the existing literature on Italian intonation, literature that for the vast majority employs this system. The present chapter is not going to extensively deal with all the variations and the different uses in the regional varieties of the different types of pitch accents and boundary tones. The interested reader is referred to Gili-Fivela et al. (2015).

[3] Note that this notation of starred tones is to some extent controversial (see Arvaniti, Ladd & Mennen, 2000 for a discussion).

From 1990 a ToBI system for Italian (ToBIt) and its varieties started to be developed (e.g., Avesani, 1990, 1995; Caputo & D'imperio, 1995; Grice, 1995 and many others; see Gili-Fivela et al., 2015 for an overview) and continued to be implemented in the following years (see e.g., Gili-Fivela, 2002; Gili-Fivela et al., 2015; Grice et al., 2005; Savino, 2013 and many others). The difficulty with the description of the intonational phonology of Italian is the fact that the Italian language has reached unity only in its written usage. In spoken language, Standard Italian is a theoretical reference and not an actual language spoken in Italy (Berruto, 2007). In addition, Standard Italian is not even a model to which existing varieties seek to conform (Lepschy & Lepschy, 1998:62). Indeed, spoken Italian, with the exception of professionally trained speakers coincides with one of its regional varieties (cf. Cangemi, 2014:19).

Within the AM analyses, several studies have formulated labelling proposals in order to identify commonalities between varieties, differentiating only the relevant divergences, namely avoiding the use of different labels for patterns that are the same, but still considering the actual data and not performing an analysis motivated only by cross-variety comparison (cf. Gili-Fivela et al., 2015:144-145). The most relevant works done in this direction are by Grice et al. (2005), D'Imperio (2002) and Savino (2012, 2013), all pointing out key aspects about Italian varieties. For example, they report the existence of finally falling contours for yes/no-questions in some varieties of Italian (spoken in Naples, Palermo and Bari) and the presence of post-focal pitch accents in these varieties. More recently an extensive attempt to characterise the highest possible number of varieties has been made by Gili-Fivela et al. (2015), who have analysed thirteen varieties (spoken in Milan, Turin, Florence, Siena, Pisa, Lucca, Rome, Pescara, Naples, Salerno, Cosenza, Bari and Lecce), offering a description of a set that is representative of the main areas of the peninsula (cf. Gili-Fivela et al., 2015:8). The authors came to a schematic representation of pitch accents that is assumed to be the set of pitch accents needed to account for the intonational description of Italian, offered in Table 2.

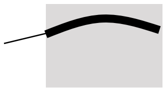
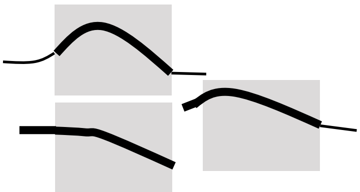
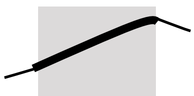
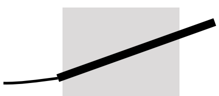
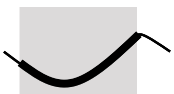
PITCH ACCENTS	
L* Low	Low tone realised at the minimum of the speaker's range. No significant difference with the F0 level of the syllable preceding the stressed one.
H* High 	High tone realised as high in the speaker's range. No significant or small difference with the pretonic F0 level.
H+L* Fall 	Falling pitch accent realised as F0 fall from a high tonal target before or at the beginning of the tonic syllable.
H*+L Rise-Fall 	Rise-fall pitch accent realised as a rise to a peak around the middle of the tonic syllable and a fall that reaches the end within the tonic syllable. In some cases, the peak might correspond to the end of a shallow rise or a plateau.
L+H* Rise 	Rising pitch accent realised as a F0 rise with the peak at the end of the tonic syllable.
L+>H* Rise late peak 	Rising pitch with a late peak realised as a F0 rise with the peak in the post-tonic syllable or later.
L*+H Fall-Rise 	Fall-rise pitch realised as a F0 rise fall within the tonic syllable and a peak in the post-tonic syllable.

Table 2. Inventory of monotonal and bitonal pitch accents proposed by Gili-Fivela et al. (2015) for the varieties investigated. Adapted from Gili-Fivela et al. (2015:148).

As in the original AM model, the starred ('*') tone of the pitch accents is placed within the accented word, at the local F0 minima or maxima associated with the lexically stressed syllable. Note that the variety of Italian spoken in Udine, that is of interest for the experiments in Chapter 5 and Chapter 6 has not been investigated in this study. By contrast the variety of Bari, which is also of interest in the present thesis (experiments in Chapter 6 and 7), is part of the Gili-Fivela et al.'s study and was considered in creating the inventory that the study proposes. However, as will be clear in the experimental part, some characteristics that are described in the Gili-Fivela et al.'s study are of use to analyse also the variety spoken in Udine. Therefore, both data of Udine and Bari will be examined referring to the inventory presented here. For the variety of Udine, additional reference will be made to the inventory provided by Roseano, Vanrell and Prieto (2015) for the Friulian language spoken in Udine (note that this thesis is concerned with the variety of Italian spoken in Udine and not with Friulian, but processes similar to the ones occurring in the dialect may occur in the variety).

Some of the pitch accents described are used only in specific positions and conditions: L* is attested in post-nuclear (or post-focal, see 2.3; e.g., evidence in Florentine and Siena varieties in Bocci & Avesani, 2011; Bocci, 2013) and in pre-nuclear position (e.g., L* in Neapolitan, see Grice et al., 2005); a downstepped compressed rising accent is attested in post-focal position for questions of some varieties, in particular the southern varieties spoken in Naples and Bari. These features will be of interest for the discussion in subsequent sections and chapters. The next section will describe in more detail the dimensions used to define a pitch accent type and the variability within types.

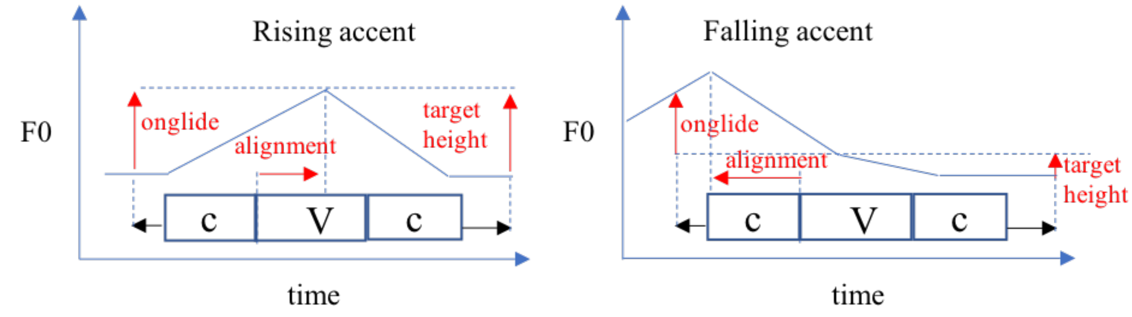
The definition of the categories in the labelling system has been conceptualised as marking communicative functions. Categories that are described as such are considered to be perceptually recognised (this point will be deepened in both Chapter 2 and Chapter 3). The discussion regarding the issues of a rigid classification of categories is unfolded throughout the following sections. The link between the perception of pitch accent categories and prominence will be of interest in the understanding of the classification of prominence relations by listeners and will be referred to in the experiment in Chapter 6.

1.2.4 Phonetic properties of pitch accents

1.2.4.1 Continuous measures

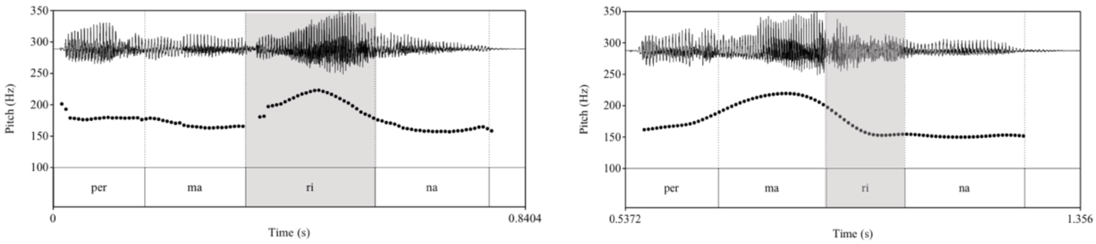
A number of continuous phonetic properties are modulated to differentiate pitch accent categories (Ladd, 2008). Therefore, as recently underlined by Grice et al. (2017), defining an inventory of pitch accents involves analysing phenomena that are continuous using discrete categories. Thus, transcribers who use the ToBI system need to draw a line between different categories and are trained to do so based on their auditory impressions (cf. Grice et al., 2017). Nonetheless, within the AM tradition the practice of using continuous parameter measures to capture the phonetic characteristics of pitch accents is common (see, e.g., Arvaniti, Ladd, & Mennen, 1998; Barnes Veilleux, Brugos & Shattuck-Hufnagel, 2012; Kügler & Gollrad, 2015; Liberman & Pierrehumbert, 1984; Peters, Hanssen, & Gussenhoven, 2015). In fact, AM has resorted to the extraction of continuous parameters in order to provide a solution for the notoriously difficult process of adequately describing and distinguishing intonational categories (cf. Cangemi, Albert & Grice, 2019). Indeed, the difficulties encountered by the labellers, especially when annotating spontaneous speech are manifold. Speech very often presents regions that have more than one possible transcription making the agreement among annotators on the adequate label difficult. In addition, when a portion of the utterance is ambiguous, labellers usually do not consider all the competing analyses and very often they report only on one. Both these operations might therefore cause loss of information and, to some extent, induce arbitrary decisions (cf. Brugos, Veilleux, Breen, & Shattuck-Hufnagel, 2008:273). Moreover, all the procedures can be disproportionately time consuming and error prone.

The procedure to extract continuous parameters regarding, for example, the alignment of a pitch accent involves the annotation of the temporal boundaries of the stressed syllables or of the stressed vowels and the subsequent calculation of the relative temporal location of the turning points of the F0 contours. The dimensions used to define a pitch accent usually involve: the above-mentioned *alignment*, which corresponds to the F0 peak of the accent as measured in the temporal relation to the onset of the stressed syllable to which it is associated (see Figure 4; see e.g, Pierrehumbert & Steele, 1989); *target height*, namely the height of the F0 corresponding to the starred tone (see Figure 4; see e.g., Ladd & Morton, 1997 for English, Kügler & Gollrad, 2015 for German), which corresponds to the maximum or minimum of the F0 in the syllable; and *tonal onglide*, which indicates whether the F0 movement is rising or falling towards the accented syllable at the location where the starred tone is placed (cf. Ritter & Grice, 2015). Figure 4 shows the parameters used to capture the phonetic details of two accent types: L+H*, which indicates a rise in the F0 contour, and H+L*, which indicates a fall in the F0 contour. Figure 5 provides F0 contours for examples of these two accent types in two Italian utterances.



a. b.

Figure 4. Parameterisation of pitch accent properties: a. rising accent, b. falling accent. Adapted from (Grice et al, 2017:7)



a. b.

Figure 5. Examples of the Italian utterance per Marina (for Marina) produced with a. rising and b. falling contour. The accented syllable is indicated in grey.

Peak alignment is considered to be a fundamental cue to indicate the type of pitch accent in several languages (see Arvaniti, 2011; D'Imperio 2011; Prieto 2011). Pierrehumbert and Steele (1989) show that in English a distinction between L+H* and L*+H could also only be a difference in the timing of the F0 peak relative to the segments, changing the position from medial to late (see Figure 3). Further, a distinction between H+L* and H* could also be made only by changing the position of the peak from early to medial (cf. Grice et al., 2017; see also Dilley & Heffner, 2013 and Kohler, 1987). Target height plays a less important role in the categorisation of the pitch accents (Ladd & Morton, 1997 for English, Kügler & Gollrad, 2015 for German), however, it is still crucial in the comparison between H* and L+H*. Moreover, it can affect the perception of alignment (cf. Grice et al., 2017). Tonal onglide is useful in that it is not limited to the F0 of the syllable which is accented, but also of the preceding syllable (cf. Ritter & Grice, 2015), therefore it considers the part of the contour corresponding to the leading tone (see 1.2.2).

Even though the presence of various studies that make use of the continuous parameters may indicate that the continuous phonetic dimension is well attested and accounted for in the literature, many studies argue for a need of a broader consideration of continuous aspects. Traditionally, the investigation of the continuous parameters has focused on the implementation and motivation of the categories, primarily aiming to describe the variations between categories (cf. Schweitzer, 2019) and between functions that these parameters convey. For example, it has concentrated on the regularities found in a category to define it as such and on the characteristics that it needs to have in order to convey a certain function. By contrast, Cangemi and Grice (2016), and Grice et al. (2017), among others (see e.g., Roessig & Mücke, 2019) have found the presence of within-category variation of the continuous parameters, both within and between speakers. The findings of these authors and the consequent proposal in the debate on the definition of categories will be deepened in the next section.

1.2.4.2 Variation in the categories

Within the AM framework, several works (Cangemi et al., 2016; Cangemi & Grice, 2016; Cangemi, Krüger & Grice, 2015; Cole & Shattuck-Hufnagel, 2016; Grice et al., 2017) have emphasised that intonation research should consider continuous aspects not only for defining categories by the presence or the absence of specific features, but also to account for variation within a category. Intonational categories are characterised by many dimensions, arguing for the need of a different conception of the relation between the continuous multidimensional parameters and discrete phenomena (Barnes et al., 2012; Cangemi et al., 2015; Grice et al., 2017; Niebuhr, D'Imperio, Gili-Fivela & Cangemi, 2011; see also Roessig & Mücke, 2019). The within category variation shown by Grice et al. (2017) proved to be related to different functions than the speakers wanted to convey (see 2.3 for a more detailed examination of the relation between intonation and linguistic functions). In fact, in the experiment, pitch accents belonging to the same category could differ in their phonetic implementation and this would sometimes lead to a different linguistically functional interpretation of the same category. The authors concluded that meaning can to some extent be dependent on continuous parameters and not only on determined categories. In their experiment on German, which investigated the prosodic realisation of three different focal structures (broad focus, narrow focus and contrastive focus; see 2.2 for an explanation of these concepts), the authors showed that a category might be realised in different ways depending on the pragmatic function speakers want to convey.

In their study Grice et al. (2017) found that when considering the labels obtained by a (G)ToBI labelling procedure, not all 5 speakers made a distinction between the focal condition using different accent types. In particular, one speaker almost exclusively used one category of pitch accent across all the focus types, while two other speakers seemed to not differentiate between two of the three conditions. This might lead to the assumption that some speakers fail to express some functions intonationally. By contrast, when describing the focal structures through the continuous parameters, all the speakers revealed to have similar tendencies in the distinction of the focal structure even though for some speakers the differences between the realisations of the structures were more extreme than for others. This shows that all the participants within their range of variation conveyed the relevant communicative function. The authors conclude that:

> A purely categorical (pitch accent-based) account would miss the continuous differences across and within speakers. Crucially, such an approach would also miss the similarities in the expression of the different focus conditions. (Grice et al., 2017:102)

A point that the study emphasises is the fact that the labels were assigned relying on discrete decisions that the transcribers made. While for some speakers the difference in the parameters usually used to distinguish between categories was great enough to conform to a discrete shift from one category to another, for other speakers the change remained within the limits of a single category because they were not considered enough by the transcribers to configure as a discrete change (cf. Grice et al., 2017). These results show the variability of the continuous parameters within a category and how failing to take variability into consideration may cause loss of information.

A similar result on inter-speaker variability in the signalling of different categories has been found by Niebuhr et al. (2011). Their study has shown that participants can be distinguished between *shapers* and *aligners*: to encode the contrast between two categories (H+L* vs. H* in Standard Northern German, L+H* vs. L*+H in Neapolitan Italian and H* vs. H*+L in Pisa Italian; cf. Niebuhr et al., 2011) some participants used peak alignment while others used the shape[4] of pitch movement. These findings are in line with preceding and subsequent work showing that the alignment of F0 peaks can

[4] The shape index was defined as the duration of the rise divided by the duration of the fall.

interact with the F0 shape (Knight, 2008; Barnes et al., 2012). Furthermore, variability within categories has been shown in the study by Cangemi and Grice (2016) on the Neapolitan variety of Italian. Although the literature usually shows that peaks in this variety align earlier in declaratives than in interrogatives, they have found that under some specific contexts the opposite trend occurs: interrogative peaks are aligned earlier than declaratives. Considering these results, they argue that

> differences in how variably a prosodic category is encoded can be dealt with in an intonation transcription system, as long as this system relates phonological form (the choice of pitch accent in this case) both to phonetic substance and to meaning in a transparent way. (Cangemi & Grice, 2016:1)

According to the authors, in the classical way of making use of the transcription system in AM phonology the transparent relation that they refer to does not happen, since the transcription system does not account for the variability in the mapping between the form, the function and the phonetic implementation of the two (see also 2.3.4). They admit that one possibility is reporting on the most typical behaviour of speakers in percentages, as employed by Baumann, Grice and Steindamm (2006) among others (cf. Cangemi & Grice, 2016; see above discussion of Grice et al., 2017). However, the authors argue that the variability that is attested in the studies points to a different conception of categories. Instead of being classified as the presence or absence of certain features, they can be conceived as clusters in a *multidimensional phonetic space* and can account for simultaneous and different changes of the different parameters as a consequence of the function they want to convey and of inter-speaker variability (cf. Cangemi & Grice, 2016). They thus argue for a distributional approach to the definition of categories and the mapping to functions and treat variation as an inherent property. In this approach, clusters would define a category: this approach expects categories to be different in the internal structure and in the degree of closeness among its members. Thus, the multidimensional representation proves to be crucial also for accounting for speaker specific behaviour.

In addition, a crucial aspect to take into consideration is the way in which the categories are perceived by listeners and mapped by them onto functions (see Cangemi et al., 2015). The distributional approach considers this aspect with *exemplar theory*, which explains categories in the perceptual realm. This theory is gaining success in the perceptual dimension of speech in the form proposed by Pierrehumbert (2016; Schweitzer, 2019). In exemplar theory, all the instances and their multidimensional expressions are stored in listeners' memory: similar exemplars are stored in a close space and form clusters of similar elements. Given that the exemplars of one category are considered to be perceptually similar, all the exemplars of a category are expected to form a cluster in the multidimensional space of perception. The categorical knowledge would then emerge in the form of the probability distribution of each category in the perceptual space (cf. Schweitzer, 2019). Schweitzer (2019) explains that knowledge of a category would arise through the process of associating the exemplars that belong to similar regions in the perceptual space to the abstract categories of meaning:

> If phonological knowledge arises by abstracting over clouds of exemplars stored in memory, then speech acquisition is initiated by accumulating those exemplars in memory and by starting to label these exemplars with meaning. As more and more exemplars are stored, implicit phonological knowledge begins to build up when exemplars associated with the same abstract meaning categories exhibit similar perceptual features, i.e. when they are located in similar regions in perceptual space. (Schweitzer, 2019:4)

This discussion can apply to the perception of prosodic prominence and of the mapping between categories and functions. These concepts will be elaborated on in Chapter 3.

Acknowledging the discussion so far, this thesis will provide both a description in terms of ToBI labels and their percentage of use in conveying different functions and an analysis of continuous parameters (experiments in Chapter 5 and Chapter 6). This is in line with the view of Grice et al. (2017, see above), who argue that the categorical classification does not need to be abandoned, but needs to be implemented with a deeper analysis of the continuous parameters and of the variation in the multidimensional space, advocating a system able to express both continuous and discrete phenomena (see also Roessig, Mücke & Grice, 2019). Various continuous phonetic parameters will be the objects of investigation in the experiment in Chapter 5, in order to account for a distributional approach in the definition of the mapping between the form and the function. This approach will also be useful for the definition of the degree of prominence of the post-focal (post-nuclear) position (see 1.3.1.1 and 2.3), since, as will be made clearer later, prominence is conveyed by a bundle of multiple phonetic parameters. The continuous dimensions chosen to be measured are not the traditional ones referred to in this section. As will be made clear in subsection 1.2.5, these parameters present some critical issues that can be overcome with a new methodology recently made available and involving periodic energy (Albert, Cangemi & Grice, 2018). A discussion centred around these issues and the outline of the advantages of the new methodology are provided in the following sections.

1.2.4.3 Problems with F0 turning points

The discussion in the previous sections has made clear the difficulty in describing intonational categories and drawing a line between one category and another. The use of the continuous parameters exemplified in Figure 4 in subsection 1.2.2 is traditionally what research on intonation has been using to account for this problem. However, as recently emphasised by Cangemi, Albert and Grice (2019) and previously pointed out by Barnes and colleagues (Barnes, Veilleux, Burgos & Shattuck-Hufnagel, 2012; Barnes, Brugos, Shattuck-Hufnagel & Veilleux, 2011; see also Bregman, 1994 among others), these chosen parameters are problematic in several respects. One problem that can easily be thought of and experienced when doing research on intonation is related to practical matters: these parameters require the location and the annotation of landmarks both at the tune level (tonal targets) and at the segmental level (the *text* e.g., segmental boundaries; cf. Cangemi et al., 2019). As discussed in subsection 1.2.4.1, these operations are time-consuming. Although some automatic annotations to locate the turning points of the F0 curve are available, they are also error-prone (Del Giudice, Shosted, Davidson, Salihie & Arvaniti, 2007). Moreover, the notion of tonal target and its specification in the turning point of the F0 curve has been regarded as problematic also within the AM model, for example by Barnes et al. (2012), who point out the necessity of developing a "more robust and perceptually realistic model of tonal timing and scaling patterns" (Barnes et al., 2012:342) than the ones based on the identification of the turning point in the F0. The authors underline as a problem not only the fact that in some cases the precise location of the F0 turning point is ambiguous, making the identification of the target extremely difficult, especially in the cases of high plateaux (D'Imperio, 2000b) and for low tones (cf. Barnes et al., 2012), but also the fact that the procedures established in order to deal with these problems mask how listeners deal with them in perception. Moreover, the authors emphasise the problem that in cases where the F0 curve is missing due to voicelessness the practice is still to identify turning points in the F0, while, by contrast, listeners do not restore the missing turning points during perception and perceive pitch only on the basis of the information present in the signal, namely based on the voiced parts of the signal (cf. Barnes et al., 2012).

In addition to stating the problems regarding the turning points in the F0, Barnes et al. (2012) develop an approach that considers the global shape of the contour without the need of referring to the turning point. This approach is based on the notion of *tonal center of gravity* (TCoG). The TCoG model is seen by the authors as a more robust and perceptually realistic model of the timing and the scaling of the F0 than the ones traditionally available (cf. Barnes et al., 2012), since it allows them to determine a time value that represents the center of the area under the F0 curve in a specific region, therefore taking into account alignment and scaling of the turning points without directly referring to them. Moreover, in a single measure it considers the shape of the contour. This approach has been proven successful to better account for the speaker-specific variability than the measure of individual turning points and to be a good measure in accounting for the distinction between accents produced in different focal structures (Bruggeman, Cangemi, Wehrle, El Zarka & Grice, 2017).

1.2.5 Modelling intonation using Periodic Energy

A similar approach to the TCoG is the one proposed by Albert, Cangemi and Grice (2018) and described and tested by Cangemi at al. (2019). Following the observations and the claims of Cangemi et al., experiments in Chapter 5 and in Chapter 6 will use this alternative method. The procedure that the authors propose is considered particularly useful because it not only provides measurements for the continuous parameters that can relate to the description of a pitch contour (see below), but also the dimension of energy mass (see 1.3.3). This latter dimension is crucial to convey prosodic prominence and subsequently convey pragmatic functions, as will be explained in more detail in the following sections and chapters (1.3.3 and in Chapter 2).

The novelty of the procedure proposed by Albert et al. (2018) and further developed by Cangemi et al. (2019), is that of using periodic energy, which not only allows a better estimation of the perception of pitch, but also avoids the need for text segmentation. In fact, the procedure is based on modelling both the tune and the segmental composition of an utterance (the text) without discretising them. The first step is the extraction of the periodic energy curve, where periodic energy is the part of the acoustic signal that carries pitch information. It reflects the periodic component of energy, which is very usefully analysed in association with F0, in that it estimates the strength of the signal producing the F0 and therefore reflects the intelligibility of pitch better than the F0 contour does alone (cf. Albert et al., 2018). Periodic energy is an intensity measure that selectively chooses only the periodic/harmonic components. Those are characteristic of vowels and syllabic nuclei. By contrast, the part of intensity that is constituted by aperiodic energy, namely the noise content of the signal, is characteristic of obstruent consonants and syllable margins. Moreover, periodic energy is crucial for adequately modelling the perception of tonal events, given that sensitivity to F0 is greater when periodic energy is stronger (cf. Cangemi et al., 2019; see Oxenham, 2012). The periodic energy curve typically exhibits a sequence of fluctuations over time, defining *periodic energy cycles*, which correspond to the intervals between two local minima along the curve (cf. Albert et al., 2018). Within each periodic energy cycle, peaks tend to align with syllabic nuclei. Thus, these periodic energy cycles can be used in lieu of syllabic cycles, reducing the requirement for segmenting the text (cf. Cangemi et al., 2019).

As far as the tune is concerned, the model, rather than only considering the turning points of the F0 contours, considers all the details of the shape in the regions that are relevant for the signal. Each cycle of the periodic energy normally has one peak occurring around the cycle's middle. Thus, in this curve the parameter of *center of gravity* (CoG) usually calculated for the F0 curve (TCoG, see above) can be calculated for the periodic energy curve as well. This parameter is called *center of mass* (CoM, Albert et al., 2018). This calculation takes each point in time within a periodic energy cycle, multiplies it for the periodic energy at that point in time and sums up each of these products. This sum is then divided by the sum of the periodic energy in each of the same time points within the cycle. This is explained in the formula reported in (8), taken from Albert et al. (2018). This measure allows them to find the point of equilibrium between the peak and the shape (as for the TCoG in reference to the F0).

(8) $(\sum_i per_i\, t_i) \,/\, (\sum_i per_i)$

The authors further calculate the classical center of gravity of F0 (CoG) within the syllable, in order to obtain the measure that they call *synchrony* (expressed in milliseconds), given by the distance between the CoG and the CoM. This measure is indicative of the overall F0 trend within the syllables. It gives the trend of the F0 movement locally and shows how much the F0 is moving within the syllable. This helps in characterising the movement as rising (positive values of synchrony) or falling (negative values of synchrony) *within* the syllable (see Albert et al., 2018 and Cangemi et al., 2019). An additional measure is that of *scaling* (expressed in Hertz). To obtain this measure the F0 at the CoM for one syllable is obtained and then subtracted from the same measure in the previous syllable. Therefore, scaling indicates the value of F0 on one syllable as compared to the previous one and reflects the F0 direction *across* syllables (see Albert et al., 2018). A more detailed explanation of the procedure will be given in the method section of Chapter 5.

Figure 6 provides a visual exemplification of the parameters described so far. The figure shows for three realisations of the word *limone* (lemon): the F0 modulated by periodic energy (first curve from the top; *periogram*, see below), the periodic energy curve (second curve from the top), the CoG and CoM for each syllable, and the values of scaling and synchrony for each syllable[5] (numbers at the bottom). The first realisation of the word *limone* (Figure 6a.) features a rising contour associated with the stressed syllable *mo*. Accordingly, this stressed syllable presents positive values of synchrony (44 ms) and positive values of scaling (42 Hz). By contrast, the following realisation of *limone* (Figure 6b.) presents on the stressed syllable a rise compared to the preceding syllable (*li*), indicated by positive values of scaling (30 Hz), while the majority of the movement within the syllable is falling, as indicated by negative values of synchrony (-20 ms). In the third realisation of the word (Figure 6c.), values of scaling and synchrony describe a rise on the stressed syllable, whose peak is realised closer to the center of the syllable compared to the first two realisations: lower values of synchrony compared to the first realisation (a.) and higher values of synchrony compared to the second realisation (b.). It should be noted that in all the three realisations, the values of periodic energy tend to be higher on the stressed syllable than in the preceding and following syllables. This aspect will be further discussed in 1.3.3.

[5] Note that the value of scaling for the first syllable is not present, since there is not preceding syllable to which the calculation can refer.

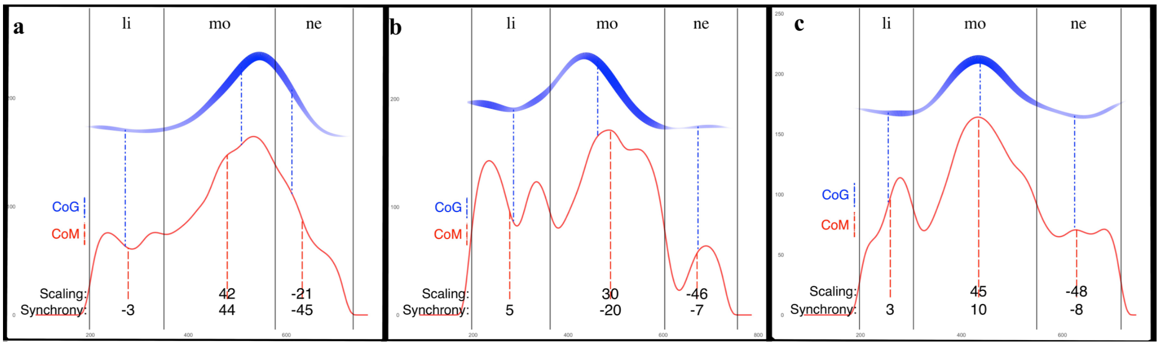


Figure 6. Periograms (F0 modulated by periodic energy) and periodic energy curves for three different productions of the word limone (lemon). Black solid lines signal syllable boundaries, dotted-dashed vertical lines denote the center of gravity (CoG) between syllable boundaries, dashed lines denote the centre of mass (CoM) between syllable boundaries. Numerical values of the bottom indicate scaling (measured as the difference between values of F0 at the CoM of syllables that are consecutive) and synchrony (measured as the distance in time between CoG and CoM of the same syllable).

As can be seen in the figure, an additional advantage that the procedure of Albert et al. (2018) allows is having an informative visual representation of both the F0 and the periodic energy curves. This is achieved through what the authors call *periogram* (first line from the top in Figure 6) and reflects the modulation of the F0 with periodic energy data. Through this modulation, a more informative alternative for the visual inspection of the F0 is achieved (cf. Albert et al., 2018). In fact, the periogram reflects the strength of each time point in the F0: the F0 curve is thus represented with a dynamically changing width and transparency, being wide and solid at the most periodic parts, and becoming narrower and more transparent when periodic energy drops (cf. Albert et al., 2018:807). The parts with no periodicity are not displayed in the periograms. One advantage is to distinguish basing only on the visualisation differences among rise-falling accents, mainly falling accents and mainly rising accents (cf. Albert, 2018a). This is because the periograms allow for the display of a perceptually motivated delineation of the pitch contour: the wider and thicker parts of the curve are the ones that are more strongly perceived. For example, in Figure 6b. the darker and wider part of the periogram is the one pertaining to the falling part of the contour. For a more detailed discussion of the issue and for more examples pertaining to the issue of distinguishing a rising from a falling contour see Albert (2018a).

A further advantage of the model is that it allows for fewer theoretical assumptions than the traditional annotation of the AM model and thus represents a promising approach for the unification of the research on intonation, because it can be applied across frameworks (cf. Cangemi et al., 2019). It is also useful to account for the variability within and between speakers in the phonetic implementations of functions (Cangemi et al., 2019).

This feature will be of use in the experiment in Chapter 5, which will deal with the prosodic implementation of different focal structures and particularly with the dimension of prosodic prominence, which is crucial in conveying pragmatic functions and focus structures in particular (see 2.3). Taking into consideration the results and the discussion in Grice et al. (2017), variation has to be expected and to be regarded as a source of information and this procedure helps in this regard.

Now that the features of pitch and the matters regarding its description have been delineated, the next section will address a central concept in prosody: prosodic prominence. It will first outline the relations and the degrees of prominence as seen within the AM model and then deal with its acoustic correlates. These matters will be applied in the discussion on Italian.

1.3 Prosodic prominence

1.3.1 Prominence relations

Highlighting is one of the main functions of prosody, marking the *prominence relations* within the prosodic structure. The conceptualisation of prominence implies its relational property which refers to any unit that somehow *stands out*. Terken and Hermes (2000:89) define prominence "as a property of a linguistic entity relative to an entity or a set of entities in its environment". Liberman and Pierrehumbert (1984) identify prominence as a local degree of emphasis. Prominence is an essential criterion of a strong element among elements of the same level of the prosodic hierarchy. Although this and other characteristics of prominence can refer to other levels of linguistic description (see Himmelmann & Primus, 2015; von Heusinger & Schumacher, 2019), the use of the term prominence in this thesis refers primarily to prosodic prominence (see Cangemi & Baumann, 2020 for an in-detail description of the concept of prosodic prominence from the early days to very recent literature). The description of prominence and of the relation between levels of prominence that is provided in this section concerns only intonation languages.

Before deepening the description of prominence, it should be mentioned that since the definition of prosodic prominence implies the understanding of prominence in terms of a relational property (see below), this results in the fact that syllables can be prominent partly independently of their actual phonetic realisation: a non-prominent syllable in shouted speech will be more perceivable than a prominent syllable in whispered speech, but the latter will be still more perceivable compared to non-prominent syllables in whispered speech (cf. Baumann & Cangemi, 2020:1-2). In addition, the understanding of prominence is also connected to the organisation of the environment around the prominent element, to which the speaker wants to attract the listener's attention (see Baumann & Cangemi, 2020). These concepts will be addressed in more detail in the following chapters (Chapter 2 and Chapter 3). These two aspects characterising prominence give a hint as to the complexity of the definition of prosodic prominence. Indeed, although there is a large body of research investigating prominence, its conceptualisation and adequate standardised measures to define it are still a matter of debate. For example, Wagner, Ćwiek and Samlowski (2019:1) argue that prosodic prominence to this day still seems to provide an umbrella term for lexical stress, pitch accent, sentence stress, prosodic focus, rhythmic alternations, paralinguistic emphasis or loudness. Therefore, a discussion of the definitions and the levels of prominence that are relevant for this thesis is necessary.

The notion of prominence proposed in the AM theory involves strength (prominence) relationships among the units of the *metrical grid,* a concept taken from the tradition of *Metrical Phonology* (Leben, 1971; Liberman & Prince, 1977). An exemplification of the grid and of its prominence relations is made in Figure 7. Each constituent presents a *metrical head*, thought to be the most prominent element of the constituent (Beckman & Edwards, 1994). This is a hierarchical structure that resembles the tree of the prosodic structure (which has been depicted in Figure 2) and consists of a series of beats which determine an abstract representation of rhythm. To understand how prosodic prominence in intonation languages works, it is important to distinguish between stress and accent. Beckman (1986) explains this distinction stating that stress corresponds to the syllables that are strong at the level of the word. Therefore, stress corresponds to lexical information that a speaker has to learn (e.g., the distinction between PERmit and perMIT in English or between ANcora and anCOra in italian). By contrast, within an intonational phrase the prominence of one word in comparison to the others is determined by pragmatic and functional factors (see 2.3) and is realised by the association of a pitch accent to the stressed syllable of the designated word. The stressed syllable of a word provides a potential place where prominence at the sentence level may occur (cf. Ladd, 2008:51). Weinreich (1954), Lehiste (1970) and Ladd (2008) distinguish between abstract word stress and concrete sentence stress: “Word-level stress is the capacity of a syllable within a word to receive sentence-stress when the word is realised as part of the sentence [..]” (Lehiste, 1970:150).

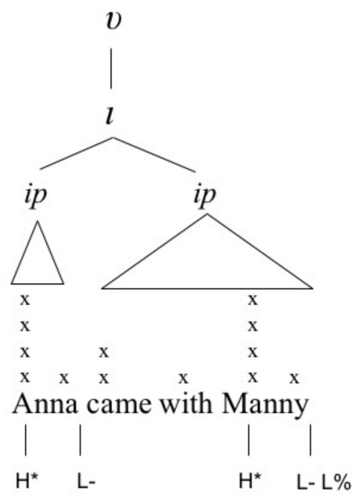


Figure 7. Metrical grid of prominence relations. Adapted from Beckman (1996:35)

Syllables stand in a prominence relation with each other (see Beckman & Edwards, 1994; Vanderslice & Ladefoged, 1972). The syllable that is *stressed* at word level is more prominent than the other syllables within the word. All stressed syllables are potential docking sites for tonal movements. However, not every stressed syllable needs to be accented. Thus, the accented syllables in an ip are more prominent relative to the stressed syllables that are not accented. An example of the metrical grid for one possible prosodic realisation of the utterance *Anna came with Manny* is shown in Figure 7[6]. Here, the strength relations in the hierarchy are illustrated: the relative strength of each beat is represented by the number of ‘x’ on the vertical axis. Figure 7 shows the realisation of the utterance with two ips and one IP, which corresponds to the level of the utterance (since the IP has the same dimension of the whole utterance). At the level of the word, three syllables are the metrically strongest syllables. These are lexically stressed syllables (*an*, *ca* and *ma* in the figure). In this case the stressed syllables are the metrical heads of the words that contain them. In addition to the strength relation within each word, a strength relation is also present at the level of the ip. The words *Anna* and *Manny* are prominent within their ip, both bearing a prominent pitch accent at the level of the IP.

[6] For exposition purposes, the level of the foot is excluded.

As seen so far, there are different linguistic levels of highlighting. An approach which distinguishes various levels of prominence is proposed by Weinreich (1954) and Lehiste (1970) and embraced by Grice and Baumann (2007). The levels of highlighting are exemplified in Table 3. Here, a distinction is made between *abstract prominence* and *concrete prominence*. Abstract prominence is considered the one at the level of the word and is represented by lexical stress. Thus, lexically stressed syllables are merely seen as a potential place where the actual prominence at sentence level may occur (cf. Ladd, 2008:51). The concrete notion of prominence further distinguishes two types of post-lexical prominence: *post-lexical stress* and *accent* (Grice & Baumann, 2007; see also Ladd, 2008). Post-lexical stress and accent both represent a concrete prominence at the level of the sentence. Ladd (2008:53) explains the idea stating that stressed syllables "may or may not be actually prominent in an utterance; if they are actually prominent in an utterance, they may or may not be pitch accented". In Ladd's understanding of actual prominence, this feature is characterised by a complex interaction of phonetic features, which reflect rhythmic regularities and greater force of articulation, while pitch accent is considered an additional feature.

lexical stress:	word level (abstract prominence); potential position for utterance level prominence (concrete prominence)
post-lexical stress:	phrase-level, concrete prominence
accent:	phrase-level, concrete prominence

Table 3. Levels of prominence description. Table adapted from Grice and Baumann, 2007:3.

The difference between post-lexical stresses and accents entails a difference in the degree of (post-lexical) prominence. This distinction involves differences in the phonetic implementation (for example, the contribution of pitch) and will be explained in subsection 1.3.2.

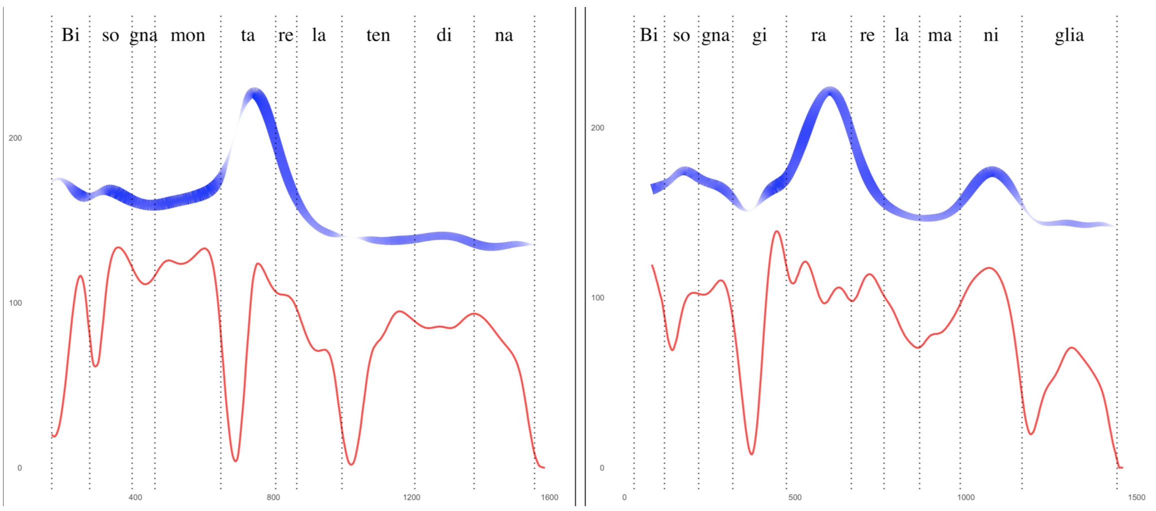
1.3.1.1 Nuclear pitch accent and definition of further degrees of prominence

The approach described above and referred to by Beckman and Edwards (1994; see also Hayes, 1995), assumes that the levels in the hierarchy are headed by prominences. Traditionally, for both West-Germanic languages and Romance languages, the head of each constituent is considered to be the rightmost prominent element in each unit. In the AM model, the last accent of an intonation unit has a special status. This is referred to as the *nuclear accent* and is considered to be the strongest accent in a phrase, not only in a structural sense, but also from the perspective of perception, since this has also been attested by perception studies (see Baumann & Winter, 2018; Cole et al., 2019; Silverman & Pierrehumbert, 1990; see also 3.4). As seen in the description of Figure 7 both the accents associated with *Anna* and *Manny* are marking prominences at the level of the utterance. However, at this level there is a further relation of prominence: the last accent at the level of the utterance is usually considered the most prominent accent by virtue of its position (see also the notion of edge placement in Himmelmann & Primus, 2015). There is also a further distinction between the nuclear accent at the level of the ip and the one at the level of the IP (or of the utterance).

The definition of the nuclear accent is taken from the tradition of the British school of Intonation (Crystal, 1969), in which the notion of nuclear tone is a fundamental concept. The nuclear tone is defined as the last major intonation movement in the utterance and following movements, defined as *tail,* are determined by the nuclear tone (cf., Xu, 2011). In the AM the components that come after the nuclear accent are characterised as deaccented, lacking recognisable F0 movements. However, this definition of nuclear accent is not uncontroversial, as will be explained in the following section.

1.3.1.2 Issues in the definition of nuclear pitch accent

The AM model implies that the structural realisation of nuclear prominence is realised on the word that is the head of the prosodic phrase and additional prominences are considered to commonly occur in prenuclear position. This conception can be regarded as valid, but variation has to be accounted for. Indeed, languages vary in the patterns of prominences (Cole et al., 2019; see 2.3). For example, Italian seems to represent an exception to the nuclear assignment rule: interestingly, in some varieties of Italian, prominence associated with a pitch accent is attested after the nuclear pitch accent of an IP. For example, a down-stepped compressed rising accent is attested in post-nuclear position of questions in some southern varieties (Bari, Naples and Palermo; Cangemi & D'Imperio, 2013; Grice, 1995; Grice & Savino, 2003; Grice et al. 2005; Savino, 2012 among others; see also 2.3.2). Figure 8 presents the periograms (and the periodic energy curve) of a statement (a. *Bisogna montare la tendina.* One needs to mount the curtain.) and a question (b. *Bisogna girare la maniglia?* Does one need to turn the handle?) produced by a speaker of the variety of Italian spoken in Bari. Both utterances present the nuclear accent, marking a contrastive focus (see 2.3) on the verb (*montare* and *girare* respectively), which in both cases occurs as the second word in the utterance. In the statement (Figure 8a.) after the pitch movement characterised by a great F0 excursion (realised on the verb *montare*), the contour on the following constituent is flat and low (*la tendina*). By contrast, in the question (Figure 8b.) after the great F0 excursion on the verb (*girare*), the contour on the following constituent (*la maniglia*) features a consistent movement in pitch. The periogram allows us to see that, while in the last constituent of the statement the signal producing the F0 is very low in strength (narrow and with high transparency), in the last constituent of the question the signal is stronger (wider and thicker lines of the curves). Indeed, the periodic energy curve (line on the bottom), reaches concomitantly with the stressed syllable of the last constituent of the question (*ni*), higher values than the values reached concomitantly with the stressed syllable of the last constituent of the statement (*di*).



a. b.

Figure 8. Periograms (upper curves) and periodic energy curves (bottom curves) of a. the statement Bisogna montare la tendina (One needs to mount the curtain) and b. the question Bisogna girare la maniglia? (Does one need to turn the handle?), both produced by a speaker of the variety of Italian spoken in Bari.

The presence of these accents has raised discussions about the status of the nuclear accent, which has been described as *the last fully-fledged accent in the IP* (see Grice et al., 2005). The definition of nuclear accent, in fact, comprises the notion of fully-fledged accents, which have been attested to typically occur in prenuclear or nuclear position and are considered as fully realised accents (i.e., not compressed). The accent on *maniglia* present in Figure 8b. is therefore defined as non-nuclear (or post-nuclear) accent, since it is compressed. In addition to what has been reported for the variety of Italian spoken in Bari, in the varieties spoken in Tuscany (in Florence and Siena) an accent with no pitch movement (labelled as L*) has been claimed to occur in post-nuclear position in cases in which this is the metrical head of an intermediate (phonological) phrase (Bocci & Avesani, 2011).

Bocci and Avesani (2011) found prominence lending acoustic characteristics in the metrically strong position in the post-nuclear region. The acoustic characteristics that they found will be discussed in the following subsection (1.3.2). Within AM theory, this has initiated the discussion about the structural assignment of prominences and about the accentual status of the post-nuclear material (Bocci, 2013; Bocci & Avesani, 2011). The authors argue, against Vallduví (1992) and Szendröi (2002), that the post-focal material is prosodically phrased (and not extrametrical, as suggested by the two previously mentioned authors) and that constituents in Italian (and in general in Romance languages) are assigned phrasal metrical prominence, which prevents the nuclear accent from being aligned with the rightmost phrasal prominence. Moreover, the authors argue that the allocation of phrasal-level prominence in post-focal position found in the Tuscan variety of Italian, does not derive from discourse-related properties (see 2.3), but is determined by the default rules of prominence assignment, which demand every prosodic constituent to be assigned a head. The structural debate in Bocci and Avesani's study is not central to this thesis. However, their findings might be considered interesting in the discussion of the degrees of prominence and of the acoustic correlates of prominence. Indeed, results found by Bocci and Avesani foster the idea that the phrasal prominences are not always realised via pitch movement.

The presence of these post-nuclear accents can be to some extent related to Kohler's (2006) concept of *duration accent* and *force accent*, allowed only in pre- and post-nuclear position. Another similar concept is the notion of *phrase accent*, proposed by Grice, Ladd and Arvaniti (2000). These concepts are elaborated on in the following section, which will focus on the acoustic correlates of prominence and will further the understanding in the different levels of prominence and in Chapter 2, where a discussion on phrase accents is provided.

1.3.2 Acoustic correlates of prominence

The acoustic correlates of prominence are language dependent (Andreeva, Barry & Koreman, 2014; Rosenberg, Cooper, Levitan & Hirschberg, 2012) and can vary in different dialects (Smith & Rathcke, 2020). However, F0 movements have been described in many languages as the primary cue for prominence (Beckman, 1986; Ladd, 2008; Liberman, 1975; Pierrehumbert, 1980; Roca & Johnson, 1999 among others). In an early study, Bolinger (1958) claims that actual prominence at sentence level is only realised through pitch movement on designated stressed syllables. Along this line, several studies have investigated the relation between F0 variation and prosodic prominence at utterance level. As correlates to prosodic prominence, they have indicated the size of F0 excursion and the variation in the shape (Baumann & Röhr, 2015; Fry, 1958; Gussenhoven & Rietveld, 1988; Hermes & Rump, 1994; Mahrt Cole, Fleck, & Hasegawa-Johnson, 2012; Niebuhr, 2009; Terken, 1991 and many others; see also 3.2), and relative pitch scaling (Xu, 2011; Xu & Xu, 2005; see also Wagner & Watson, 2010). However, the role of the pitch accents as correlate to prominence is not easy to define, since the accent may be realised with several different F0 contours, which include cases where the stressed syllable has a small F0 excursion (e.g., the above-mentioned case of the L* pitch accent). The difference in the acoustic implementation of pitch accents leads to a scale of different degrees of prominence within pitch accents (Baumann & Röhr, 2015; see 3.2.2), where a higher pitch excursion has been attested to increase the degree of prominence of a word (Baumann & Röhr, 2015; Ladd & Morton, 1997; Rietveld & Gussenhoven, 1985; see also 3.2.2).

However, although F0 variations are a strong acoustic cue to prosodic prominence, many more recent studies have attested that speakers employ various phonetic dimensions to signal prominence. There is evidence that prominence at sentence level is not exclusively associated with pitch variations. With regard to accents, they are expressed not only by F0 variations alone but by a complex interaction of parameters such as increased intensity, increased duration, spectral emphasis and articulatory effort (see e.g., Avesani, Vayra & Zmarich, 2007; Baumann & Winter, 2018; Batliner, Nöth, Buckow, Huber, Warnke, & Niemann, 2001; Beckman & Edwards, 1992; Fry, 1955; Kochanski et al., 2005; Roessig & Mücke, 2019; see also Wagner & Watson, 2010). Influence on the level of prominence of a given stressed syllable is also given by prominence of the adjacent syllables (Arnold, Wagner, & Baayen, 2013) and the position of the syllable in the utterance (e.g., nuclear pitch accent, see above; Gussenhoven & Rietveld, 1988). How all the cues to prominence interplay in the expression of prominence is still not clear (cf. Wagner et al., 2019). Not only is prominence not exclusively conveyed by pitch movement, but speakers are also able to use multiple features in different combinations to convey the same prominence degree (cf. Roessig & Mücke, 2019:3).

In relation to the prominence hierarchy discussed in subsection 1.3.1 and exemplified in Table 3, post-lexical stressed syllables have been proven to be reliable at utterance level also when not associated with pitch accents. The parameters that identify lexical stress are duration (e.g., Campbell, 1993) and spectral tilt (e.g., Huss, 1978; Sluijter & van Heuven, 1996; Beaver, Clark, Flemming, Jaeger & Wolters, 2007), corresponding to the energy distribution in the spectrum. Sluijter and van Heuven (1995, 1996) studied Dutch, aiming to find the acoustic correlates of the two different levels of prominence: stress and accent. They employed the measure *spectral balance,* corresponding to intensity differences at the higher frequencies of the spectrum and found that this measure correlates more strongly with lexical stress than with accent (see also Astruc & Prieto, 2006 for evidence on Catalan).

These studies suggest that the structural distinction made in Table 3 has its acoustical basis: different acoustic measures are related to different levels of prominence. Duration and spectral balance correlate more likely with (post-lexical) stress and overall intensity is more reliable as a correlate of accent. Baumann (2006) reports that further support for the relative independence of post-lexical stress from accents is given by the study of Shattuck-Hufnagel, Ostendorf and Ross (1994) and by the study of Harrington, Beckman, Fletcher and Palethorpe (1998). In the former, the authors investigate the phenomenon of shift in the perception of prominence of the word *MassaCHUsets*[7] when put in rhythmic clash context, for example *MassaCHUsets* MIracle. In this latter case, the syllable perceived as prominent in the state's name is not the lexically stressed syllable, but the first one (*MA*; *MassaCHUsets* → *MAssachusets*; cf. Baumann 2006:10). Shattuck-Hufnagel et al. found that the position of the stress actually remains constant and it is just the F0 movement that causes the perception of prominence on the non-stressed syllable (*MA*). Moreover, this movement does not cause the increase of duration on this syllable.

[7] Capital letters indicate the prominence syllable.

Harrington et al.'s (1998) further show the independence of stress from accent. In this study, the authors investigated the post-nuclear deaccented position. They showed that in this position, the effect of accent is neutralised, while the same does not hold for the effect of stress (cf. Baumann, 2006:10; see also Howell, 2016).

From this discussion and the previous one on the status of nuclear accent, at least all the levels of prominence listed in Table 4 can be identified.

No stress/accent:	no degree of prominence
(Postlexical) stress:	Prominence at the word level, perceived also at the phrasal level. A stressed syllable is louder, longer and more strongly articulated, with less vowel reduction than an unstressed syllable.
Pitch accent:	Prominence at phrasal level. A syllable bearing a pitch accent (i.e., accented syllable) has additional tonal movement on or in the direct vicinity of a stressed syllable.
Nuclear pitch accent:	The nuclear syllable is the last pitch accent in an intonation phrase, usually perceived as the most prominent one in the phrase (see 3.2.2.).

Table 4. Degrees of prominence. Adapted from Grice and Baumann (2007:3).

A further level in the prominence hierarchy is suggested by Kohler (2005, 2006). He states that there can be accents that are not marked by pitch variation and therefore introduces further concepts of prominences at utterance level. He proposes a distinction between three types of sentence accents: *pitch accents*, *duration accents* (Kohler, 2006) and *force accents* (Kohler, 2005; see also Gagliardi, Vallauri & Tamburini, 2012). In his account force accents are not related to pitch and are based on increased intensity and increased duration, given by the *phonatory* and *articulatory force* (cf. Kohler, 2005:99). On the contrary, pitch accents and duration accents are related to pitch features but the realisation of the latter can be independent from pitch. Kohler (2006) assumes that the sentence accents comprise four distinct levels: unaccented, default accented, partially deaccented and reinforced (cf. Röhr, 2016:79). He argues that the phonetic manifestation of the default and the reinforced accent levels are signalled primarily by F0 movements and are thus called *pitch accents*. On the contrary, the partially deaccented level is primarily signalled by duration and increased energy, even though it can be accompanied by an F0 peak inflection that is below the F0 peak declination[8] (cf. Kohler, 2006:749). The accents characterised by these features are called *duration accents* and may occur both in prenuclear and post-nuclear position (cf. Baumann, 2006:12). In Kohler's view, post-lexical prominences, like the one described by the duration accent, are not to be interpreted simply as expressions of word stress. In addition, they are also different from fully-fledged accents. This concept seems to resemble the one presented by Bocci and Avesani (2011), who argue for the possibility of having (pitch) accents without pitch movement in the post-nuclear position (labelled as L* pitch accents). This argument stems from the fact that they found a syllable occurring after the nuclear accent in a metrical strong position to be higher in some prominence lending feature (duration and spectral tilt) than both the syllable occurring postfocally in a non-metrically strong position and the syllable occurring in a non-metrically strong position, but being part of a broad focus domain, therefore bearing a pitch accent (see 2.3).

The fact that a syllable in post-nuclear position can present some prominence characteristics that are similar to the ones of a syllable occurring in broad focus (see next subsection), has also resulted from Mücke and Grice's (2014) study on German. The authors investigated parameters connected with lip kinematics and found that accented words in broad focus were not articulated in a systematically different manner from unaccented words occurring in post-nuclear position (as part of the background; see 2.2.1 for clarification of the term background). By contrast, systematic modifications of the oral articulators (involving the strategy of *sonority expansion*, see below) were found for the differentiation between words in broad focus and in contrastive narrow focus. The authors conclude that in the case of broad focus, since the examined word is not the only constituent under focus, meaning it does not occupy a particularly prominent position (see Chapter 2 for a clarification), its realisation involves only modifications at the level of the larynx, that is to say in terms of pitch accent. The modification of oral articulators such as the lips o the jaw is not necessary (cf. Mücke & Grice, 2014:59). Likewise, the study of Roessig and Mücke (2019) supports the view that the expression of different degrees of prominence conveyed through different modifications of different articulatory systems. Nevertheless, it should be noted that this study found systematic modifications of the supra-laryngeal articulatory system between words in post-nuclear position and words occurring in broad focus.

[8] F0 declination refers to the descending trend of F0 during the course of an utterance.

To sum up, these studies attest that the laryngeal system is not the only one engaged in conveying prominence, but there are several strategies of the supra-laryngeal system that help in conveying the contrast among different degrees of prominence. One of these strategies is called *sonority expansion* (Beckman, Edwards & Fletcher, 1992) and involves the wider opening of the mouth to convey prominence: when the oral cavity is more open, this allows for a greater radiation of acoustic energy. This in turn results in the production of louder and more sonorous syllables. A second strategy is the *localised hyperarticulation* (de Jong, 1995) which involves the more extreme articulation of the tongue body in the production of vowels. As thoroughly explained in Roessig and Mücke (2019), in producing low vowels, the strategies of sonority expansion and hyperarticulation are not competing with each other: the low position of the tongue and the jaw with a further higher degree of lip opening increase specifications of manner and place targets. On the contrary, in high vowels these two highlighting strategies seem to compete against each other: sonority expansion requires a more open vocal tract to produce louder values, while localised hyperarticulation leads to smaller degrees of constriction to increase the place feature of the vowels (cf. Roessig & Mücke, 2019:3). Nonetheless, different articulatory subsystems can coordinate and combine the two strategies. This leads to syllables that are louder and longer and are characterised by more peripheral formant frequencies (cf. Roessig & Mücke, 2019:3).

The discussion so far points to the possibility of different levels of prominence and of different contributions to the degrees of prominence. Given these different contributions to the degrees of prominence and the findings reported in the literature, the experiment in Chapter 5 and part of the experiment in Chapter 6, will investigate energy, duration and F0 in the post-focal (post-nuclear) position in corpora of two varieties of Italian. The aim is to see whether there can be a manifestation of a certain degree of prominence in this position and in the corpora collected, which can be treated as an accent, despite there being no F0 movement. The methodology used to investigate the degrees of prominence is explained in the following subsection (1.3.3).

1.3.3 Periodic Energy Mass

Although the features of prominence connected to an increase of energy are connected to articulation, the research on intonation has involved few articulatory studies due to the demanding processes of collecting, maintaining and quantifying data (but see the study of

Roessig & Mücke, 2019, which involves an extraordinarily high number of participants and see Mücke & Grice, 2014 among others). As opposed to this, measures on the F0 are easy to collect. To account for this problem Albert et al. (2018; see also Cangemi et al., 2019) propose to rely on the measure of periodic energy, which has been proven to correlate well with sonority (cf. Albert et al., 2018; see also Ladefoged, 1997; Heselwood, 1998) and can account for strategies like sonority expansion, which implies an increase in energy (see above). Moreover, the periodic energy is directly connected to the strength of perceived pitch and can therefore account for phrasal prominence connected to pitch features, as the one described by Kohler's duration accents and by the L* accent proposed by Bocci and Avesani (2011).

As Albert (2018b) suggests, acoustic cues to prosodic prominence are related to the perception of pitch. In line with this claim, Albert et al. (2018) argue that periodic energy offers an advantage in comparison with measures of overall intensity since, in contrast to intensity, periodic energy is directly connected to the strength of perceived pitch. Periodic energy cycles highly equate to syllabic units, thus, it would be possible to measure the overall intensity and duration of each periodic cycle, and this would equate the standard measure of overall intensity and duration of the syllable, as usually done in prosodic research (cf. Albert et al., 2018). However, as the authors note, these calculations result in two reciprocally exclusive values which would not directly correlate to the perception of the strength of the cycle. In fact, duration is indifferent to energy and (band-pass) intensity combines periodic and aperiodic components of speech, making it less reliable as a measure of pitch intelligibility, since the perception of pitch is based only on the periodic component. As Albert (2018b) argues, the different acoustic variables of energy, duration and F0 have the shared goal of manipulating prosodic prominence, and using periodic energy instead of intensity allows the perceptual mechanisms that contribute to prominence to be better described.

A particularly useful measure that can relate to prominence is that of *Periodic Energy Mass* (PEM; see Albert et al., 2018:806). This measure corresponds to the area under the periodic energy curve for each periodic energy cycle, indicated by the filled part in Figure 9. Corresponding to the sum integral of periodic energy and duration, this measure accounts for both these dimensions. Since prosodic strength is characterised by the variation of both energy and duration dimensions, PEM reflects the relative strength of the periodic cycle (cf. Albert at al., 2018). Figure 9 depicts the F0 contour (upper line), the intensity profile (line in the middle) and the periodic energy curve (line on the bottom) of the question *Hai girato per il mondo?* (Have you travelled the world?), pronounced by a speaker of the variety of Italian spoken in Bari in broad focus (see 2.2.1 for a clarification of the concept). The figure clearly shows the cycles of periodic energy, which correspond to each syllable of the utterance. The area under the curve of each cycle corresponds to the periodic energy mass related to that cycle (and to the relative syllable). The example in Figure 9 shows only one filled part under the periodic energy curve for explanatory purposes, but for each cycle the PEM value can be calculated. Indeed, at the bottom of Figure 9 for each cycle are reported the values of the area under the curve (AUC), relative to each syllable in the utterance. Note that there is a rising-falling F0 movement on the stressed syllable *mon* of the word *mondo*, which indicates the sentence modality of question in this variety, and which is associated with the highest values of PEM in the utterance. The interaction of these acoustic cues makes the syllable *mon* the acoustically most prominent one in the utterance.

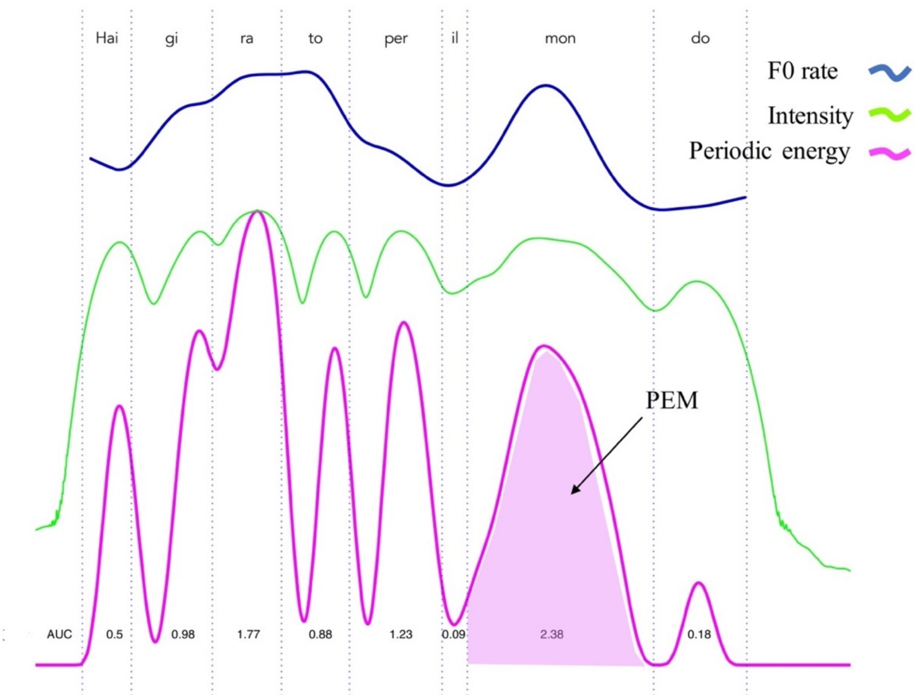


Figure 9. F0 contour (upper line), intensity profile (line in the middle), periodic energy curve (line on the bottom) and AUC values of the question Hai girato per il mondo? (Have you travelled the world), uttered in the variety of Italian spoken in Bari. Dotted vertical lines indicate the syllable boundaries.

PEM will be used in the experimental part of this thesis (Chapter 5, Chapter 6 and Chapter 7) to investigate prominence profiles related to pitch in different focal structures.

The question to be discussed is whether different degrees of prominence are produced and perceived by speakers and listeners.

1.3.4 Categorical versus continuous aspects

There is an issue regarding the conception of prominence as either a discrete binary feature or a gradient one. One possibility would imply that words are categorised by means of a binary distinction as prominent or non-prominent. Pitch-accentedness is often considered a categorical phenomenon (e.g., Bruce, 1977, Ladd, 2008, Pierrehumbert, 1980), whereas prominence is assumed to be more gradient (cf. Schweitzer, 2019:2; see also Wagner et al., 2019). A categorical view regarding pitch accent can be justified by the distinction between accented and unaccented words that listeners and speakers seems to be able to make (Baumann & Winter, 2018). A view that considers both the gradual and the categorical perspective can be seen in metrical stress theory, which entails a multi-layered concept of prominence: the degree of prominence of an element depends on the level of embedding in the prosodic structure of the constituent to which it belongs (cf. Mahrt, Cole, Fleck & Hasegawa-Johnson, 2012:2422). However, the assignment of the head in this theory is seen as a categorical process, in which the status of nuclear accent is assigned to the accent which is associated to a specific syllable in the structure. A change in the position of the head (e.g., in the case of an early focus, namely focus fronting, see 2.3) implies a reorganisation in the prominence distribution and in the prosodic prominence structure, as the nuclear accent is placed on the first element and the head status is moved from the rightmost element to a non-final element. By contrast, Roessig and Mücke (2019) propose a more gradient distribution of prominences, which is not configured by the reassignment of the heads of the constituent in the IP and is more in line with what happens in the perception realm (see 3.2.2). The authors provide evidence that prominence is expressed gradually, that it is crucial to analyse categorical and continuous aspects simultaneously and that those aspects do not need to be separated but need to be considered as a single system (cf. Roessig & Mücke, 2019:15-16).

1.4 Summary

The present chapter has introduced relevant concepts of intonation and prosodic structure. These notions were elaborated within the Autosegmental-Metrical model. Critical issues within this model were presented, in particular the shortcomings of relying only on discrete categories when investigating the mapping between form and function. This problem will be further considered throughout Chapter 2 and Chapter 3, where the lack of a one-to-one mapping between prosody and pragmatic functions will be further addressed. Innovative ways to measure continuous parameters relating to intonation have also been discussed. The procedure involving periodic energy has been presented not only as a perceptually more appropriate way to describe pitch accent in comparison to the procedure based on turning points, but also as a way to measure prosodic prominence (through periodic energy mass). These continuous parameters will be used in the experimental part of the thesis (Chapter 5, Chapter 6 and Chapter 7).

Chapter 2

Prosody, information structure and information status

2.1 Introduction

The previous chapter addressed concepts regarding the prosodic organisation of speech, particularly the prosodic function of highlighting and the acoustic characteristics related to it. The present chapter will address the role of prosody and prominence in marking information structure. Indeed, prosody reflects the organisation of the discourse driven by the evolution of mutual expectations and beliefs of the participants, and the degree of givenness of the discourse entities (i.e., information status), which indicates the activation of an entity in the mental representation of the interlocutors. The aim of the chapter is to provide an overview of studies on the mapping between prosodic prominence and the discourse functions that it conveys. Special consideration will be given to the different degrees of prominence assigned to elements in relation to their degree of givenness in the discourse and of their position in the utterance (with particular interest in the post-focal position). This discussion will focus on German and Italian, in order to provide a theoretical background for the experiments on Italian, both in production (Chapter 5 and part of Chapter 6) and perception, in particular of the post-focal region (Chapter 6 and Chapter 7). Part of this discussion will include the marking of the post-focal position in different sentence modalities (questions and statements). The difference in prominence relations between these two modalities in the two varieties of Italian involved in this thesis (Bari and Udine), and in Italian and German, will be of interest both for Chapter 6 and Chapter 7.

Firstly, section 2.2 will provide a brief introduction to information structure, introducing the basic concepts of focus and background and the degrees of activation of referents. Particular attention will be devoted to the concept of givenness (high activation). Secondly, section 2.3 will discuss the prosodic marking of information structure and of different degrees of givenness, specifically concentrating on the similarities and differences between Italian and German. Particular emphasis will be directed to the variability of the mapping between prosodic form and communicative function and between the distribution of prominences within the utterance.

2.2 Basic concepts of information structure

Speakers engage in speech to fulfil an aim. This aim can be manifold: share news, connect with others, amuse, provoke an action. The organisation of the discourse people engage in reflects the content and the goals that speakers want to convey. This organisation is called information structure or information packaging (cf. Arnold, Kaiser, Kahn & Kim, 2013; see also Lambrecht, 1994; Chafe, 1974). Information structure is concerned with the relation between the distribution of information units of a message over a sentence and the linguistic form that they assume (Krifka, 2007; Lambrecht, 1994). It is defined as the internal structure of the utterance, which reflects how the utterance relates to the discourse context, in terms of the degree of activation of its content in the discourse, the attentional states of the participants in the discourse, and their knowledge, beliefs, intentions and expectations (cf. Kruijff-Korbayová & Steedman, 2003:250). One of the dichotomies that information structure subsumes, which is of interest to the present thesis, is the division of the utterance into focus and background, namely in a more informative and less informative part (see 2.2.1). Several accounts describing information packaging (von Heusinger, 1999; Krifka, 2007; Fèry & Krifka, 2008; Zimmermann & Fèry, 2010) consider both the roles of speaker and listener and follow Chafe in arguing that the packaging of the information corresponds to "ways in which a speaker accommodates his speech to temporary states of the addressee's mind" (Chafe, 1976:28). Through information structure the participants' knowledge state is instantiated in the discourse (Chafe, 1974; Büring, 2007).

Indeed, discourse structure is driven by the intentions of the speakers involved and influenced by the attentional states, namely the focus of the participants' attention, both changing over the course of time (cf. Grosz & Sidner, 1986). The modulation of the focus of attention by the speaker as the discourse unfolds is strictly connected with the notion of information packaging (see above): the speaker tailors an utterance in order to meet the assumed needs of the receiver (cf. Prince, 1981:224), with the goal of fulfilling his/her own intentions during the discourse. The fact that the speaker accommodates to the receiver involves the possibility of influencing the attentional states of the receiver according to the aims of the communication (cf. Bornkessel-Schlesewsky & Schumacher, 2016:585). The concept of attention orienting and the processing of information structure and status will be dealt with in more depth in Chapter 3 (3.6).

The concept of focus and background are relative to the packaging of information and the direction of attention for discourse goals. These notions are now briefly explained in the next subsection (2.2.1).

2.2.1 Focus and background

Sentences are usually partitioned in a more informative part and in a less informative one. To refer to this partition the literature has used the *focus/background* dichotomy. The term *focus* refers to the element of the sentence that expresses relevant communicative information (Lambrecht 1994; Uhmann, 1991; Krifka, 2007). This is in contrast to *background*, which generally is deduced by the context (Lambrecht, 1996). Rooth (1985, 1992) proposes the theory of alternatives, defining focus as the marking of elements to indicate that alternatives to these elements are relevant for the interpretation of the utterance. In (9) the questions of both examples represent the context for the answer (e.g., Uhmann, 1991; Krifka, 2007; Culicover & Rochemont, 1983) and the part in focus is the one that directly answers the question. The answers in this example are morphosyntactically identical and denote the same proposition but differ with regard to which argument is in focus.

(9) a. Q: What did Mary eat for lunch? A: [Mary ate]$_{\text{BACKGROUND}}$ [pasta]$_{\text{FOCUS}}$
b. Q: Who ate pasta? A: [Mary]$_{\text{FOCUS}}$ [ate pasta]$_{\text{BACKGROUND}}$

In (9a.) *pasta* is in focus, implying that there are alternatives to it that are relevant for the interpretation of the utterance; the same holds for *Mary* in (9b.), where *pasta* is part of the background. To better understand the theory of alternatives, one can think that in (9a.) the speaker is contrasting pasta with, for example, rice or soup. On the other hand, in (9b.) the speaker is saying that it was Mary who ate pasta and not somebody else.

A typological distinction that is assumed in the literature on intonation pertains to the domain of focus. Following Ladd (1980), if the domain of focus comprises only one word, as in (9a.) and (9b.) the focus is *narrow*, i.e. only one constituent corresponds to the focus of the sentence. The logic formalisation of the narrow focus is the following: there is at least an x, such that Mary ate x (x= *pasta*). Possible narrow focus types are contrastive focus, which entail an explicit contrast with a referent in the aforementioned sentence, and corrective focus, which entail a correction of the aforementioned sentence.

Examples of these two types of narrow focus are given in (10) and (11). (10) reports an example of contrastive focus, in which *movie* is opposed to the *tv series*. (11) reports an example of a corrective focus, where speaker B corrects the statement of speaker A (correcting *Mary* with *Lisa*). Contrary to narrow focus, there is the possibility of focusing more than one constituent in an utterance (or even the utterance as a whole), without pragmatically singling out any specific element of it (Ladd, 1980), as in (12). In this case the focus is called *broad.* The formalisation of this domain of focus corresponds to: there is at least an x, such that x happened (x= Mary ate pasta).

(10) Context: The possibility to choose between a movie or a TV series.

Utterance: I want to watch a [movie]$_{\text{FOCUS}}$.

(11) A: Mary went to the cinema with John last night.

B: [Lisa]$_{\text{FOCUS}}$ went to the cinema with John.

(12) Q: What happened? A: [Mary ate pasta]$_{\text{FOCUS}}$

Elements in focus and in background are also distinguished by their information status within the utterance, which influences their production (see 2.3). Crucially, there is no one-to-one mapping between the focus/background partition and the new/given distinction of information status. The next subsection will delineate some basic concepts relating to information status.

2.2.2 Information status

In (9), discussed in the previous subsection, the focus of the sentence overlaps with the part that the literature refers to as *new*, being not recoverable from the context; the background is referred to as *given*, since, in this case it is recoverable form the question (Halliday, 1967). The concepts of new and given information refer to the information status of the entities in the utterance, where the information status of an entity reflects its activation in the mental representation of the interlocutors. Information can be described as something new that is opposed to something that is already taken for granted or has been previously introduced in the discourse context (cf. Lambrecht, 1994:51). Thus, in discourse the entities have a different information status which unfolds dynamically from the evolution of the mutual beliefs of speaker and listener during the discourse (Grosz & Sidner, 1986).

In more recent studies, the dichotomy between given and new is no longer regarded as such, but is rather seen as a continuum in a scale that extends from the pole of maximally given to the one of maximally new (see Baumann & Riester, 2012; Chafe, 1994; Hajicova, 1993). This continuum refers to the cognitive accessibility of the information (Prince, 1981) and can be seen as a gradient notion of givenness. In fact, the concept of givenness is a very subtle one, since it does not only refer to the fact that an entity has been previously mentioned in the discourse, but there is the need of a more fine-grained description of the concept. The descriptions that have been provided by the literature are outlined in 2.2.3.

2.2.3 Concepts of givenness

Baumann and Riester (2012) report that in the literature the concept of givenness is explained relative to three cognitive dimensions: (i) *knowledge* shared by speaker and listener, (ii) the assumptions of the speaker relative to the listener's *consciousness* at the time of the utterance and (iii) what the speakers considers of *importance*. Regarding the first state, given information has been defined (among others) by Clark and Haviland (1977), Prince (1981), Clark and Marshall (1981) as known information that is not required to be active in the listener's mind at the time of the utterance. This knowledge is referred to with the terms *shared knowledge*, *assumed familiarity* or information that is *culturally present* (cf. Baumann & Riester, 2012). The second dimension is the one of consciousness and entails the presupposition by the speaker that the listener has a particular entity in his/her consciousness at the time of the utterance. This concept of givenness is particularly supported by Chafe (1976, 1994), who considers the idea of activation as fundamental in order to define this concept. In his approach, givenness is determined in terms of activation costs needed to convert a concept from a state of consciousness in the listener's mind into an active state (cf. Baumann & Riester, 2012; see also Gundel, Hedberg & Zacharski, 1993 and 2.2.1.). Another notion of givenness is the one by Kuno (1972) and implies the concept of *predictability* which corresponds to the ease of recovering a previously mentioned entity. The third possible definition of givenness involves the intentions of the speaker: a speaker can present an item as given or new with regard to what is *important* or not for s/he (e.g., Halliday, 1967; cf. Baumann & Riester, 2012).

Here, following Baumann and Riester (2012), the concept of givenness is considered to refer to the cognitive dimension shared by the interlocutors at the time of the utterance (cf. Baumann & Riester, 2012:4; in line with Chafe's proposal), since the concept of importance comprises a considerable amount of subjectivity and the dimension of knowledge is not subject to immediate contextual changes. Accordingly, givenness is regarded in terms of Chafe (1994:71), who defines the information status of an element in the discourse as induced by the assumption of the speaker of what is present in or immediately available to the listener's consciousness. Thus, an element's degree of givenness is derived by its degree of cognitive activation and it is equal to the speaker's effort to transfer it into an active state. Even though the categories between given and new are numerous (see Baumann & Riester, 2012), a simplification can be made considering a minimal extra category between these two poles: information that is accessible (Chafe, 1994:73). Hence, if at the time of the utterance an element is already active in the listeners consciousness, it is given. If it becomes activated from a previously inactive state, it is new. If a referent becomes activated from a previously semi-active state, it is accessible (see Figure 10). Accessible entities have not been directly mentioned in the discourse but are linked to ones present in the consciousness through lexical or conceptual association.

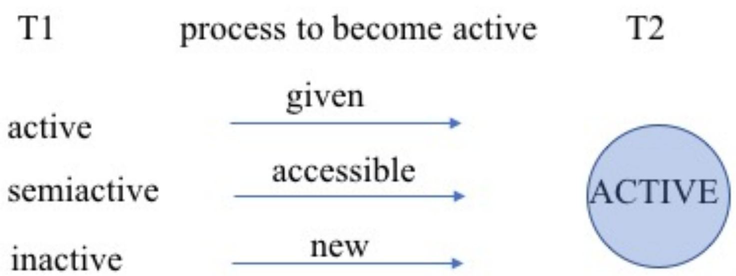


Figure 10. Chafe's degree of givenness. Figure adapted from Chafe (1994:73).

Numerous studies on West-Germanic languages have reported the existence of the distinction of the degrees of givenness (or accessibility) in the minds of speakers, reflected by the differentiation of the degrees of acoustic (prosodic) prominence assigned to entities in the uttering of the discourse (Baumann & Grice, 2006; Baumann & Riester, 2013; De Ruiter, 2015; Röhr & Baumann, 2010 among others). These concepts and relative studies are going to be delineated in 2.3. The next subsection will firstly refer to the concept of prominent information in the discourse, in order to better understand its relation with prosodic prominence.

2.2.4 Discourse prominent information

Determining the concept of discourse prominence is helpful in understanding the conditions underlying prosodic prominence (cf. Wagner, Breen, Flemming, Shattuck-Hufnagel, & Gibson, 2010:1). Von Heusinger and Schumacher (2019; but see also Himmelmann & Primus, 2015) define prominence as an organisational principle for discourse and characterise prominence in discourse as: (i) following the criterion of activation in the listener's mind, where the most active entities are most prominent (are singled out compared to the other discourse's entities of the same type and structure); (ii) shifting in time during the course of the discourse, i.e. prominence status of entities shifts in time; (iii) operating as structural attractor, i.e. prominent elements serve as anchors for the larger structures that they are part of, and licence more operations compared to other elements (cf. von Heusinger & Schumacher, 2019:119; see also Himmelmann & Primus, 2015).

The relation between prosodic prominence and prominent information from the discourse point of view is an inverse relation. Information that is more active in the speaker's and listener's mind is prosodically attenuated while information that is less active (that has not been mentioned recently) is less prominent in the discourse but more prominent prosodically. Studies concerned with discourse prominence consider the degrees of information accessibility as determining the importance (or prominence) of the information (cf. Luchkina, 2016; see Chafe, 1976, 1994; Calhoun, Nissim, Steedman, & Brenier, 2005). This is in line with the view of Information Theory (see Baumann & Riester 2012, 2013). In Figure 11 examples in A describe the relation between the activation in discourse of words that are new, accessible and given (underlined in the examples) and their prosodic prominence. Information that is the least active in the discourse (new; A.1 in Figure 11) is marked with higher prosodic prominence compared to accessible information (A.2 in Figure 11), which in turn is prosodically more prominent than the most active information in the discourse (A.3 in Figure 11).

Further factors of discourse prominence that have been argued to exert an influence on the prosodic prominence relations of an utterance are presented in the hierarchy of discourse prominence proposed by Wagner et al. (2010; cf. Luchkina, 2016), illustrated in the examples of Figure 11 B. Discourse new information which evokes alternatives (B.1. in Figure 11) is marked by speakers as more prominent than discourse given information (active in the speaker's and listener's mind) that evokes alternatives (B.2. in Figure 11), which in turn is marked by speakers as more prominent than discourse given information that does not evoke alternatives (B.3 in Figure 11). It should be noted that discourse given information can be activated in a contrastive way, as for the second occurrence of the word *scarf* in Example B.2. in Figure 11 (see Wagner et al., 2010:2 for a more detailed exemplification). Watson (2010) underlines that there are several factors correlating with prominence in discourse and should be viewed as an interaction of speaker-based and listeners-based components that are a reflection of the processes involved in the generation and perception of discourse. Wagner et al. (2010) found distinctions in prosody to mark the relative information prominence in the discourse. The subsequent section will discuss this issue.

Information status	Examples	Activation in the discourse	Prosodic prominence
A			
1. **New**	What did you buy? I bought a GUITAR.	**least active**	**most prominent**
2. **Accessible**	I wanted to bake pizza, but the oven was broken.		
3. **Given**	Our neighbour owns a dog. A dog is a really nice pet to have.	**most active**	**least prominent**
B			
1. **Discourse-new contrastive entities** (evoking alternatives)	Gramma has a lot of grandchildren, but she only gave a present to MARY.	**least active**	**most prominent**
2. **Discourse-given contrastive entities** (evoking alternatives)	Grandma gave presents to John and Mary. She gave a scarf and a sweater to John, only a scarf to Mary.		
3. **Discourse-given entities** (given entities not evoking alternatives; have the most mentions)	Grandma likes making presents. She gave a lot of presents to Mary.	**most active**	**least prominent**

Figure 11. A. Examples of new, accessible and given enities. B. Hierarchy of information prominence adapted from Luchkina (2016:2). Underlined words in example A and B are the ones to which the description in the first column refers to.

2.3 The prosodic marking of information structure and information status

Studies on West Germanic languages have shown that speakers employ prosody to mark the information structure of the utterance and the information status of elements within the utterance (see Baumann & Grice, 2006; Féry & Kügler, 2008; Pierrehumbert & Hirschberg, 1990). Degrees of prominence are used to signal both the level of focus and background, as well as the degree of accessibility of the elements of an utterance. Thus, prosodic features convey information about the attentional and intentional structure of the discourse (Hirschberg & Pierrehumbert, 1986; Pierrehumbert, 1980; Pierrehumbert & Hirschberg, 1990, among others) and are relevant to discourse and sentence processing, as will be discussed in Chapter 3 (see e.g., Arnold, 2008; Cutler et al., 1997; Dahan et al., 2002; Kjelgaard & Speer, 1999; Kraljic & Brennan, 2005; Nakatani, 1997; Terken & Nooteboom, 1987).

As explained in Chapter 1 (1.3), prosodically prominent units (e.g., syllables) are characterized as more prominent than adjacent units by virtue of their acoustic characteristics, such as major excursions in F0, increased duration or loudness, or, more generally, increased articulatory effort. Differences in prosodic prominence relations within an utterance typically reflect differences in the focal structure of the utterance. For example, an utterance with a nuclear accent occurring at the end, and presenting one or more pre-nuclear accents is likely to express a broad focus utterance. An utterance with an early nuclear accent is usually a narrow focus utterance (although narrow focus can also occur late in an utterance). The focal exponent of the utterance is associated with the nuclear accent of the sentence, often regarded to be the most prominent. Following the classical definition of the nuclear accent, this is the last accent on the IP (see 1.3.1.2). In utterances with a broad focus structure the last argument of the verb in one utterance always bears the nuclear accent. By contrast, in narrow focus utterances the nuclear accent can shift further to the beginning of the utterance, which in the syntactic literature has been called focus fronting (see Fanselow & Lenertova, 2011). Usually, this construction occurs for contrastive/corrective narrow focus as in the second sentence in the following example: *I was not the one who finished the chocolate. MARY finished the chocolate*. To convey an early narrow focus the region after the focussed word (henceforth, post-focal) must be attenuated. In German, words occurring in this region are considered to be deaccented (see 2.3.1). Subsection 2.3.2 will show that in Italian more variability has been reported for the prosodic realisation in the post-focal position. An exception in German is second occurrence focus, which will be addressed in section 2.3.1.

In addition to presence or absence of pitch accents, pitch accent position and pitch accent type can distinguish the domain of focus and the type of focus. Some studies report that pitch excursion differences are also used to signal focus (e.g., Braun, 2006; Grice et al., 2017; Liberman & Pierrehumbert, 1984; Pierrehumbert, 1980): accents presenting greater pitch excursions are used to distinguish contrastive narrow focus from non-contrastive narrow focus (cf. Roettger, 2017; see below). Moreover, a focussed constituent also presents temporal and spatial expansion that can additionally affect the unaccented syllables inside a word. In sum, focus is expressed through prosodic mechanisms that convey prominence: tonal events, phrasing, lengthening and strengthening of the segments within the focused constituent. In subsections 2.3.1 and 2.3.2 studies on German and Italian will be presented.

Early work (conducted mostly on West-Germanic languages) considered the relation between prosodic prominence and information status as binary, entailing that speakers signal the new and unpredictable information making it more prominent in comparison to shared and expected information, which is, by contrast, attenuated. However, the discussion above (2.2.3 and 2.2.4) showed that this binary distinction is not applicable to discourse prominence, which needs a more continuous description. At the same time, the straightforward mapping between pitch accent types and information structure or information status exemplified by early works (Brazil, Coulthard, & Johns, 1980; Halliday, 1967; Pierrehumbert & Hirschberg, 1990) has been supplemented by a probabilistic relation between pitch accent types and information structure and status (cf. Chodroff, Arthurs, Kurian, Pazol & Cole, 2019; see for English: Terken & Hirschberg, 1994; Bard & Aylett, 1999; Ito, Speer, & Beckman, 2004; for German: Baumann & Grice, 2006; Baumann & Riester, 2013; De Ruiter, 2015; Röhr & Baumann, 2010 and many others cited throughout this chapter). Moreover, prosodic marking is not only expressed in terms of pitch accents. Given that prominence is expressed through a bundle of phonetic parameters (see 1.3.), there is a highly probabilistic and complex relation between correlates of phonetic prominence and information structure and status (cf. Chodroff et al., 2019), which has been reported in studies on English and German. Namely, given information is not always realised as deaccented and new information can also be deaccented (Bard & Aylett, 1999; Ito et al., 2004; Riester & Piontek, 2015; Schweitzer, Riester, Walsh & Dogil, 2009; Terken & Hirschberg, 1994). Even though rising tones very often signal contrastive focus and newness, this does not always occur (cf. Chodroff & Cole, 2019:1).

Chodroff and Cole (2019) in a study on American English analysed words in nuclear position and found that it was more probable for given information to be unaccented or be conveyed through low pitch accents (L*) and more probable for contrastive information to be marked by high (H*) and rising accents (L+H*). The relation that they found was, however, not one-to-one: given items were also found to occur with high or rising accents, new and contrastive items were also found to be realised with low accents or without accent. Information structure was not only conveyed by pitch accents but also by a number of acoustic correlates: given items were more likely shorter, quieter, and realised with creaky voice during the duration of the word; contrastive information was more likely to be louder.

The studies mentioned so far show that the distinction in the information status of the elements is prosodically marked by different degrees of prominence. As many studies have shown, prominence signalled through prosodic means is a psychologically relevant property of discourse both to signal and perceive the level of activation of a referent (see, e.g., Watson et al., 2008; Mo, Cole & Lee, 2008, Kaland, Krahmer & Swerts, 2014; Baumann & Hadelich, 2003 and many others; see also 3.3). In addition, as will be discussed more in depth in Chapter 3, the probabilistic relation between prominence and information structure in production has important implications for the perception of the relations between prominence and information structure (cf. Chodroff et al., 2019).

An important clarification to be made is that, as seen in the brief explanation of discourse prominence in 2.2.4 (see Figure 11), given information can still be the focus of the sentence and evoke contrast. When this occurs, the given referent is also marked through prominence that conveys contrastiveness (Riester & Piontek, 2015). Indeed, accentuation does not only depend on the degree of activation (cf. Baumann & Grice, 2007:1641): speakers can decide to present a constituent as newsworthy and highlight it regardless of its degree of givenness. For example, when contrasting an element with another, also elements in the maximally given pole of the scale are highlighted with particularly prominent accents. This prosodic marking signals which of the elements are the focal exponents of the utterance. As pointed out by Riester and Piontek (2015) the claim that is often made for English and German (and other West-Germanic languages), that subsequent mention of a referent has to be deaccented is a misinterpretation of the concept of givenness. The authors emphasize that given elements can in some cases answer the question under discussion, as in the case of *tea* in the following question-answer pair: *Would you like tea or coffee? I would like tea.* Moreover, when realised in prenuclear position (occurring before the nuclear accent and before the focus exponent), given elements very often carry pitch accents (Féry & Kügler, 2008; Baumann & Riester, 2013).

Baumann and Riester (2012) argue that for an adequate analysis of the information status of an item and its consequent prosodic realisation, two levels of givenness have to be taken into account: a referential level and a lexical level. This necessity appears clear when considering Example (13) and (14), taken from Büring (2007; Example 7 adapted from Ladd, 1980).

(13) A: *Did you see Dr. Cremer to get your root canal?*
B: *Don't remind me. I'd like to STRANgle the butcher.*

(14) A: *Why do you study Italian?*
B: *I'm MArried to an Italian.*

In (13) and (14) the types of givenness involved in *the butcher* and *Italian* are different but the prosodic realisation can be hypothesised to be the same. In (13) *the butcher* is lexically new (is not mentioned in the question), but can be realised as deaccented because it is interpreted as referring to the entity that has previously been mentioned (*Dr. Cremer*). In (14) *Italian* can be realised as deaccented because it is lexically given (mentioned in the question), even though the referent is not mentioned in the previous context: *Italian* in the question refers to the language, in the answer to the wife. At least in German, when empirically testing this hypothesis, the picture is more complex (see Baumann & Riester, 2013). Nonetheless, the lexical and referential level prove to have an impact on the degree of prominence of the element:

> […] an item's information status *at each level separately* has an incremental effect on the degree of prosodic prominence assigned to the item. That is, referentially and lexically new expressions are most prominent, and referentially and lexically given expressions are least prominent […]. (Baumann & Riester, 2013:33).

In addition, data on German has shown that the difference between the prosodic realisation of elements that are lexically new but referentially given (as in (13)), and lexically given but referentially new (as in (14)) is mostly levelled out, even though a tendency of the lexical level to predominantly influence prosodic realisation is registered (see Baumann & Riester, 2013). Authors also argue that not only the categories of given and new have to be distinguished, but also intermediate classes of these categories need to be identified and show a stepwise increase in prosodic prominence going from given to new (see also Baumann & Grice, 2006 and Chafe, 1994).

The statistical predictability of an information in the discourse also influences its prosodic realisation. Words statistically predictable from the preceding context have a higher probability of being realised with shorter duration (Bell et al., 1999; Gregory et al., 1999; see also Jurafsky, Bell, Gregory & Raymond, 2001) and without a pitch accent (Watson, Arnold & Tanenhouse, 2008). Moreover, Aylett & Turk (2004), and Cole, Mo, & Hasegawa-Johnson (2010) argue that lexical frequency reliably triggers variation in acoustic-prosodic expression. Aylett & Turk (2004) propose a model in which acoustic redundancy compensates for the low level of redundancy in the language (e.g. low frequency words) and they also argue that languages with looser constraints on word ordering can use, to some extent, position as a substitute to acoustic prominence. For example, instead of presenting an attenuation of repeated mentions, these languages could place these repeated mentions at the beginning of a phrase where they are less predictable from the context (cf. Aylett & Turk, 2004:54).

The distribution of probabilities in the prosodic marking of information status and information structure are language-dependent. The marking of information structure and information status in German and Italian is of interest for the perception experiment in Chapter 6. The two languages are claimed to present differences in this respect, and these differences may have an impact on the perception of prominence, as will be made clearer in Chapter 3. Therefore, a discussion of studies and results found in the two languages will be the subject of the next subsections (2.3.1, 2.3.2 and 2.3.3). In particular, the divergences that are of interest here are the ones concerning the marking of background and givenness and of the post-focal region.

2.3.1 German

In German[9], pitch accents are assumed to be the primary correlate of information status (cf. Schweitzer et al., 2009; see Halliday, 1967). The H* accent is considered to be the standard marking of newness, and H+L* accent the standard marking of discourse new but to some extent expected information, whereas deaccentuation is considered to be the most commonly occurring strategy to mark given information (Baumann, 2006). However, as discussed at the beginning of this section, the variability is high and these associations have to be considered as probabilistic preferences rather than a fixed

[9] Note that the studies on German mostly involve speakers from north of the Benrather isogloss, speaking Standard German.

description of what speakers produce. For example, Baumann, indicates that H+L* is also found as a marker of givenness. Indeed, the relation between given-new and prosodic marking has to be more fine-grained to better describe the preferred associations. Analysing a corpus of read speech (taken from the economics section of a newspaper and read by a speaker of standard German) Baumann and Grice (2006) observed that H+L* is used by speakers to indicate information that is partially given, or accessible.

Schweitzer et al. (2009), adopted a more specific classification of the degrees of accessibility of the entities (following Riester, 2008) to label a corpus of German Radio News. Among the category of given information, they found that deaccentuation occurred in 70.73% of pronouns, 28.57% of short forms of expressions already mentioned in the discourse, 34.41% of expressions that are given at the referential level but are new at the lexical level, and 41.18% of exact repetitions (although the corpus contained few cases of short forms of mentioned expressions and exact repetitions).

Röhr and Baumann (2010) conducted a reading study labelling information status as new, given, textually accessible and inferentially accessible. Their results show that new referents were marked with H* and L+H*. By contrast, when increasing their degree of givenness, the probability of finding L* accents, early peak accents (H+L* and H+!H*) and deaccentuation increased. Figure 12 shows the stepwise change from lack of pitch movement to falling movements and from falling movements to rising movements while moving along the continuum from given to new entities.

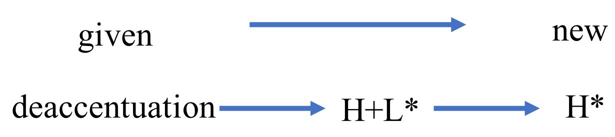


Figure 12. Proposed mapping between information status and pitch accent types in German. Adapted from Baumann & Grice (2006:1655).

However, as expected, they also found variability in the data, with accessible referents marked also by H* accents and deaccentuation, and new information marked with falling accents (early peaks) 20% of the time. A study similar to the one by Röhr and Baumann (2010), involving the same annotations of the degrees of givenness, but analysing spontaneous speech rather than read speech, was carried out by De Ruiter (2015). This study reports that new referents were always accented: L*+H accents occurred 25.5% of the time, H* 21.6% of the time and L+H* 17.6% of the time. When the referent was accessible, its realisation was mainly associated with L+H* and L*+H accents. Given referents were deaccented 32.6% of the time. Röhr and Baumann, together with other colleagues, conducted further experiments involving spontaneous speech (Baumann & Riester, 2013; Röhr, Thies, Baumann & Grice, 2016 and Thies, Röhr, Baumann & Grice, 2018). They annotated the information status of referents based on the RefLex scheme proposed by Baumann and Riester (2012), thus marking the information status at the referential and lexical level (note that Thies et al. adopted a simplified version only marking the referential level). The results found by all these studies indicate a tendency towards a stepwise increase in prominence: given < accessible < new, both at lexical and referential levels.

In addition to information status, many studies have also investigated information structure, or focus-background structure. These studies have confirmed that speakers make use of a large variety of prominence-lending cues to encode differences in type and size of focus and to mark given elements that are part of the background (Baumann, Becker, Grice & Mücke, 2007; Mücke & Grice, 2014; Roessig & Mücke, 2019). Mücke and Grice (2014) analysed words in three focal conditions, namely broad focus, narrow focus and contrastive narrow focus, and words occurring in background, after the focus (post-focal position). As in the studies on information status, this study found preferences for particular accent types in each focal condition: speakers predominantly used early peak falling pitch accents to mark broad focus (H+!H*), while contrastive focus was predominantly marked by late peak rising pitch accents (H* and L+H*). Nonetheless, they confirmed the absence of a one-to-one relation between focus and accent type and found different strategies for different speakers (as later pointed out by Grice et al., 2017 for the same dataset). Words occurring in the background were by contrast never accented. However, they found a clear differentiation between the duration of the stressed syllable in the background compared to contrastive and narrow focus, but the systematic difference in duration was not present in the comparison with stressed syllables occurring in broad focus (with only one speaker making the differentiation in this parameter). Moreover, stressed syllables in broad focus always involved the presence of a pitch accent, which differentiated them from the ones in the background, but did not present systematic differences in the supralaryngeal parameters.

Roessig and Mücke (2019) have further confirmed that speakers use a combination of laryngeal and supra-laryngeal cues to enhance prosodic prominence in different focal conditions, corroborating the presence of modifications not only between words in the background and words that are focus exponents, but also within this latter category. In addition, they found clearer modifications of the cues between background and broad focus, attributed to the fact that the dataset of this study was larger than the one of the previous study, thus allowing for more robust results, less influenced by individual variability. Specifically, they found for background information a distribution of onglide values (the F0 movement towards the tonal target on the stressed syllable, see 1.2.4.1) around zero (indicating the absence of tonal movement). Conversely for broad focus the results revealed almost an equal presence of falling and rising onglides. For narrow and contrastive focus, they reported onglide values indicating a prevalence of rising accents, with a progressively larger excursion going from the rising accents in broad focus to narrow focus to contrastive focus. In the articulatory domain (see sonority expansion in 1.3.2), for the production of the low vowel /a/, the authors found a clear increase in the aperture of the lips from background to broad focus condition. This strategy was also adopted to distinguish broad from narrow and contrastive focus and narrow from contrastive focus. However, the difference between background and broad focus was the greatest. The same hierarchy can be found for the lowest tongue body position: background > broad focus > narrow focus > contrastive focus (opposite to the lip aperture because when the tongue body is lowered, the values decreased and become negative; see Roessig & Mücke, 2019:9 and recall that these results pertain to the vowel /a/).

An important distinction has to be made in the realisation of background as occurring after or before the focus of the utterance. Studies on German have shown that background information occurring before the focus of the utterance is mostly realised with pre-nuclear accents. These have a prominence status that is lower than the nuclear accents, but are nonetheless fully-fledged accents (see 1.3.1.2). Background information occurring after the focus of the utterance in German is usually realised as deaccented, but deaccentuation is not always present. In second occurrence focus, for example, stressed syllables of informationally focussed words in the same IP of a previous focus have been shown to lack a pitch accent, but to have increased duration and increased spatial modifications of the opening and closing gestures of the lips in comparison to stressed syllables of words occurring in background, non-focussed conditions (Baumann, 2016; Baumann, Mücke & Becker, 2010).

Another alternative to deaccentuation is the occurrence of pitch movements categorized as phrase accents (Grice, Ladd & Arvaniti, 2000), which are different from pitch accents. Phrase accents have been described as edge tones with a primary association to a prosodic edge and a secondary association that can be either to a stressed or to an unstressed syllable (depending on the language or variety considered). Phrase accents accordingly describe the contour between the nuclear accent and the boundary tone. Grice et al. based their argument on the analysis of the intonation of questions of Eastern European languages that contain a low pitch accent (L*) on the focus exponent followed by a high tone (H-) before a low final boundary tone (L%), as presented in Figure 13, adapted form Grice et al. (2000).

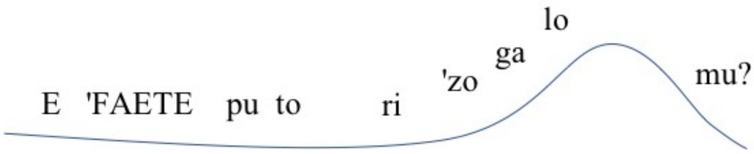


Figure 13. Stylized contour of the question 'HAVE YOU EATEN some of my rice pudding?' uttered in Cypriot Greek *E 'FAETE pu to ri 'zogalomu?* Picture adapted from Grice et al., 2000:154. The utterance is realised with the focus on the first constituent (E FAETE). The following constituents therefore occur in post-focal position. Note that in this variety the rising movement is not associated with a stressed syllable (stressed syllables are indicated with ' before the syllable of interest).

In post-focal position in German questions, a fall is produced after a L+H* contrastive accent associated with the stressed syllable of the focal word. Low pitch then extends until the end of the utterance where a rise is produced to indicate sentence modality. Contrary to questions in East European languages, the low phrase accent in German questions does not seem to necessarily be stress seeking. Figure 14 illustrates this behaviour, where the rise and the fall are realised on the focal word (*Wohnungen*, apartments) and the low target reached by the fall stretches throughout the post-focal domain until the final rise signalling the modality (see also 2.3.3.1).

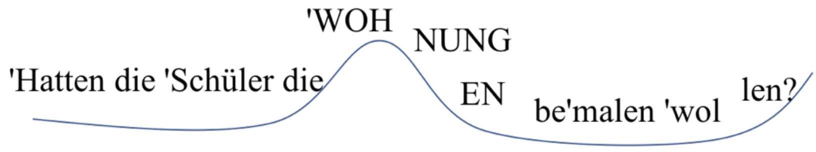


Figure 14. Stylized contour of the question 'Had the pupils wanted to decorate the FLATS?' uttered in German *'Hatten die 'Schüler die 'Wohnungen be'malen 'wollen?* Picture adapted from Grice et al., 2000:166. The utterance is realised with the focus on WOHNUNGEN, the following constituents therefore occur in post-focal position. Stressed syllables are indicated with ' before the syllable of interest).

Grice et al. (2000) argue that, differently from pitch accents, these secondary associations of phrase accents do not signal focus and do not signal prominence to the same extent as pitch accents. The situation seems to be different in questions of some varieties of Italian, as explained in 2.3.3. In fact, in German the falling-rising contour after the nuclear accent comprises a post-focal falling movement (directly realised after the nuclear accent) and a subsequent rising boundary tone, which can associate with a stressed syllable but does not have to (see 2.3.3). By contrast, the rising-falling-rising-falling contour occurring in some varieties of Italian comprises a post-focal falling movement directly realised after the rise on the nuclear accent, a rise that systematically associates with a post-focal stressed syllable and a subsequent fall. This latter process seems more likely to be linked to a downstepped pitch accent, since it also seems to present cues to prominence other than F0 movement (see 2.3.2).

Kügler and Féry (2017) argue that German, too, can lack deaccentuation in post-focal position. The accents in this position would not be suppressed, only downstepped. In their study, they analysed statements and showed downstepped relations occurring among the two (or three) constituents after the focus. They reported these pitch movements to be above the threshold for perception and thus to be perceived as having a certain degree of prominence. Nevertheless, they did not conduct a perception test and did not compare the prominence of these accents in relation to other material occurring in post-focal position. Moreover, no other cues to prominence were investigated in the experiment. The authors argued that the asymmetry found between the pre-focal and post-focal regions is such that the pre-focal accents are just slightly compressed whereas post-focal accents are strongly compressed. This would be a consequence of the fact that the nuclear accent has to be the last strongly prominent in the IP: the compression of the post-focal part would then enhance the effect of finality of a nuclear pitch accent (see also Féry & Kügler, 2008). By contrast, slight compression of the pre-focal material would be related merely to the degree of givenness.

The experiments reported in Chapter 5, Chapter 6 and Chapter 7 will investigate prominence cues in post-focal position in Italian (with a method that would be interesting to apply to German data as well, see 1.2.5). The possibility for German to have post-focal pitch accents and prominences in the post-focal position is left open here. Nonetheless, the next chapter will show that the post-focal region in German is usually perceived with a low level of prominence and, if these pitch accents occur, they might not convey prominence at all, i.e., listeners might not perceive the words with these pitch accents as prominent. However, this thesis will neither confirm nor deny the presence of post-focal compressed accents in German, since the question as to whether there is a difference between Italian and German is beyond the scope of the research undertaken.

2.3.2 Italian

There are only very few studies in the Italian literature on the prosodic realisation of information status that thoroughly consider the prosodic marking of different degrees of accessibility. There is only one study that investigates the use of pitch accents in relation to the tripartite distinction of new, given and accessible[10] information and is concerned only with questions (realised in the Bari variety). This study is the one by Grice and Savino (2003), which analysed semi-spontaneous dialogues elicited through MapTasks (Anderson et al, 1991). Results showed that when speakers thought that the information was mutually inactive (new), the nuclear accent in questions was L+H*, while when they believed the information to be mutually active (i.e., the question was a confident confirmation seeking one) the nuclear accents were realised by H*+L or H+L* accents.

Other attempts to examine the prosodic realisation of information status in Italian, this time only marking information status of referents either as new or given (but not as accessible), have been conducted on radio broadcast speech by Avesani (1997) and on (semi-)spontaneous dialogues by Avesani and Vayra (2005), Swerts et al. (2002) and Avesani et al. (2015). The two latter studies have investigated the production of noun phrases (NP) in a card game involving noun-adjective pairs (e.g., *triangolo rosso*, lit.

[10] With accessible information in turn distinguishing between textually accessible, inferentially accessible and situationally accessible.

triangle red). However, only Avesani et al. (2015) have reported results in terms of pitch accent types. These results indicate that the majority of new referents are marked with H*, H+L* and L+H* accents. Given referents were unaccented 33% of the time when occurring in prenuclear position (i.e., when the given element was the noun), while they were always accented when occurring in final position (i.e., when the given element was the adjective). Accentuation was present also when the given element occurred after a focal accent. This result is in line with Swerts et al. (2002) who have also found that the final word of the NP in Italian was always accented, even when realised after a contrastive focus. In addition, the two studies compared the production of Italian speakers with Dutch and German speakers. They found a different distribution of deaccentuation among given referents between Italian and Dutch (Swerts et al., 2002) and Italian and German (Avesani et al., 2015). In these studies, lexically and referentially given elements in Italian occurring in final position are reported to be always accented. Conversely, Dutch (Swerts et al., 2002) and German (Avesani et al., 2015) constantly realised these words as deaccented, both as a consequence of their givenness and of their position (after the focussed element). These findings are in line with Ladd (1996:176-177), who argues that Italian exhibits a preference towards the absence of deaccentuation. Nonetheless, these studies report some problems. In Swerts et al.'s study, Italian speakers produce a hat pattern stretching over the entire NP for all the conditions, thus failing to signal the differences among information status. Yet, the analysis of continuous parameters could possibly show that the conditions are differentiated in terms of peak alignment or pitch excursion. Indeed, Avesani et al. (2015), although reporting for Italian a 100% rate of accentuation of the given element of the noun phrase in the last position of the phrase, found that these words were marked by different accent types, which can reveal a differentiation among the conditions, about which the study is, nevertheless, unclear.

Results reported by Avesani (1997) and Avesani and Vayra (2005) refer to sentence-length utterances. Avesani (1997) found for an Italian data set that clauses and entire NPs in post-focal position (as part of the background after the focal accent) were produced without a pitch movement, resembling the pattern for English and German. However, she found that elements that belong to the same discourse segment when identically repeated can be reaccented even when occurring in the same grammatical function and irrespectively of their position in the utterance. In Avesani and Vayra (2005) speakers of the variety of Italian spoken in Rome participated in a spot-the-differences game. The authors coded the referring expressions using a given-new dichotomy based on two definitions of givenness (in order to explore potential differences between the two codings): givenness as shared knowledge (following Prince, 1981) and givenness as the dynamic unfolding of mutual beliefs during the discourse (following Grosz & Sidner, 1986). The authors found that for both codings, the percentage of given referents realised without a pitch accent was very low, especially when compared with the percentage found for comparable studies in English. Indeed, the percentage of decaccented given referents was 6.5% in Italian in contrast with 20% found by Bard and Aylett (1999) in English. The authors conclude that the informational status of entities is not related to deaccenting and that speakers do not mark as given the repeated mentions of the entity in the same discourse segment. These results are in line with both the previous study by Swerts et al. (2002) and with the subsequent study by Avesani et al. (2015) who also found a difference in the distribution of accents in the given elements between Italian on one side and Dutch and German on the other. However, remember that the assignment of given vs. new category in Avesani and Vayra (2005) was binary. A more fine-grained analysis might have revealed different patterns. Nonetheless, the lower percentage of deaccentuations found in Italian with respect to other comparable studies on West-Germanic languages at least suggests that in Italian the relation between givenness and deaccentuation is different.

None of the studies mentioned in this subsection take into account continuous parameters. Nonetheless, they still reveal that for Dutch and German the mapping between given elements in post-focal position and deaccentuation is more straightforward, and that Italian allows more variability. This higher variability is further supported by the findings of a comparable study investigating NPs in English. This study reported that 61% of the occurrences of the second element within the NP when given in the discourse segment were deaccented, whereas 50% of these occurrences were deaccented when given in the larger domain of the whole discourse[11] (Ito, Speer & Beckman, 2004). Hence, in comparing studies on Italian and studies on German, prosodic realization in German is

[11] In this experiment participants had to collaborate in order to decorate a tree. The domain of discourse and discourse segments were defined in relation to the task in which participants were involved. The discourse domain was considered to be represented by the interactions between participants during the decoration of a whole tree. Discourse segments were defined in relation to the completion of a subsection of the tree, thus in relation to the achievement of a local communicative goal.

more straight-forward (see the abovementioned studies of Schweitzer et al., 2009; De Ruiter, 2015; Röhr & Baumann, 2010).

These results could be ascribed to the fact that in Italian there is a different mapping between given constituents and deaccentuation. Moreover, the tendency of Italian native speakers to keep accenting the given post-focal elements is found also in Italian native speakers speaking German as L2 (Avesani et al., 2015 and Dahmen, 2013 for Italian speaking German as L2). In addition, Dahmen reports the persistence of this tendency for the L2 speakers also after a training in German prosody.

As far as the marking of focus and background are concerned, there are several studies conducted in a number of varieties of Italian. An overview of the behaviour in different varieties can be found in Gili-Fivela et al. (2015). For all varieties investigated the authors report that the nuclear pattern found in broad focus statements is H+L* L-L%, with the high leading tone reported to be highly variable. The pattern used in marking non-contrastive narrow (informative) focus occurring at the end of an utterance in the variety spoken in Florence is reported to be the same as broad focus (Avesani & Vayra, 2003), while in other varieties (spoken in Napoli and Palermo) the difference between the two is reported to be signalled through different pitch accent types (D'Imperio, 2003; D'Imperio, 1997; Grice, 1995). Narrow contrastive focus is reported to be either realised with the broad focus pattern which alternates with the pattern that is typically realised for contrastive-corrective focus and is reported to be: L+H* for Florence and Siena, H*+L in Pisa, Bari and Lecce (and in Pescara as a secondary option), and ¡H+L* (upstep-fall) in Pescara (cf. Gili-Fivela et al., 2015). Contrary to German, for Italian there are no articulatory studies conducted on the distinction of different sizes and types of focus. However, there are studies attesting that also in Italian post-focal background is realised with a flat and low contour, as reported for German (Bocci & Avesani, 2011). In addition, as well as for German, studies on Italian argue for the lack of deaccentuation of this post-focal region. Nonetheless, in some varieties of Italian the compression for questions is different from statements. Indeed, some studies report the presence of a post-focal rise in pitch for questions (even if compressed). The presence of this post-focal rise is ascribed to its function of conveying sentence modality (question). This situation will be elaborated in the following paragraphs. To understand the occurrences of these rises in post-focal position, a brief overview of the question nuclear contour in the varieties of Italian spoken in Naples, Palermo and Bari is provided in the next subsection.

2.3.3 Prosodic realisation of post-focal region in different modalities

Questions in Italian have traditionally been regarded in terms of a geographical division between Northern and Central varieties on one side and Southern varieties on the other (cf. Savino, 2013:80). Whereas yes/no questions in most Central and Northern varieties of Italian are reported to be more frequently characterised by a terminal rise (Avesani, 1995; Marotta & Sorianello, 1999), the nuclear accented syllables of questions in Southern varieties of Italian are reported to be characterised by a rise followed by a fall (cf. D'Imperio, 2002:1; see also Grice et al., 2005; Savino & Girce, 2011). Recently, Savino (2013) contributed to other studies (Endo & Bertinetto, 1997; Gili-Fivela, 2003; Marotta & Sorianello, 2001) challenging this traditional division by the comparative analysis she conducted on 15 varieties of Italian (spoken in Turin, Bergamo/Brescia, Milan, Venice, Genoa, Parma, Florence, Perugia, Rome, Cagliari, Naples, Bari, Lecce, Catanzaro and Palermo) based on Map Task dialogues (Anderson et al., 1991) contained in the CLIPS corpus (*Corpora e Lessici di Italiano Parlato e Scritto*, Corpora and Lexicons of Spoken and Written Italian, downloadable at www.clips.unina.it). Results of her analysis show that the rising-falling accent to mark questions, usually considered as characteristic only of some Southern varieties, is in fact widespread also in the majority of varieties of Italian (Turin, Venice, Parma, Naples, Bari, Catanzaro and Palermo), independently of whether they are Southern or Northern. In the varieties mentioned, yes/no questions are mostly realised with an accentual rise (either L+H* or L*+H), indicating questions, followed by a falling boundary (L-L%; see also D'Imperio, 2002; Grice et al., 2005 and Grice, 1995) and only occasionally by a final rise (H%). In the remaining three varieties (Genoa, Rome, and Florence) the accentual rise is followed equally by either a low or a high edge tone. By contrast, in the varieties of Milan and Bergamo the tune of questions systematically has a falling-rising shape (H+L* L-H%).

In the rise-fall question contour the peak of the pitch is considered to be part of the rising pitch accent (the high tone in the L*+H o L+H* accent), which associates with the head of the IP and the following fall is represented as a low edge tone (cf. Roettger, 2017:133). Figure 15 shows the rising-falling functional contour of the question *Si deve prendere l'aereo?* (Does one have to catch a plane?), uttered in the Bari variety of Italian. The question is produced in broad focus and the sentence modality is signalled by a great excursion in pitch associated with the word *aereo* (plane).

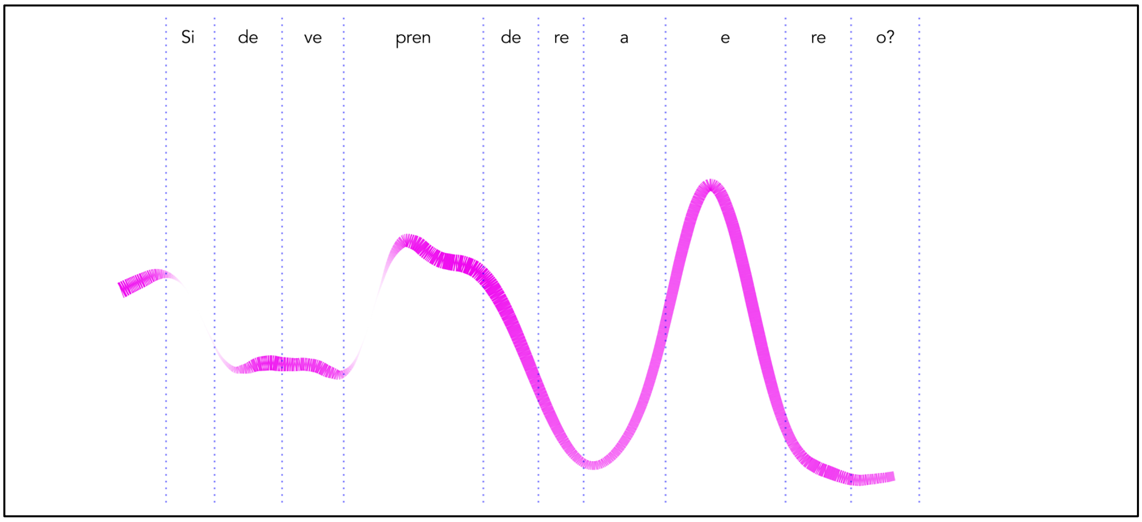


Figure 15. Periogram of the question 'Does one have to catch a plane?' uttered in the variety of Italian spoken in Bari *Si deve prendere l'aereo?* The question is realised in broad focus. The periogram shows that there is a pitch movement (rising-falling) associated with the stressed syllable of the last noun (*mon*).

For yes/no questions that present an early focal accent, studies on the Southern varieties of Italian have reported that after the nuclear accent the contour is not flat and low. Instead, the functional rise (indicating sentence modality) in post-focal position is not subject to suppression, as shown in Figure 16. This configuration of the contour is considered by D'Imperio (2002) similar to the one in Swedish, in which the nuclear accent does not have to be the last accent occurring in the phrase. Indeed, early focal questions such as the one presented in Figure 16, show a peak on the last stressed syllable of the IP, occurring after the preceding focal accent (see D'Imperio, 2002:7 for other examples in the Neapolitan variety;see also Cangemi & D'Imperio, 2013; Grice, 1995; Grice & Savino, 2003; Grice et al., 2005; Savino, 2012; Gili Fivela et al., 2015). For Neapolitan, D'Imperio (2001) does not consider this a fully-fledged pitch accent and labels it as !H* (downstepped H accent).

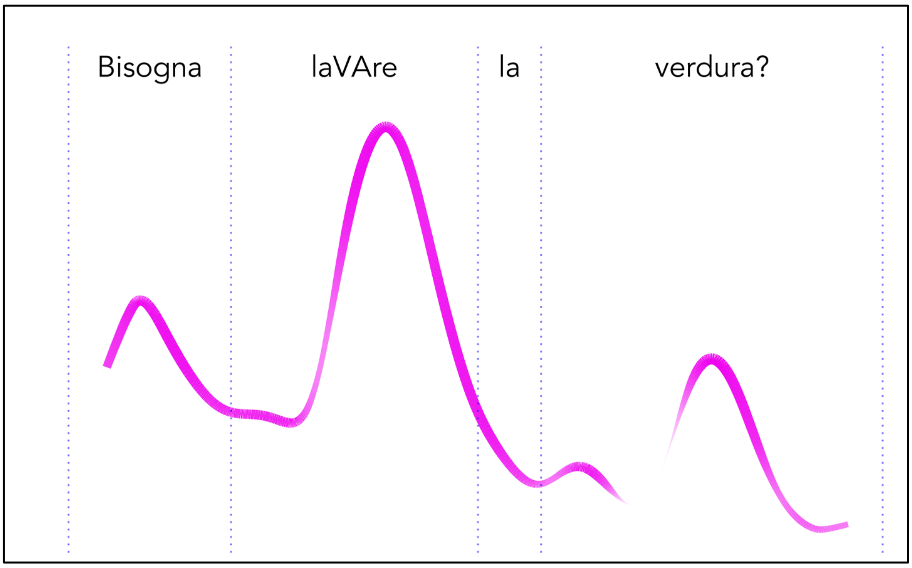


Figure 16. Periogram of the question Should one wash the vegetables? uttered in the Bari variety of Italian, with an early narrow focus on the verb *lavare*.

The occurrence of rising movement to mark questions is similar to the phrase accent described for Eastern European languages by Grice et al. (2000; see above). However, in Italian varieties, it is always associated with a stressed syllable and occurs after a very prominent movement associated with the focus (see Figure 16). By contrast, in the examples of Grice et al. (2000) in Eastern European languages the focus was marked with low pitch accent rather than a rising one. This characteristic of Italian is of interest in the discussion on attention orienting in Chapter 3, which will present the rationale for the experiment in Chapter 7. Among the varieties collected in the CLIPS corpus and analysed by Savino (2013), the variety spoken in Udine is not present. Next subsection (2.3.3.1) will present a comparative study conducted on semi-spontaneous speech in the varieties of Bari and Udine, with the aim of attesting whether there are differences between the two varieties in the production of questions, focusing especially on the contour of the questions after an early focus (i.e. the part in post-focal position).

While there is consensus on the presence of a post-focal pitch accent in questions in some varieties of Italian, the picture regarding the accentutation/deaccentuation of declaratives is less clear, as delineated in the previous sections (2.3.1 and 2.3.2). While Swerts et al. (2002), and Avesani et al. (2015) report a lack of deaccentuation in post-focal position within NPs, other studies show the absence of F0 movements in post-focal position of sentence-length utterances (Avesani & Vayra, 2005; Bocci & Avesani, 2011; D'Imperio, 1997). Nonetheless, Bocci and Avesani (2011) argue for the presence in (Tuscan) Italian of a pitch accent (L*) in the post-focal position of declaratives. They report that a given element (part of the background and exact repetition of the element in the question), when it is the metrical head of an intermediate phrase (thus occupying a metrically strong position), shows prominence features. It exhibits enhanced spectral emphasis, more extreme F1 trajectories and longer duration compared to a new item occurring in broad focus utterances and compared to an item in post-focal position, but not the metrical head of the intermediate phrase. Based on these observations, the authors claim that Italian allows words in post-focal position to bear prominence not reducible to the effects of (post-) lexical stress.

In addition, D'Imperio (2002) indicates that in Italian declaratives, accents in the post-focal region are not subject to the automatic suppression of accentuation. The author explains that the post-focal accents in statements would be a downstepped version of an H+L* accent and this would explain why the accent is almost undetectable in most of the cases (D'Imperio, 2002:17). She argues that considering the presence of a post-focal accent for Italian appears to be supported by observations such as the one in Figure 17, which shows a stylised contour of the utterance *Vedrai NONO e mamma* (You will see NINTH and mom), where *NONO* is in focus. This figure is adapted from D'Imperio (2002). The contour in the post-focal region is not completely flat. In the analysis conducted by D'Imperio, the pitch in this position (on the word mamma) falls from 163.5 Hz to 142 Hz. The author proposes to transcribe this fall as a !H+L* accent (downstepped H+L* post-focal accent).

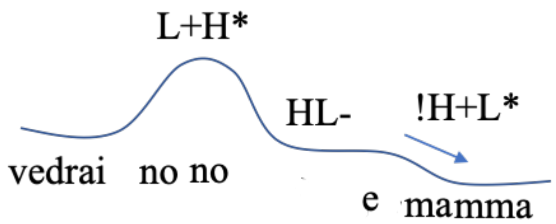


Figure 17. F0 curve, word and tone labels of the statement *Vedrai [NONO] e mamma* (You'll see [NINTH] and mom), with narrow focus on *nono*. Image adapted from D'Imperio (2002:18).

D'Imperio argues that the finding for Italian gives rise to two issues. The first concerns the definition of the concept of accent in Italian. The second concerns whether pitch excursion is a part of the definition of an accent or whether duration and intensity suffice to define accentuation (see the discussion in 1.3.2.). Although this thesis will not be able to answer these far-reaching questions, both the production and the perception of the post-focal region will be investigated, with the aim of gaining a more fine-grained understanding of prominence in Italian and its connection with information status and information structure. Crucially, the findings of D'Imperio (2000a) suggest that in the perceptual domain prominence is not only conveyed by F0, in line with the discussion in Chapter 1 of this thesis (1.3.2) and with the discussion that will be conducted in Chapter 3 (3.2.1).

2.3.3.1 Comparison between Udine and Bari in the production of yes/no questions

Literature on question realisation of the variety of Italian spoken in Udine is very scarce. The only studies investigating it are the ones by Canepari (1980) and Romano and Miotti (2008), which investigate read speech and report the presence of questions with a rising final contour. However, the study of Canepari (1980) reports that in all varieties yes/no questions are only realised with a final rising contour, which has been shown not to be the case in (semi-)spontaneous speech (Savino, 2013). Contrary to the Italian variety spoken in the region of Udine, a more in-depth investigation has been conducted on Friulian, another language spoken in the region of Udine (D'Agostin & Romano, 2007; Finco, 2007; Roseano, Vanrell & Prieto, 2015). Among these studies, in Roseano et al. the presence of both final rising contours and functional rising-falling accents to signal questions is reported for semi-spontaneous speech. Also D'Agostin & Romano's (2007) study seems to indicate the possibility of the presence of a rising-falling contour to indicate questions in the Udine variety of Italian, although the authors are not clear on this. However, results of both studies hint at the possible presence of both strategies to indicate questions in the variety of Italian spoken in the region.

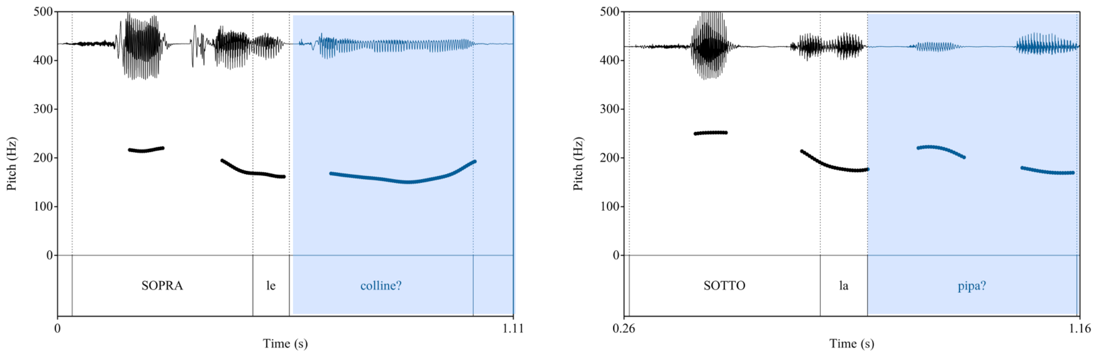
To compare the production of questions, specifically of the post-focal region of questions, in the varieties of Bari and Udine, I have analysed two comparable corpora of semi-spontaneous speech of the two varieties, collected with the Map Task technique (Anderson et al., 1991). The corpus of Bari that has been used is part of the CLIPS collection (and was already analysed by Savino 2013; therefore, the present experiment needs to be considered as a replication experiment), while the corpus of Udine was collected separately. For collecting the corpora, speakers of the Bari variety (16 speakers, overall duration of 2 hours and 19 minutes) and speakers of the Udine variety (22 speakers, 5 hours and 25 minutes) were paired and instructed to interact with the aim of reproducing as accurately as possible the route on the partner's map (see also Savino, 2013 for a more detailed explanation on the procedure). The corpora were analysed following the AM framework (see 1.2.1). The analysis of Bari data was conducted referring mainly to Savino (2013), whereas the analysis of Udine data was conducted taking into consideration the existing description of Friulian (Roseano et al., 2015). Despite referring to Friulian and not to the variety of Italian, this description was substantially retained. This suggests that the intonation of yes-no questions does not extensively differ between the variety of Italian spoken in Udine and Friulian (see also D'Agostin & Romano, 2007).

Results of the replication of the study of the Bari variety confirms that the majority of the questions are realised with rising-falling contours (L+H* L-L% in the 84.27 % of the cases). The analysis of the Udine corpus has showed that the most common pattern is the (falling-)rising one (H+L* L-H% in the 46.98% of the cases, L* L-H% in the 17.93% of the cases). However, the presence of the rising-falling shape (L+H* L-L%) is also attested, although the occurrence of this pattern is much lower than in the Bari dataset, corresponding to 16.53% of the cases. In addition, 15.93% of the rising-falling accents in the Udine variety were followed by an extra rise at the end of the utterance, whereas in the Bari variety the percentage was lower, corresponding to 7.3%. In the post-focal region, the difference between the two varieties becomes more evident. The presence of the post-focal rising-falling movement to indicate the modality is found in the Bari corpus in 100% of the occurrences of an early narrow focus, as attested by previous studies on Southern varieties spoken in Bari, Naples and Palermo (Cangemi & D'Imperio, 2013; Grice, 1995; Grice et al., 2005; Grice & Savino, 2003, among others). By contrast, in the Udine corpus, the occurrences of early narrow focus were always followed by a flat contour terminating in a final rise (L* L-H%[12]). In line with these findings, we can conclude that the presence of post-focal rising F0 movements is not attested in this variety.

Figure 18 shows the difference between the two varieties in the F0 contour of the post-focal position in questions. Figure 18a. shows F0 contour of the question *Sopra le colline?* (Above the hills?), realised in the variety of Italian spoken in Udine and produced with the adverb *sopra* (above) in contrastive focus. The noun *colline* (hills)

[12] Note that the actual presence of the post-focal accent (L*) has not been analysed in the present corpus (no measurements of duration, spectral properties or energy have been conducted). The reported labelling as L* has only the aim of clarify the presence of a low and flat contour and to ease the description.

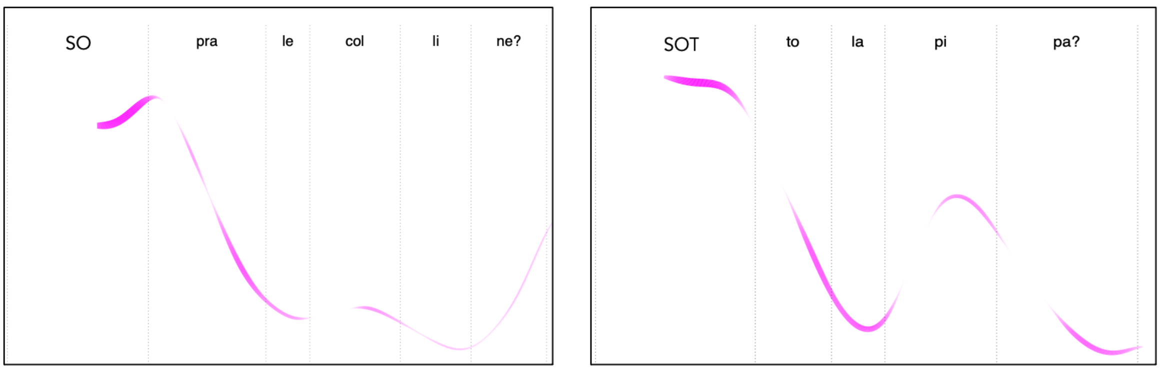
follows the constituent in focus and presents a low and flat contour ending in a rise, which conveys sentence modality. Figure 18b. shows the F0 contour of the question *Sotto la pipa?* (Under the pipe?), uttered in the variety of Italian spoken in Bari. Also in this utterance the adverb (*sotto* under) is realised in contrastive focus, but, contrary to the utterance in the Udine variety, the stressed syllable (*pi*) of the word following the focus is associated with a rising pitch movement.



a. b.

Figure 18. a. F0 curve and waveform of the question *[Sopra]F le colline?* (Above the hills?) with narrow focus on the adverb *sopra*, realised in the variety of Italian spoken in Udine. b. F0 contour of the question *[Sotto]F la pipa?* (Under the pipe?) with narrow focus on the adverb *sotto*, realised in the variety of Italian spoken in Bari.

In addition, the periograms in Figure 19 show that the energy in the word occurring postfocally in the Bari example (*pipa*) is higher than the one in the Udine example (*colline*), as reflected by the width and transparency of the curve (see 1.2.4.4).



a. b.

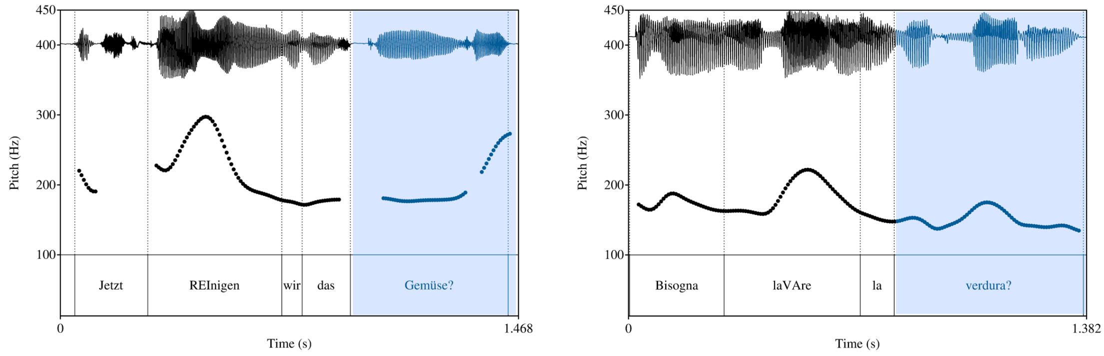
Figure 19. a. Periogram of the question *Sopra le colline?* (Above the hills) uttered in the Udine variety of Italian, with an early narrow focus on the adverb (*sopra* above). The narrow width and the high transparency of the curve for the word *colline* (hills), together with the lack of movement on the stressed syllable, reflect low values of prominence for this word. b. Periogram of the question *Sotto la pipa?* (Under the pipe) uttered in the Bari variety of Italian, with an early narrow focus on the adverb (*sotto* under). The rather large width and density of colour for the word *pipa* (pipe), together with the pitch movement, reflect high values of prominence.

Given that the distribution of prominence in the post-focal position of questions is not the same between the two varieties, the Bari variety has a greater paradigmatic choice of F0 shape for realising the post-focal position compared to the Udine variety. Indeed, the post-focal position of statements does not differ between varieties, displaying a flat and low contour, possibly presenting a greatly compressed accent, despite the (near) lack of F0 movement (Cangemi & D'Imperio, 2013; Grice et al., 2005; Gili Fivela et al., 2015). By contrast, the post-focal position of question does differ between the two varieties, with the Bari variety showing an identifiable rising-falling pitch accent, analogous to the one characterising focused words.

The next subsection will review the literature on the production of yes/no questions in German, in order to attest differences from the Bari production, especially in post-focal position.

2.3.3.2 Comparison between Bari and German in the production of yes/no questions

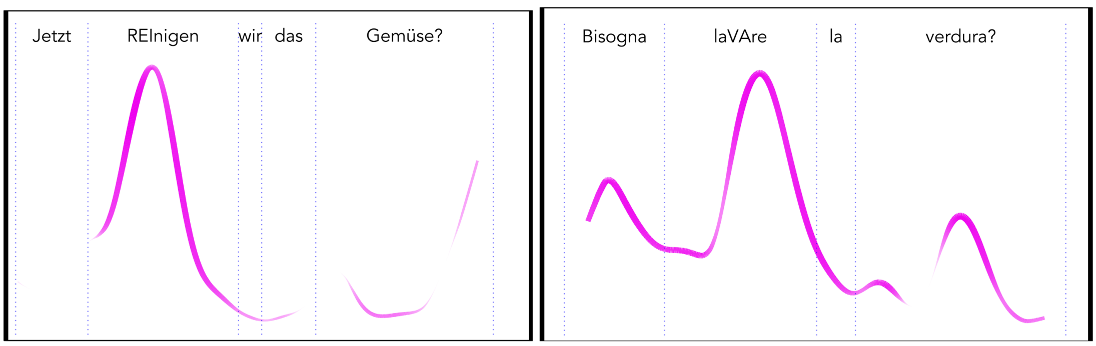
The traditional view on the realisation of questions in (Standard) German entails that the marker of questions is always the final rise (e.g., Féry, 1993; Grice & Baumann, 2000). However, rising-falling movement on questions is attested also in (Standard) German (Kügler, 2003; Selting, 1995; Wochner, Schlegel, Dehé & Braun, 2015). Nonetheless, as in the case of the variety of Italian spoken in Udine, a rising accent is never attested for German questions in post-focal position. Post-focal position in Standard German is realised with a flat and low contour ending in a rise (high boundary tone; see the examples reported in Grice et al., 2000:166-167; Grice & Baumann, 2000:15; Grice, Baumann & Benzmüller, 2005:21). Figure 20 shows the differences in the realisation of the F0 contour in the post-focal position of questions in German and in Bari Italian. Figure 20a. shows the F0 contours of the question *Jetzt reinigen wir das Gemüse?* (Now do we clean the vegetables?) uttered in German with a contrastive focus on the verb *reinigen* (to clean). Here *Gemüse* (vegetables) is realised in post-focal position and is produced with a low and flat contour and with a final rising boundary tone, which conveys sentence modality. Figure 20b. shows the F0 contour of the question *Bisogna lavare la verdura?* (Should one wash the vegetables?) realised in the variety of Italian spoken in Bari and produced with a contrast on *lavare* (to wash) and with a rising pitch accent on *verdura* (vegetables). Note that the rise associates with the stressed syllable *du*. Therefore, the German and the Bari Italian production of this utterance differ in that the stressed syllable of the word occurring postfocally in the German example does not associate with a an F0 movement, while in the Bari example it does.



a. b.

Figure 20. a. F0 curve and waveform of the question *Jetzt [reinigen]$_F$ wir das Gemüse?* (Now do we clean the vegetables?) with narrow focus on *Gemüse*. b. F0 curve and waveform of the question *Bisogna [lavare]$_F$ la verdura?* (Should one wash the vegetables?) with narrow focus on the verb *lavare*.

In addition, the periograms in Figure 21 show that the energy in the word occurring postfocally for the Italian utterance (*verdura*) is higher than the one in the German utterance (*Gemüse*), as reflected by the width and transparency of the curve (see 1.2.4.4).



a. b.

Figure 21. Periogram of two questions uttered in German (a. *Jetzt reinigen wir das Gemüse?*, Now do we clean the vegetables?) and in the Bari variety of Italian (b. *Bisogna lavare la verdure?*, Should one wash the vegetables?), with an early narrow focus on the verb (*reinigen* for German and *lavare* for Italian). The narrow width and the high transparency of the curve for the word *Gemüse* (a.), together with the lack of movement on the stressed syllable, reflect low values of prominence for this word. The rather large width and density of colour for the word *verdura* (b.), together with the pitch movement, reflect high values of prominence.

2.3.4 Interim summary

To summarise, so far this section has underlined the lack of a one-to-one relation between givenness and deaccentuation.Despite this variability, some tendencies emerge. Italian shows a behaviour similar to German in terms of deaccentuation of given post-focal clauses and entire NPs (cf. Avesani & Vayra, 2005; Avesani, Hirschberg & Prieto, 1995; Hirschberg & Avesani, 1997; D'Imperio, 1997; Farnetani, & Zmarich, 1997; Ladd, 1996), but it is reported to be different in the behaviour within NPs (Swerts et al., 2002; Avesani et al., 2015). Within NPs, the percentages reported in the literature suggest that Italian and German differ in the probabilistic distribution of the prominence in lexically and referentially given elements. The higher variability between the mapping of givenness and deaccentuation that is attested for Italian may influence perception (see 3.4). In addition, for the variety of Italian spoken in Bari, the presence of pitch accent and high energy in the post-focal position of questions leads to a higher probability in this variety to find cues to prominence in post-focal position in comparison to the variety spoken in Udine and to German. This too may have an impact on prominence perception (see 3.4).

To conclude with the overview offered in this section, Table 5 presents a summary of the differences between the two varieties of Italian and German in the realisation of post-focal position that have been outlined so far.

	German	**Bari**	**Udine**
Within NPs	post-focal deaccentuation of given elements	**possible lack** of post-focal deaccentuation of given elements	**possible lack** of post-focal deaccentuation of given elements
Sentence-long statements	post-focal downstep/ deaccentuation	post-focal downstep	post-focal downstep
Sentence-long questions	Lack of pitch accents in post-focal position	Rising pitch accents in post-focal position	Lack of rising pitch accent in post-focal position

Table 5. Summary of the landscape found in the literature regarding the prosodic realisation of the post-focal region of questions and statements in Italian and German.

Another characteristic of a language that can affect the production and the perception of prominence is the relation between prominence and word order (for a clarification on the perceptual domain see Chapter 3). The following subsection (2.3.5) will discuss the reported differences between Italian and German (and in general between Romance and West-Germanic languages) in the use of word order and intonation to mark prominence.

2.3.5 Preferred position of prominence: differences between Italian and German

The literature reports different constraints for languages in the location of prominence. For example, Spanish and Italian, when compared to English are traditionally claimed to be more rigid in the position of the main prominence and to use only word order to convey focus. Spanish and Italian have been described as languages that can change word order more freely to signal prominence than English and German (cf. Face & D'Impero, 2005; see Büring, 2009). See (15) below for examples. By contrast, it has been argued that Italian and Spanish present greater constraints to signal focus through accentuation, therefore showing resistance to change the main prominence position from the last element of the utterance to other positions in the utterance (e.g., Bolinger, 1954; Contreras, 1980 for Spanish and Antinucci & Cinque, 1977; Benincà, Salvi & Frison, 1988 for Italian). This distinction is traditionally addressed with reference to plastic vs. non-plastic languages (Vallduví, 1991), where plastic languages are defined as languages that can modify their prominence pattern in order to highlight information occurring anywhere in the utterance (e.g., English), while non-plastic languages are defined as languages that have fixed prosodic prominences and need to modify word order to mark an element as prominent (e.g., Spanish and Italian). Contrary to this traditional account, Face and D'Imperio (2005) propose to describe this typological distinction as a continuum, loosening this rather rigid distinction between plastic and non-plastic languages. This less rigid distinction is preferred here to account for the data found in the languages and especially for a more precise characterization of the status of Italian. In fact, the notion of the continuum allows us to account not only for differences between the two typologies but also for similarities.

(15) a. The COFfee machine broke.
b. Die KAFfeemaschine ist kaputt.
c. Si è rotta la caffetTIEra.

Italian shows the tendency to use alternations in word order to a greater extent, in order to allow the focal word to occur at the end of the utterance (contra West Germanic languages). This is true especially in comparison with English, whereas German is characterized by a less free use of intonation and a less rigid word order than English (see e.g., Büring, 2009; Koster, 2000). However, in sentences like in (15) in English (15a.) and in German (15b.), even if the information structure entails that the focus occurs on the subject (*coffee machine* or *Kaffeemachine*), the word order would still require the verb to come in final position and therefore the early narrow focus can only be marked by means of intonation. By contrast, in Italian the same sentence would be generally realised as in (15c.) with the focal word (*caffettiera*) occurring at the end of the utterance (cf. Ladd, 1996:191).

Along these lines, studies on word order in Spanish and Italian (Antinucci & Cinque, 1977; Benincà, Salvi & Frison, 1988; Zubizarreta, 1998) identify the sentence-final position as having a special status, due to the fact that given information tends to precede new information, which would then come at the end of the sentence. This suggests a preferred position for prominence due to information structure, as in (16) and (17), adapted from Antinucci and Cinque (1977). The answers in each example are the preferred ones for the paired questions, but are acceptable also as answers for the questions in the different examples.

(16) Question: Che fa Giovanni?
What is Giovanni doing?

Answer: Giovanni viene.
Giovanni is coming

(17) Question: Chi viene?
Who is coming?

Answer: Viene Giovanni.
Giovanni is coming

Indeed, Italian presents several cases in which intonation is used instead of word order to indicate an early focus of the sentence, thus showing similarities with West-Germanic languages. For example, the answer in (16) is acceptable in the context of (17) if it is realised with the nuclear accent on the word *Giovanni*, changing the canonical intonation pattern for a broad focus sentence (see Face & D'Imperio, 2005). Thus, intonation can be flexible in Italian too, in particular it can be used instead of word order to express an early narrow focus. According to Face and D'Imperio (2005), Italian is situated in the middle of the continuum they propose because it presents a higher tendency compared to English to use word order instead of changing the position of the accent from the last element of the utterance, but allows for greater interaction of word order and intonation in comparison to Spanish (Face, 2000, for Spanish; Caputo, 1997, for Neapolitan Italian; see also Face & D'Imperio, 2005; and Nava, 2010 for a more detailed account). The possibility of using intonation to shift the default prominence patterns results in a lack of alignment of the main prominence with the final position in the utterance. Nevertheless, a number of studies report the tendency of Italian to still maintain a certain degree of prominence in the last constituent, even if it does not correspond to the focus of the sentence (see discussion before).

Face and D'Imperio (2005) do not locate German in the continuum. However, considering the fact that its word order is more free than English but more rigid than Italian, its position on the continuum would be between these two languages and can be conceptualised to be the one reported in Figure 22.

Figure 22. Hypothesised placement of Spanish, Italian, German and English in the typological continuum going from languages that use only word order to languages that use only intonation. Modified version of the continuum proposed by Face and D'Imperio (2005:283).

The distance between Italian and German still implies differences in the preferred word order and preferred prosodic realisation between the two languages. The following example (18) adapted from Grice and Baumann (2007:9) shows one of these differences. Indeed, the second mention of the word *casa* for the Italian Example (18a.) and the second mention of the word *Haus* for the German Example (18b.) are both referentially and lexically given (they have the same referent and the same lexical form). However, while for German this second mention is deaccented, in Italian it is usually accented. This derives from a higher tendency in Italian to have an accent occurring on the last element of an utterance.

(18) a. È un lavoro che si fa dentro CAsa o fuori CAsa?

Is it a job that you do inside the HOme or outside the HOme?

b. Ist das eine Arbeit, die man INnerhalb des Hauses oder AUßerhalb des Hauses macht?

Is it a job that you do INside the home or OUtside the home?

Nonetheless, the distribution of accents in the example is just a tendency. In the Italian utterance there is the possibility of having accents on *fuori* and *dentro* and to have no accent occurring on the word *casa*. The discussion in 2.3.2 also emphasised that it is not accurate to say that Italian never uses deaccentuation, the strategy is just considerably less systematic and less common than in German (and English; Avesani et al., 2015; Cruttenden, 2006; Röhr et al., 2016; Thies et al., 2017).

2.4 Summary

This chapter has discussed the relation between information status and information structure on the one hand and prosody on the other hand, both in German (and more in general West Germanic languages) and in Italian. The central focus was the relation between the degrees of givenness and the distribution of prominence in German and in Italian, and the differences in the distribution of prominence within varieties of Italian (the variety of Bari and of Udine). This point is of interest because from the distribution in the production one can expect a probable mapping from form to function. This mapping might have an influence on the perception of the degree of prominence that listeners associate with particular acoustic features. Given the differences between Italian and German on the one hand, and between varieties of Italian on the other, some questions may arise. Firstly, given the high variability on the degree of prominence of post-focal position in Italian production, do Italian listeners perceive the post-focal region with a higher degree of prominence than the one encoded in the signal? Secondly, does the mapping between prominence and word order in Italian and German make a difference in the perception of the degree of prominence between native speakers of the two languages? Thirdly, does the exceptional high variability in the mapping between post-focal position and lack of pitch movement in Bari Italian (both within NPs and in questions), have an influence in the perception of prominence? Can the difference in the attenuation of the post-focal region between Bari Italian and German cause a difference in the prominence perception of this region between native speakers of Bari Italian and native speakers of German? Fourth, do the differences in the post-focal position of questions between the two Italian varieties reported in this chapter (Bari and Udine) influence the degree of prominence in the post-focal position perceived by native speakers of the two varieties? The rationale for these questions on perception will be discussed in the next Chapter (Chapter 3), where the relation between prominence in the signal and prominence perceived will be outlined. The questions previously presented will be addressed in the rating experiment in Chapter 6.

Chapter 3

Prominence perception and prominence as an attention orienting device

3.1 Introduction

The last chapter reviewed the relation between prosody and information structure. In particular, it highlighted the role of prosodic features in conveying information about the intentional and attentional structure of discourse. The present chapter will focus on the relevance of prosodic features in the perceptual domain and, especially, on their role in attention orienting. Throughout the chapter, a series of behavioural and on-line studies on the perception of phrasal prosodic prominence will be reviewed. The aim of this chapter is to set the theoretical framework for the experiments in Chapter 6 and Chapter 7. The object of the experiment in Chapter 6 will be the ratings of perceived prominence in the comparison between contrastive narrow focus, broad focus and post-focal position in two regional varieties of Italian (spoken in Udine and in Bari), performed both by Italian native speakers and by German native speakers. Listeners' perception of the degrees of prominence in these focal structures and the influence of the native language on the perception of prosodic prominence will be discussed in the present chapter. The experiment in Chapter 7 will be conducted on one regional variety of Italian (Bari Italian) and will investigate the modulation of the depth of semantic processing in relation to the degree of prosodic prominence. Thus, the role of bottom-up (stimulus-driven) and top-down (expectation-driven) inferences in the perception of prominence and subsequently in the depth of processing will be highlighted.

More in detail, this chapter comprises sections and subsections that explain how listeners perceive phrasal prosodic prominence and how the perception of an element as being prominent contributes to ease the processing of this element. This in turn is a result of the allocation of more attentional resources towards the element. Firstly, section 3.3 will present an overview on studies shedding light on the signal-based perception of prominence and discussing the relation between categorical and gradient perception. Section 3.4 will investigate the consequences of prominence perception during the comprehension of the information structure of utterances and information status of referents. Secondly, section 3.5 will address the influence of expectation-driven inferences on prominence perception. The discussion is relevant for understanding the influence of prosody on on-line processing, which will be discussed throughout sections 3.6 and 3.7. Section 3.8 provides an overview on research considering prosodic prominence as an attention orienting device, focusing in particular on studies employing event-related potentials (ERPs). Subsequently, the role of focus and accentuation on the depth of semantic processing will be discussed. Studies presented in these sections were largely conducted on West-Germanic languages or on languages other than Italian. Indeed, literature on Italian investigating the perception of prominence (especially the online perception) is largely missing. Therefore, the last subsection (3.8.1) will more extensively focus on language-specific driven expectations on the perception of phrasal prosodic prominence and will discuss the scarce literature available on Italian. The aim of this last section is to provide the basis for understanding the perception of the prominence degree in post-focal position in Italian. This represents the novelty of the perceptual experiments in the present thesis.

3.2 Different perspectives on prominence: signal-based and expectation-based perception

Prominence perception needs to be seen from two different perspectives: a signal-based one and an expectation-based one. On the one hand, prominence can be conveyed by the form of an element (e.g., the acoustic features of an element). This is what is referred to as signal-based cues to prominence. On the other hand, prominence profiles are dynamically constructed as a discourse unfolds and based on this information, expectations can be generated for upcoming elements (e.g., the acoustic make-up of an upcoming element). These progressively generated expectations influence the perception of prominence and constitute what is referred to as expectation-based cues to prominence (non-signal-based factors). It should be further noted that expectations regarding the prosodic form of upcoming elements in the discourse are connected to previous knowledge built on language experience.

Signal-based and expectation-based perception of prominence are strictly connected to the concepts of top-down and bottom-up processing. These two types of processing are included in perception in general. Bottom-up processing is based on the instant input from incoming data (integration of information from the world that stimulates one's receptors). Top-down processing is based on previous experience (integration of information from the individual cognitive system, cf. Goldstein, 2010:1011), which can also comprise inferences from the context. For speech perception, both types of processing are necessary (cf. Shuai & Gong, 2014; see also McClelland & Elman, 1986 and Hickok & Poeppel, 2007). Relative to prominence perception, they configure as: (i) processing that depends on the instant auditory stimulus (bottom-up), which corresponds to the signal-based processing, and (ii) processing that depends on language experience, which for example allows us to generate inferences from the context and corresponds to the expectation-based processing.

Section 3.3, 3.4 and 3.5 will deepen the two aspects comprising prominence perception (signal-based and expectation-based). In these sections, the effects that the interplay (i) of acoustic characteristics of the stimulus, (ii) of the context in which it occurs and (iii) of the knowledge of the language generates on prominence perception will be outlined. These effects are central to understand the question addressed in the perception experiments presented in this thesis, since they aim at deepening the understanding of how the probabilistic distribution of prominence in a language (generating top-down expectations) interacts with the presented acoustic stimulus (bottom-up inferences) in the perception of prominence (experiment in Chapter 6) and in online language processing (experiment in Chapter 7).

3.3 Signal-based perception of prominence

3.3.1 Acoustic features in phrasal prominence perception

Prosodic features play an important role in sentence and discourse processing (e.g., Arnold, 2008; Cutler, Dahan & Van Donselaar, 1997; Dahan, Tanenhaus, & Chambers, 2002; Nakatani, 1997; Terken & Nooteboom, 1987). Several studies have pointed out that listeners are sensitive to the acoustic changes that speakers make to convey phrasal prosodic prominence (Cole, Goldstein, Katsika, Mo, Nava & Tiede, 2008; Eriksson, Thunberg, & Traunmüller, 2001; Honorof & Whalen, 2005; Krahmer & Swerts, 2001; Turk & Sawusch, 1996; among others, see also following sections). For instance, listeners, being speakers themselves, might, at some level, be aware of the effort needed to raise the F0 and therefore recognise it (Hsu, Evans and Lee, 2015). The effort entailed by speakers in producing an entity as prosodically prominent has consequences on the perceptual domain, enabling the prominent elements to be perceived as ‘standing out’ in relation to their neighbours (Terken, 1991).

A large body of work has been conducted on the role of various acoustic features in the perception of phrasal prominence. Studies employing the Rapid Prosody Transcription (RPT) paradigm, developed by Cole and colleagues (Cole, Mo, & Baek, 2010; Cole, Mo, & Hasegawa-Johnson, 2010; Mo, Cole & Lee, 2008; see also Cole & Shattuck-Hufnagel, 2016) are especially relevant. In this task, participants with no training in prosodic transcription and no specific knowledge about prosody or speech listen to spontaneous recorded speech samples. Listeners are provided with a transcript of the speech and are asked to underline in real-time the words that they hear as prominent (another version of the task consists in marking the prosodic boundaries, however this is not of interest in this chapter). The instructions given to the transcribers are minimal and correspond to the following sentence (with some variations depending on the study): “mark as prominent words that the speaker has highlighted for the listener, to make them stand out” (Cole & Shattuck-Hufnagel, 2016: 8; for diverging instructions depending on the aims of the study, see Cole, Mahrt & Hualde, 2014).

Studies employing the RPT task and conducted on American English (Cole et al., 2010a, b; Mo, 2011, among others), French (Smith, 2011), Spanish (Hualde, Cole, Smith, Eager, Mahrt & Napoleão de Souza, 2016), German (Baumann & Winter, 2018) and Russian (Luchkina & Cole, 2014) show that there is a strong positive correlation between words that are considered prominent by raters and the acoustic features that are usually attributed to prominence (see Riesberg, Kalbertodt, Baumann & Himmelmann, 2018; You, 2012 for RPT studies on languages that show different prosodic characteristics, which might cause different results and would need a separate discussion). In these languages pitch accents are typically, although not always, realised with a pitch excursion and combined with metrically prominent positions and have been shown to correlate with perceived prominence (Cole et al., 2019; Baumann & Winter, 2018). In fact, words assigned with a ToBI pitch accent label in each language are rated as more prominent than unaccented words.

It is worth noting that more than static properties of F0 (F0 maximum and minimum), F0 dynamics, indicated by range (F0 excursion) and slope (F0 excursion per second) are particularly important in the perception of prominence (cf. Baumann & Winter, 2018; see also Baumann & Röhr 2015; Niebuhr, 2009; Rietveld & Gussenhoven, 1985). However, not only F0 movements but also acoustic measures related to duration and overall energy and their interaction contribute to the perception of prominence (Arnold, Wagner, & Baayen, 2013; Baumann & Winter, 2018; Campbell, 1995; D'Imperio, 1998; Kochanski, Grabe, Coleman & Rosner, 2005; Mahrt et al., 2012; Mo, 2008; Turk & Sawusch, 1996; Wagner et al., 2015). This indicates that the perception of prosodic prominence is not based on a single acoustic feature and that the gradient modification of several acoustic features contributes to the perception of prominence. The next subsection will provide a discussion regarding the contribution to prominence perception of gradient changes in the parameters conveying prominence.

3.3.2 Gradient perception of prominence

In the RPT studies mentioned in the previous section participants rate prominence categorically, with a binary distinction between prominent and non-prominent words. This categorical distinction is in line with metrical phonology in which words that are prominent are categorically and structurally distinct from non-prominent words, and words that are in nuclear position are categorically different from the ones in pre-nuclear position (Cole et al., 2019; see also 1.3.1.1). Moreover, this distinction is motivated by the presence or absence of an accent, that, as seen in the previous section, is a good predictor of whether words are considered prominent or not. However, the acoustic properties of syllables in prominent positions cannot only be described as categorical: the variation of different acoustic cues and the change in their interaction signal gradient prominence distinctions which are perceived by listeners (Ayers, 1996; Arnold, Wagner and Möbius, 2010, 2011). These gradient properties suggest that the perception of prominence, in turn, is not strictly binary (prominent vs. non-prominent) but also gradient. The same studies using the RPT paradigm have found both a categorical and a continuous effect on prominence perception (Mahrt et al., 2012; Cole at al., 2019). Indeed, Cole and colleagues found the effect of categorical structural prominence (for example, the occurrence of a prosodic boundary increasing the likelihood of prominence ratings) but also a continuous variation of the likelihood of prominence ratings in relation to continuous variation in the acoustic cues: the greater the enhancement of the acoustic cues of a word, the more likely it is rated as prominent (cf. Cole et al., 2019:130).

Although in many experiments on prominence perception subjects rate prominence only on a binary scale, studies have shown that prosodic prominence is perceived at a fine-grained level, i.e. listeners are able to distinguish various degrees of prominence (Ayers, 1996; Baumann & Röhr, 2015; Cole et al., 2019; Watson, 2010; among others). These studies show that properties of prosodic prominence are systematically ranked in terms of perceived prominence. For example, Baumann and Röhr (2015) tested the perceptual prominence of seven different accent types that are present in German. They recorded utterances in which a proper name was realised with different accent types and asked listeners to give their judgments on how *highlighted the name sounded*. Participants had to provide their judgments on a continuous horizontal line (a visual analogue scale) that allowed for gradient ratings. This study revealed that rising accents are more prominent than falls, with steep rises and falls perceived as more prominent than shallow ones. Moreover, rising accents were perceived as more prominent than down-stepped accents and low (shallow falling) accents. This study also took into consideration deaccentuation, showing that it produced the lowest values of perceived prominence. This is not surprising, since deaccentuation requires the least effort for the speaker and does not present acoustic features related to phrasal prominence (see 2.2.2). Cole et al. (2019; see also Hualde et al., 2016) reported for English, Spanish and French that pitch accents that in a language are commonly associated with (contrastive) narrow focus are more likely to be perceived as prominent in comparison to other accent types, since accents marking (contrastive) narrow focus are characterised by greater acoustic prominence (see 2.2).

To summarise, listeners are able to perceive gradual, fine-grained changes in the F0 dynamics, duration and energy and their interaction and to rank these changes in different degrees of perceived prominence. This consideration constitutes the basis for the experiment in Chapter 6, in which the listeners' fine-grained perception of prominence will be investigated, with particular interest in the perception of the post-focal position. Additional evidence for the fact that listeners are able to perceive the changes in prominence and that they can further relate these changes to different information in the discourse is provided in the following section (3.4). Studies presented in this section will refer mostly to West-Germanic languages, which the majority of the literature on this topic has focussed on.

3.4 Prosodic prominence influences the perception of information status and information structure

As extensively discussed in the previous chapter, in many languages, speakers employ different degrees of prominence to mark information status and information structure. Production studies suggest that there is a large amount of variability in the mapping between the intentions of speakers and their intonation realisation (see 2.2), in the acoustic dimension and in the characteristics of the speakers (see 1.2.4.2). In addition, the context in which a specific utterance occurs influences the perception of prominence, as further discussed in the next section (3.4.; see Bishop, 2012, 2016). Hence, for listeners, understanding the mapping between the form used by a speaker and the function that this speaker wants to convey might not be straightforward.

For example, Gussenhoven (1983) found that English listeners could not reliably distinguish whether an utterance was uttered in a context eliciting narrow informational focus or in one eliciting broad focus. Accordingly, in Welby (2003) the distinction of which question was appropriate for the utterance *I read the DISPATCH* with a contrastive pitch accent on *dispatch*, was not easily teased apart. The questions *Which newspaper do you read?* (eliciting narrow focus on the word *dispatch*) and *How do you keep up with the news?* (eliciting broad focus) were both considered suitable contexts for the critical utterance. A similar result was found in Breen et al. (Breen, Fedorenko, Wagner, & Gibson, 2010). In this latter study, English participants were paired. One of them was asked to silently read a question and then to utter the answer, while the other participant was required to select, among a set of seven questions, which was the most suitable one for the utterance heard. Listeners were often not accurate in identifying the correct question. However, in these studies results could have been influenced by the difficulty of the task, arising either due to the high number of questions among which the choice was possible (Breen et al., 2010), or by the rather high ambiguity in the context (Gussenhoven, 1983). Moreover, in Welby (2003) participants had to give appropriateness judgments, which might have hidden preferences that would have arisen in a forced choice task.

Indeed, Cangemi, Krüger and Grice (2015) showed that (in German) listeners are clearly able to detect focus type, distinguishing among broad, contrastive narrow focus and no focus. In addition, Röhr and Baumann (2011; see also Röhr, 2016) examined the perceived level of referent's givenness (see 2.2.2) indicated by different accent types and positions. They have found that deaccentuation is associated with a high degree of referent's activation, whereas nuclear accents were mostly interpreted as encoding new or less activated referents: "An increase in prominence-lending cues on the target referents triggers a decrease in the degree of perceived givenness" (Röhr & Baumann, 2011:1706). Their findings were valid for both utterances in isolation and in context. This study supports the previous one by Baumann and Grice (2004) which revealed high pitch accents (H*), usually a marker of new information, to be indeed judged as the most appropriate accents of new information. By contrast, deaccentuation was considered the most appropriate for given referents.

Compelling evidence for (American) English is presented by Roettger, Mahrt and Cole (2019). The study investigated utterances (*Damon fried the omelet*) in broad focus (elicited by the question: *Do you know what happened yesterday?*), with the subject narrowly focused (elicited by the question: *Do you know who fried the omelet?*), with the subject contrastively focused (elicited by the question: *Did Pam fry the omelet?*), as well as completely given utterances (elicited by the question: *Did Damon fry the omelet?*). Results showed a high degree of correct matching between prosodic form and the information structure that utterances were intended to convey. Though the distinction between utterances signalling contrastive focus and the ones signalling broad focus was easier than the distinction between completely given utterances and narrow focus, these differences can relate to the greater acoustic overlap in the F0 contours of the completely given utterance and the narrowly focused object.

Taken together, these studies suggest that for West-Germanic languages listeners are capable of interpreting the referent's information status building on its degree of prosodic prominence. Speech signal seems to be perceptually mapped into discourse functions, even though the exact mechanisms on how this happens are still not completely understood (cf. Roettger et al., 2019).

Psycholinguistic work has also provided support for the claim that listeners are able to interpret the information status of a word relying on its acoustic characteristics. Further, it has demonstrated that listeners can anticipate speaker intentions based on intonational information. Many studies have shown that in English, Dutch and German, language comprehension is facilitated by the accenting of focused information and the deaccenting of non-focused information (Birch & Clifton 1995; Bock & Mazzella, 1983; Terken & Nooteboom, 1987). When new information is accented the understanding of the content is faster and more accurate. For example, in an eye-tracking study Dahan et al. (2002) investigated the number of fixations on referents in a series of two types of instructions. They showed that an accented referent on the second instruction lead to more fixations on the corresponding new (not previously mentioned) object in comparison to the given (previously mentioned) one. Namely, in the instruction *Move the candle above the triangle. Now move the candy below the square*, an accent on *candy* increased the number of fixations towards the image of the candy at the time point in which the two words could not yet be distinguished (during the first syllable *can*). Accordingly, an accented adjective in the second instruction (e.g., *Click on the purple violin. Now click on the RED violin*) lead to more fixations on the same object of the preceding instruction compared to a different object (Ito & Speer, 2008; for evidence on German see Weber, Braun, & Crocker, 2006). These results indicate that prosodic information allows us to interpret the information status of the referent before the signal unambiguously identifies it, accelerating the comprehension of whether the speaker is introducing a new referent or is continuing with the previously mentioned one. This is important in the processing of discourse as it helps fast updating of the mental model which listeners create during language comprehension (see 3.5.1 and following). In addition, Watson, Tanenhaus and Gunlogson (2008) showed that during processing of information a steep rising pitch accent (L+ H*) creates a bias towards contrastive referents, while a less steep accent (H*) signals both new and contrastive referents. Finally, Roettger and Stoeber (2017) replicated previous results on on-line integration of intonation. They used the mouse-tracking paradigm which, in comparison to the eye-tracking paradigm, more precisely detects the time course of referential processing by charting the hand-mouse movement trajectory (see Spivey, Grosjean, & Knoblich, 2005).

To summarise, evidence has been collected on the correct perception of intonation events and their intended functions, in that it has been shown that listeners can correctly map acoustic prominence to information structure (or information status of elements). This evidence also comprises the anticipatory eye and hand movements in response to intonational events that allow to predict upcoming words before the lexical disambiguation occurs (Dahan et al., 2002; Ito & Speer, 2008; Roettger & Stoeber, 2017; Watson, Tanenhaus, & Gunlogson, 2008; Weber et al., 2006). The degree of prominence of the signal is, however, not the only factor influencing the perceived prominence. The next section will present studies on the other factors that play a role in prominence perception. This discussion will be of interest in the experiments in Chapter 6 and in Chapter 7, which aim to disentangle top-down from bottom-up inferences.

3.5 Expectation-based perception of prominence

As already discussed in the previous sections, acoustic characteristics of the signal play a major role in perception. However, all studies mentioned have found great variance in the prominence ratings, suggesting that these characteristics are not sufficient to completely account for the prominence perceived. This section offers a review on the studies showing the effects of expectation-driven inferences on the perception of prominence. Such expectation-driven effects are expected to play a role in the degree of prominence of the post-focal position as perceived by native speakers of the different varieties of Italian investigated in this thesis (Bari and Udine) and by German native speakers.

A first evidence of the fact that signal-based inferences are not sufficient to explain the prominence perceived by listeners was the study by Eriksson, Thunberg, and Traunmüller (2001), which correlated prominence rating with signal-based variables relating to vocal effort, pitch and duration. Results showed that these variables described 48% of the variance in the prominence ratings. When incorporating the discrete categorisation of being accented or conveying contrast, the variance explained increased to 57%, still leaving out a high percentage of not explained variance. This study has showed, therefore, that prominence perception not only depends on a large number of different variables, but that these variables are not enough to explain prominence perception in its whole. Despite the evidence that fine-grained acoustic cues are perceived by listeners, prosodic prominence perception is not only based on the acoustic characteristics of the stimuli (i.e. the signal-based input), but is also expectation-based.

The impact of expectation-driven information on prominence perception has been demonstrated in several studies (Bishop, 2012, 2016; Baumann & Winter, 2018; Cole, Mo & Hasegawa-Johnson, 2010; Heldner & Strangert, 1997). In these studies, expectation comprised intrinsic properties of the words such as frequency of occurrence as well as external factors from the context of utterance. Crucially, Cole et al. (2010b:428) indicated that “the perception of prominence is more complex than can be predicted by a simple signal-based model where acoustic cues are the primary influencing factor".

As far as intrinsic properties of words are concerned, studies have shown that word frequency influences listeners' judgment of prominence: words that occur frequently in a language are less likely to be rated as prominent than words with lower frequency, (partly) independent from the fact that frequently mentioned words are subject to reduction phenomena (Cole, Moe & Hasegawa-Johnson, 2010; Baumann & Winter, 2018). In a study on American English, Cole, Mo and Hasegawa-Johnson (2010) have shown that the perception of prominence is dependent both on signal-driven and on expectation-driven factors, namely, the predictability of a word. They considered predictability in terms of word frequency and discourse givenness (measured as the amount of repetitions in the speech sample given to participants) and found that the less predictable a word is (less frequent and with a low repeated measure index), the more prominent it is perceived. They demonstrated that the words that participants marked as prominent could be predicted also on the basis of unpredictability, reaching the same accuracy of the prediction based on acoustic cues (see also Cole, Mo & Beak, 2010).

With regard to external contributions to perceived prominence, discourse context (Eriksson et al., 2001; Cole at al., 2010b), visual cues (Krahmer & Swerts, 2007), language-specific expectations (Baker, 2010; Cole et al., 2019; Grabe, Rosner, García-Albea, & Zhou, 2003; Huang, 2004; Qin & Mok, 2012; Shport, 2015; You, 2012; see 3.7), knowledge of the language (e.g., Huang, 2004; Jurafsky et al., 2001; Streefkerk, 2002; Cole et al., 2010b) and position of the word in the utterance (Ayers, 1996; Bishop, 2012; Cole et al., 2010b; Jagdfeld & Baumann, 2011; Ladd, 1996), all play a role in prominence perception. These effects are independent of acoustic factors, and are therefore evidence for top-down (expectation-driven) prominence processing.

Within a certain discourse context, the information available creates anticipations regarding information structure and referents that are plausible for the upcoming utterance in the discourse continuation. Arnold, Wagner and Möbius (2010) have used the priming paradigm to manipulate the expectations of prominence in an upcoming utterance. In their study participants were presented with a series of (German) semantically different utterances (4 priming utterances and one test utterance) with the same syntactic structure and the same prosodic structure, manipulating the prominence pattern of the second constituent. This constituent could be either accented in the priming utterances and not accented in the test utterance, or deaccented in the priming utterances and accented in the test utterance. Subjects were found to be influenced by the expectations elicited by the prime, showing a significant difference in the prominence ratings of the critical word on the basis of the prominence pattern of the preceding utterances.

In line with this finding, Bishop (2012) has shown that the information conveyed by the acoustic signal can be overwritten by expectations derived from the discourse context. He presented listeners with utterances that did not vary in their prosodic structure, but in the context-questions preceding them. Results showed that when utterances were preceded by questions that conveyed a broad focus interpretation, listeners reported hearing the verb and the object as similar in prominence. By contrast, when the same utterances were preceded by questions eliciting narrow focus on the object, listeners reported to hear the object as more prominent in comparison to the verb, even though acoustic information did not change. Thus, listeners' interpretation of prominence patterns was shown to be systematically different depending on the underlying information structure. Accordingly, Bishop (2016) defines perceived prominence as:

> […] the subjective impression of prosodic strength that a listener experiences in some perceptually measurable way. This contrasts with acoustic prominence and structural (i.e., phonological) prominence, although perceived prominence is sensitive to both. (Bishop, 2016:668).

This means that prominence of a word is not objectively perceived as a direct consequence of its structural prominent position and its acoustic cues. These features and their interplay contribute to the perception of prominence, but are not the only cues. In addition, listeners' sensitivity to variations in F0, duration and intensity is different between speech and non-speech stimuli (Burnham, Francis, Webster, Luksaneeyanawin, Attapaiboon, Lacerda & Keller, 1996; Hsu et al., 2015; Qin & Mok, 2012). This also proves that the relationship between prominence and its acoustic correlates is not a one-to-one relationship, but that prosody and listeners' expectations regarding phrasal prominence interact in a complex way (cf. Wagner, 2005:2381).

The prominence of a syllable and of a word is also perceived in relation to the utterance's metrical structure. The positional hypothesis implies a hierarchical order of perceived prominence: the perception of an accent as prominent follows a decreasing order, going from the nuclear accent of the Intonation Phrase being perceived as more prominent than the nuclear accent of the intermediate phrase, being in turn perceived as more prominent than the prenuclear accent. Baumann and Winter (2018) implemented this hierarchy including post-nuclear prominence, corresponding in German to the phrase accent (see 1.3), which is perceived by listeners as having the lowest level of perceived prominence. The same authors have found that accent position plays a very important role in the perception of prominence, being the second variable for importance in the prominence perception. This might have an influence on the generation of expectations of prominence: "Like in music, once a rhythm has been established, the expectation of a strong beat may be enough for people to perceive one, even without discernible phonetic cues" (Calhoun, 2007:56). Furthermore, Jagdfeld and Baumann (2011) have found that a word bearing a less prominent accent than the word preceding it in an utterance is considered accented when realised as the last accent in the utterance (nuclear position), while it is not considered accented when occurring before a clearly accented word (thus being in pre-nuclear position).

Part of speech (POS) is another factor that affects the perception of prominence due to the fact that some POS categories appear more commonly than others in sentence positions where prominence is more frequently assigned. That is, nouns in English will be more likely rated as prominent, since they frequently occur in sentence final position. For the same reason, certain POS categories can also disfavour prominence: for example, English usually avoids accentual prominence on verbs (Büring, 2016; Gussenhoven, 1983).

The previously described behaviour in rating prosodic prominence (which can be predicted by a multitude of factors) and in the association of the intonational contours to a function, might be seen as the result of probabilistic inference processes (in the theoretical framework of *inference under uncertainty*; cf. Roettger et al., 2019; see also Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Kleinschmidt & Jaeger, 2015; Kleinschmidt, Weatherholtz, & Jaeger, 2018; Norris, McQueen, & Cutler, 2003). Cole et al. (2019), underline that the association between pitch accents and information structure in speech production is only probabilistic (not one-to one mapping, see 2.2; Chodroff &

Cole, 2018, 2019; see also Cangemi et al., 2015 for German; Cangemi & Grice, 2016 for Neapolitan Italian). They posit that as the variability in production is great, the top-down (expectation-driven) inferences are similarly variable. Given the knowledge of the acoustic cues' distribution, listeners might associate each linguistic unit with the distribution of acoustic cues characterising it and might probabilistically infer the likelihood of a linguistic unit to occur in a context, given the knowledge of the acoustic cues' distribution.

In this framework, expectations are built as deriving from the probability of intonational contours/acoustic correlates of prominence to occur in a given context. This context might coincide with the immediately preceding discourse context, with the context of one speaker's idiosyncrasies (inter-individual differences) and of the different realisations a speaker might have during the discourse (intra-individual differences; for inter-intra individual differences see Cangemi et al., 2015; Grice et al., 2017; Roettger, 2017; Turnbull, 2017), and with the context of a specific language. Prosodic processing seen as *inference under uncertainty* can explain the successful interpretation of prosodic information in spite of its variability and allows for the simultaneous integration of top-down and bottom-up input (cf. Roettger et al., 2019). This interpretation is in line with the lack of a one-to-one mapping of acoustic cues into prominence categories.

Finally, numerous studies have emphasised the presence of variation among participants in the perception of prominence (Baumann & Winter, 2018; Bishop, 2016; Cangemi et al., 2015; Roy, Cole & Mahrt, 2017; Shport, 2015, among others), suggesting that one of the aspects that "matters for prominence perception is relative with respect to *who* is listening" (Baumann & Winter, 2018:35), with potential cues for prominence more noticed by some listeners than others. Listeners are, therefore, influenced by their top-down inferences in prominence ratings, which might be a consequence of differences in their processes of probabilistic association between intrinsic and extrinsic properties of a word and its degree of prominence.

In view of the discussion so far, listeners seem to infer prosodic prominence and its consequent mapping into speakers' intentions building on bottom-up acoustic cues and top-down (probabilistic) expectations. Expectations concern the likelihood of the co-occurrence of prominence with a specific word in a context. This likelihood is affected by the frequency of the word in the language, its information status, the position of the word in the utterance and by the individual differences of the listeners, which include the higher sensitivity to some cues in comparison with others, participants' knowledge of the language and the associations inferred by the context that each participant can differently make. Thus, perception of prosodic prominence, rather than being associated with any acoustic correlate, should be viewed as a complex interaction of various acoustic cues and thought of as the product of multiple sources, concerning different cognitive and linguistic aspects (cf. Watson, 2010). Prominence should be conceived as a continuous measure relating to information status, acoustic cues and the interaction between speaker-based and listener-based components reflecting cognitive processes of discourse generation and production (cf. Luchkina, 2016).

In light of these findings on the nature of expectation-driven prominence perception, given the different distribution of prominence cues in the two varieties (see 2.3.3.1), the presence of differences in the perception of the post-focal position between native speakers of the Bari and Udine variety of Italian can be hypothesised. Along the same line, a difference between native speakers of the Bari variety of Italian and German learners in the perception of the post-focal position can be expected, both for the different distribution of prominence cues in the two languages and for the reported different probabilistic mapping between information status and prosodic marking. The latter reason can be taken to hypothesise differences between the German learners of Italian and native speakers of the Udine variety. In this case, the probabilistic distribution of prominences cues is not different, but the mapping between information status and information structure is, and this could lead to differences in prominence perception. All these hypotheses will be tested in the experiment in Chapter 6.

Despite the findings of a considerable amount of studies presented in this section, there is still an ample extent of knowledge missing about the integration of signal and structural cues and about the individual strategies of listeners, both within and across languages (Wagner et al., 2019:3). The present thesis and experiments in Chapter 6 and in Chapter 7 aim to further the understanding of these issues and of the complexity of prominence perception. One way to address this complexity is to look at the online processing correlates connected to prominence. Section 3.7 presents studies that sought to measure the underlying cognitive mechanisms in the brain, while the next section (3.6) will introduce a method to investigate these mechanisms (event-related brain potentials, ERPs).

3.6 ERPs, signal-based processing and expectation-based processing

Various methods to examine cognitive mechanisms in the brain are nowadays available. Electroencephalography (EEG) and in particular the EEG-based method of event-related brain potentials (ERPs) is widely and very commonly used for studying language processing. ERPs have also been frequently employed to investigate the neural substrates of the processing of prosodic prominence and will also be utilized in the present thesis (in Chapter 7).

ERPs are small changes in the electrical activity of the brain that are temporally connected to sensory or cognitive events (cf. Bornkessel-Schlesewsky & Schumacher, 2016:589). ERP signals are obtained through electrodes attached to the surface of the scalp and can be isolated from background activity by means of an averaging procedure across multiple stimuli of the same type. The signal is averaged from the critical stimulus' onset onwards: this results in a series of positive and negative shifts of potentials over time (cf. Bornkessel-Schlesewsky & Schumacher, 2016:589; see also Luck, 2005; see Figure 23 for a graphical explanation). ERPs are then classified corresponding to four dimensions (cf. Bornkessel-Schlesewsky & Schumacher, 2016): polarity, namely whether they have a negative or positive shift in comparison to a control condition; latency, namely the time from stimulus onset to the onset or to the peak of an effect; topography, namely the sites of the electrodes where the difference was observed; amplitude, namely the difference of the area under the curve between the critical and the control condition. Polarity, latency and topography are used to classify the ERP *components*: particular deflections are attributed to a certain functional significance, while amplitude of the deflection is considered a measure of the effect size (cf. Bornkessel-Schlesewsky & Schumacher, 2016:589; see Figure 24). The nomenclature of ERPs is usually defined through polarity and latency; however, the components can also derive their names from their function (see Luck, 2005 for a more detailed description).

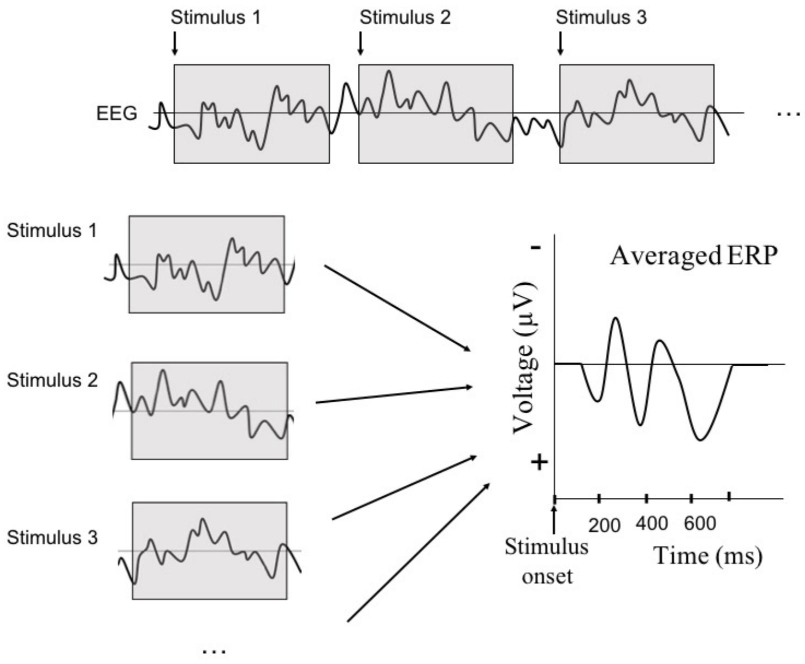


Figure 23. The top row shows the continuous EEG. The EEG segments following the onset of each stimulus are extracted (grey boxes) and averaged to obtain the ERP. Adapted from Luck, Woodman & Vogel (2000:433).

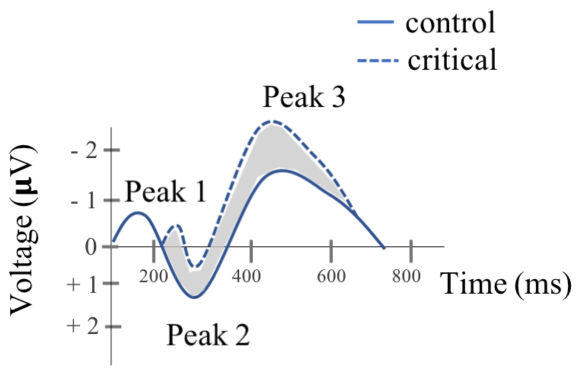


Figure 24. Idealised ERPs from control and critical condition. The filled area represents the amplitude of the component (i.e., effect size).

The advantage of ERPs in comparison to behavioural measures is that ERPs provide a continuous estimate of the response to a stimulus and, therefore, a more direct measure of brain activity. This allows us to measure a specific cognitive process and its variation in time. In fact, ERPs provide millisecond accuracy on the time course of brain processes as they unfold in time. Changes in neural processing can reveal the time point at which a specific difference among stimuli is recognised. A second advantage is that ERPs provide an online measure of the processing of stimuli without the need for a behavioural response (cf. Luck, 2005), namely an explicit task. Specifically, in behavioural tasks and in on-line eye-tracking experiments it is difficult to evaluate the processing of an unattended stimulus: if a question about an ignored stimulus is made, this stimulus may become attended instead of unattended. ERPs allow to compare the processing of attended versus ignored stimuli, since the processing of ignored stimuli can be “covertly” monitored.

There are several functionally distinct ERP components that are relevant for language processing. One component which has been widely studied and is relevant for the investigation carried out in Chapter 7 is the N400, a negative deflection that peaks around 400 ms after the onset of the critical stimulus. It is mostly regarded (along with other componenets, see 3.7.2) as indicating the degree of mismatch between a word and its previous semantic context (van Berkum, Brown, & Hagoort, 1999; van Berkum, Zwitserlood, Hagoort, & Brown, 2003; Bornkessel-Schlesewsky & Schlesewsky, 2019; Brown & Hagoort, 1993; Chwilla, Brown & Hagoort, 1995; Hagoort & Brown, 2000; Kutas & Hillyard, 1980; Kutas & Federmeier, 2011; among others), but has also been shown to be sensitive to the processing of prosodic prominence (Baumann & Schumacher, 2012; Dimitrova, Stowe, Redeker & Hoeks, 2012; Heim & Alter, 2006; Hruska & Alter, 2004; Hruska, Alter, Steinhauer, & Steube, 2001; Li, Deng, Yang & Wang, 2018; Magne, Astésano, Lacheret-Dujour, Morel, Alter & Besson, 2005; Schumacher & Baumann, 2010; Toepel & Alter, 2004; Toepel, Pannekamp & Alter, 2007; Wang, Bastiaansen, Yang & Haagort, 2011). In fact, as mentioned in section 3.3, studies monitoring fixations have pointed out that prosodic cues are processed in real-time (Dahan et al., 2002; Ito & Speer, 2008; Watson et al., 2008; Weber et al., 2006; see also Roettger & Stoeber, 2017 for evidence using the mouse-tracking paradigm). In line with these results, several auditory ERP studies also revealed that prosodic cues are computed as the sensory input unfolds and that expectations for the prosodic realisation of upcoming entities are incrementally built up.

Therefore, ERPs are useful to disentangle top-down from bottom-up processing of prosodic information. Indeed, the difference between the processing of the upcoming acoustic input (bottom-up processing) from the one predicted while the discourse unfolds (top-down processing, see 3.2 and 3.6.1) yields effects on the ERP components. For example, the N400 is one of the components that is modulated by the mismatch between the processing of the prosodic characteristics of an element and the expectations of these characteristics. The following section (3.7) will outline the literature on the effect of prosody on the N400 and on other components which revealed useful for the understanding of prominence processing.

3.7 Neurophysiological correlates of prosodic prominence: ERPs

3.7.1 The N400 component and prosody

The present subsection focuses on the N400 component, which will be of particular interest for Chapter 7 of this thesis. In addition, an overview of studies proving that the N400 also reflects access to the prominence relations that mark the information status of elements is provided. The aim of this discussion is to better understand how prosodic marking impacts processing, in order to have a clearer background to understand the relation between prosodic prominence and attention (section 3.8). Studies that are reported register effects of prosody on processing which interest not only the N400, but also other components (that are associated with different mechanisms), effects that will be described in the next subsection (3.7.2). The focus of the present subsection will consist in the processing of mismatch (either semantic or prosodic) of an element with the context. This issue will be further deepened in subsection 3.8.1 and will be of interest for the experiment in Chapter 7 in understanding the effect of the different degrees of acoustic prominence in focal and post-focal position on the semantic processing.

The processing system engages constantly in the construction of a mental model. This mental model generates predictions for the upcoming input. When the prediction fails and a mismatch between an element and its context occurs, this yields a negative ERP deflection, the N400 (Bornkessel-Schlesewsky & Schlesewsky, 2019). The N400 component is a negative potential with a centro-parietal maximum distribution over the scalp, peaking around 400 ms after the onset of a critical entity (Kutas & Hillyard, 1980). It reflects the prediction error of the critical entity in relation to the current context. This has been found for the processing of less expected words (e.g., van Berkum et al., 1999; van Berkum et al., 2003; Brown & Hagoort, 1993; Chwilla et al., 1995; Hagoort & Brown, 2000; Kutas & Hillyard, 1980; Kutas & Federmeier, 2011; among others), less accessible discourse entities (e.g., Burkhardt, 2006; Yang, Perfetti & Schmalhofer, 2007; Schumacher & Hung, 2012) and violations in regard to world knowledge (Hagoort, Hald, Bastiaansen & Petersson, 2004). For example, semantic violations such as *He spread his warm bread with socks* (vs. *butter*, Kutas & Hillyard, 1980), as well as word knowledge violations such as *Dutch trains are white* (incongruous) vs. *yellow* (congruous; Hagoort et al., 2004), elicit an N400 effect. The higher the prediction error for an entity, the more enhanced is the N400 amplitude.

A variety of studies have been conducted in the domain of speech processing, showing that, similarly to written words, spoken words elicit a higher N400 amplitude when they are semantically unappropriated with the preceding utterance context than when they are appropriate in the context of the utterance. For example, studies investigating the effect of highly predictable versus low predictable words observed an N400 (e.g., Besson, Faita, Czternasty & Kutas, 1997; Connolly, Stewart & Phillips, 1990; Connolly, Philips, Stewart & Brake, 1992) as well as studies investigating semantic anomalies (e.g., Connolly & Phillips, 1994; Friederici, Pfeifer & Hahne, 1993; Holcomb & Neville, 1991; McCallum, Farmer & Pocock, 1984; van Petten et al., 1999).

The N400 is also elicited by the interaction of focus and accentuation. A mismatch in prosodic input, such as a missing accent on new information or a superfluous accent on background information causes difficulties in the processing of the incoming information, eliciting a negativity in the N400 time window (e.g., for German Heim & Alter, 2006; Hruska & Alter, 2004; Hruska et al., 2001; Schumacher & Baumann, 2010; Toepel et al., 2007; Toepel & Alter, 2004; see also Magne et al., 2005 for French, Dimitrova et al., 2012 and Wang et al., 2011 for Dutch). The data on the interaction of prosody and focus, however, have been mixed. For example, Hruska et al. (2001, among others) have found an N400 only for missing accents and not for superfluous accents. They have shown that listeners' brains respond to the congruity of accents in the discourse, producing an N400 effect when the accent is not congruous with the focus of the sentence and is missing on a new and focussed element (*What does Peter promise Anna to do? – Peter promises ANNA to work and to clean the office*; the word *work* is the focus of the utterance but is realised as deaccented: missing accent on *work.* Example translated from German and taken from Hruska et al., 2001:426). By contrast, findings by Dimitrova et al. (2012) show a different pattern, in which only superfluous accents elicit an N400 and not the missing ones.

In an ERP study, Schumacher and Baumann (2010) have shown that different pitch accent types in German have an impact on expectation-based processing. In this study, the authors tested utterances (*Sabine repaired an old shoe. In doing so, she cut the sole.;* example translated from German) in which a referentially accessible target word (*the* *sole*) was presented in three prosodic conditions: bearing a context-appropriate falling accent (H+L*), which marks the information status of accessibility in German, bearing a shallow rising accent (H*), which corresponds to newness, and being deaccented, which indicates givenness. Results showed a three-way N400 modulation reflecting the degree of mismatch between context and prosodic marking (deaccentuation > H* > H+L*; results also showed another effect associated to repair, see 3.7.2). Results indicate that the deaccentuation of accessible referents yields different effects compared to the accentuation (by shallow rising or falling accents), as it usually marks referential and lexical givenness. Moreover, results showed that the falling accents (H+L*) were perceived as more acceptable as marking accessible referents than shallow rises (H*), in line with the higher co-occurrence of the latter accents with referentially and lexically new information (Röhr & Baumann, 2010).

This subsection has discussed how the violation of semantic information, world knowledge as well as prosodic expectations elicits N400 effects. Particularly new in the language processing landscape are the studies proving the effects engendered by prominence relations and showing that prosodic prominence cues guide real-time processing. The next subsection will consider ERP components other than the N400, which have also been found to be sensitive to prosody.

3.7.2 Other ERP components

This subsection considers other prosody-related ERP effects, in order to provide a more complete overview on the neurophysiological correlates of prosodic prominence, as well as to link these effects to attentional processing. Together with the N400 component, in particular the Late Positivity response will be of interest for the experiment in Chapter 7, as it is connected to signal-driven attention orienting. Throughout this subsection the different components will be briefly described, while the next section (3.8) will deepen the concept of attention and the effects on processing.

One of the components that has been associated with processing prosodic information is the P200 component, which represents a positive peak around 200 ms post stimulus onset. This component has been found to be relative to the processing of different pitch contours (Friedrich, Alter & Kotz, 2001). Heim and Alter (2006) found a P200 to be elicited by sentence-initial words bearing pitch accents in comparison to unaccented sentence-initial words. The authors found that the processing of accents in utterances in isolation varies as a function of their position in the utterance: while utterance-initial accents are recognised early and elicit a P200, the process of accents occurring later in the utterance depends on whether the accent is expected or not, revealing two different negativities, an early negative deflection (Expectancy Negativity, see also Hruska & Alter, 2004) and a later negativity (N400), respectively.

For unexpected inappropriate or missing accents in addition to the N400, a Late Positivity has been observed, indicating the need to resolve the conflict between discourse and prosody (Baumann & Schumacher, 2012; Schumacher & Baumann, 2010; see also Brouwer & Hoeks, 2013; Dimitrova, et al., 2012; Li et al., 2018; Magne et al., 2005; Toepel et al., 2007). Late Positivity represents a positive waveform beginning around 400–500 ms after stimulus onset and usually reaching its maximum amplitude around 600 ms post stimulus onset. It usually has a centro-posterior or (left) anterior distribution over the scalp. In the case of Schumacher and Baumann (2010, 2012), a Late Positivity emerged in response to deaccentuation on an accessible entity (deaccentuation > H*/ H+L*). This reflects a difficulty in integration (i.e. in the update of the mental model), since deaccentuation inherently signals that the entity is already given in the discourse, but in the stimuli used it represented new information. This indicates that the resolution of the conflicting information form prosody and discourse is computationally demanding.

As shown in the previously reported study by Baumann and Schumacher (2010, 2012), the Late Positivity response can be associated with the update of the mental model, which is usually required by attention orienting mechanisms (e.g., Ruchkin, Johnson, Canoune, Ritter & Hammer, 1990; Squires, Squires & Hillyard, 1975), namely mechanisms that prioritise the processing of a particular input. The concept of attention and the relative effect on processing are deepened in the next subsection. The Late Positivity component is part of the P3 family (see in particular Coulson, King & Kutas, 1998 and Sassenhagen, Schlesewsky & Bornkessel-Schlesewsky, 2014 understanding the P600 as a type of P3b), a family comprising positive ERP deflections with a peak that can occur between 250 and 700 ms post stimulus onset (the range can vary depending on stimulus modality, task condition age of the subject, cf. Polich, 2007:2129). This P300 component has been shown to be triggered by attentive processing of less probable or novel stimuli and by memory storage during discourse updating (Kok, 2001; see Polich, 2007 for a complete review; see also Gray, Ambady, Lowenthal & Deldin, 2004 and Roye, Jacobsen, & Schröger, 2007 for elicitation in hearing one's own name and one's phone ringing tone, respectively). These two processes are distinguished by scalp distribution of the component: P3a distributed over fronto-central scalp regions and related to attentive processes; P3b distributed over posterior scalp regions and memory-related.

The P300 is associated with the ease of identifying a visual and auditory stimulus as the target and to store it in working memory. Even though the P300 component does not directly reflect the processing of linguistic stimuli, it may be elicited during attentive speech comprehension (see 3.8.3 and Röhr, Brilmayer, Baumann, Grice & Schumacher, 2020; see also task-related P300 effects in e.g., Roehm, Bornkessel-Schlesewsky, Rösler & Schlesewsky, 2007). For example, Röhr et al. (2020) have found that rising accents show an early positivity (associated with the P300 component family) which falling accents don't show. With rising accents being more prominent than falling accents, these results show that more prominent accents attract more attentional resources. A similar result has been found in the preceding study by Magne et al. (2005), which showed that focal accents evoke an early frontally distributed positivity (P3a), connected to the orientation of attention. Attention and the P300 component will be further discussed in section 3.8.

As the present section has indicated, the modulation of the ERPs has been proven to be affected by attentional mechanisms, which, in turn have been argued to be modulated by prosody. The next section will elaborate on the link between prosody and attention. This discussion is particularly relevant for the experiment in Chapter 7 where the modulation of attention as a function of fine-grained differences in prominence will be investigated.

3.8 Relation between attention, prosodic prominence and focus

In the previous section the relation between prosody and attention has been briefly outlined, the concepts and the research relative to them will be deepened in this section. The concept of discourse prominent information and prosodic prominence-lending features viewed as attention orienting devices will be of particular interest. The impact on attention of different degrees of prosodic prominence in different positions of the information structure will be the object of Chapter 7.

The term attention typically refers to processes that control the flow of information through the nervous system (cf. Luck & Kappenman, 2012). Almost all ERP components are influenced by attention: attended stimuli elicit larger ERPs in comparison to non-attended stimuli (Luck & Kappenman, 2012). If an ERP component is larger for a particular stimulus because attentional resources are devoted to its processing, the processing depth of this very stimulus is greater compared to another which does not show this effect. Attention is considerate to operate

> within both the auditory and visual modalities as a *sensory gain control* — like the volume control on a sound system or the brightness control on a video monitor — that serves to boost the effective intensity of the attended stimuli and reduce the effective intensity of the ignored stimuli (Luck & Kappenman, 2012:297).

Some researchers have argued that attentional mechanisms are used to overcome the overload of inputs that the sensory system has to face by limiting processing to a subset of inputs, which will be more accurately perceived (cf. Luck & Kappenman, 2012). This process corresponds to selective attention, which, driven by expectations and current goals (top-down mechanisms), modulates cognitive processing into a deeper processing of task-relevant information, while ignoring other, i.e. irrelevant, information. For example, Karns, Isbell, Giuliano and Neville (2015) defined auditory attention as the ability to block out sound and noise by attending to some stimuli and ignoring others occurring at the same time. According to a classical model (Corbetta & Shulman, 2002), this top-down attentional selectivity is controlled by a parieto-frontal brain system, the dorsal attention network (DAN), which allows orienting, namely the ability to prioritize sensory input (see Petersen & Posner, 2012 for an accurate and extended review).

However, irrelevant information for the current task may be dangerous to miss. To provide for such an incidence, a second system is needed: the bottom-up attentional capture (or ventral attention network, VAN). This allows for automatic or reflexive (re)orienting of attention due to an unexpected stimulus: signal-based processing of an unexpected stimulus, which may coincide with new, highlighted or potentially dangerous incoming cues, causes the (re)orienting of attention (Corbetta & Shulman, 2002). Therefore, the appearance of a perceptually salient stimulus can lead to an alert state, that prevents the stimulus from being ignored: “Unexpected, novel, salient and potentially dangerous events take high priority in the brain, and are processed at the expense of ongoing behaviour and neural activity" (Corbetta & Shulman, 2002:201). However, top-down attentional processes can still minimise or inhibit this stimulus-driven attention orienting.

Taking as a starting point the notion of information packaging of Chafe (1976; see 2.2), Bornkessel-Schlesewsky and Schumacher (2016) have argued that the speaker's strategies to accommodate the needs of the addressee (i.e. prosodically focussing a word) "is indicative of the possibility to influence attention orienting in accordance with the current communicative aims" (Bornkessel-Schlesewsky & Schumacher, 2016:585). In fact, a P300 effect has been observed during the processing of new focussed information elicited by questions in reading tasks (Bornkessel, Schlesewsky & Friederici, 2003). The P300 component has been interpreted to correlate with the bottom-up reorientation of attention and, therefore, to reflect the reaction to the critical current event (cf. Bornkessel-Schlesewsky & Schumacher, 2016; see Nieuwenhuis, Aston-Jones, & Cohen, 2005). This reaction is represented by an increase of processing of the critical event, which prevents the event to be misinterpreted or ignored.

Studies have argued that focused information attracts attention and allows for more careful encoding of the information, reflected by longer reading times in comparison to non-focussed information (Birch & Rayner, 1997) and faster detection of the change from a word to a semantically related one (Ward & Sturt, 2007). In addition, Cutler and Fodor (1979) demonstrated for English that the same holds for the auditory domain (see also Akker & Cutler, 2003 for evidence on English and Dutch). Their stimuli consisted of utterances with the same prosodic pattern that were preceded by questions that evoked different focal structures. Results revealed that when the critical word was part of the focus domain, the response to the task (target phoneme detection) was faster. This faster detection is considered to be the result of listeners exploiting previous information to ease the search for the most central portion of the message. This responds to the need of identifying the focus of the utterance in order to understand what the speaker has said.

Sanford, Molle and Emmott (2006) have argued that prosodic prominence is an attention-capture device which modulates discourse processing. In their study, participants listened twice to short discourses and had to identify which of the words changed in the second presentation. The target words were realised either with a contrastive accent or with a non-contrastive accent: heightened prominence of the contrastive accent led participants to more accurately detect the changes in comparison to the less prominent non-contrastive accent. The authors interpreted this result as indicating that listeners draw more attention to the more prosodic prominent words, namely that accent can modulate the selective attention of the listeners. Their findings were in line with others (Cutler & Foss, 1977; Cole, Jakimik & Cooper, 1978), showing faster phoneme recognition and easier detection of mispronunciations in words that were focussed (answering a preceding question) and received a pitch accent.

Another noteworthy aspect of the advantage that intonation offers to processing is that listeners appear to make use of the whole contour to actively direct attention to the upcoming parts of an utterance where the accent will fall. Cutler (1976) analysed utterances in isolation realised with different focal structures. She compared normal versions of the utterance with versions in which the critical word was cross-spliced in order to be prosodically the same for each condition. Results showed that also for the spliced utterances listeners detected the target faster in the condition in which the embedded utterance implied the occurrence of an accent on the critical word than in the condition in which the embedded utterance implied the deaccentuation of the critical word. This difference is present in spite of the critical words being acoustically identical. Listeners then exploited the information encoded in the preceding contour to predict where the accent would occur. This result could be explained by the attention allocation account: listeners actively directed attention to the part of the utterance where the preceding information of the F0 contour suggested that the accent will have occurred. This account implies that when new information is accented and given information deaccented the correct allocation of attention towards the new information facilitates processing. By contrast, when a mismatch between information structure and prosodic marking occurs, more attention would be allocated to the less important part of the message (the given information) and this would lead to difficulties in processing. This account would consider accent as an attention-orienting device and could explain the difficulty in processing the mismatch between accent and information structure (see also the results of Baumann & Schumacher, 2012; Schumacher & Baumann, 2010).

This subsection has pointed to the fact that attentive stimuli are processed deeper than non-attentive stimuli. The increased depth of processing is reflected by the enhanced amplitude of the ERP components elicited by the attentive stimulus. Attention is oriented by top-down mechanisms, which allow to attend to task-relevant information and to ignore task-irrelevant information. However, the attentional system can be reoriented through bottom-up mechanisms. Studies have shown that prosody can function as an attentional orienting device, including both top-down and bottom-up mechanisms: listeners actively direct attention where the unfolding of the pitch contour indicates that the focussed information will occur (top-down) and at the same time, as the decoding of the signal proceeds, orient more attention to the more prosodically prominent parts of the utterance (bottom-up). Relating to this conception of prosodic prominence as orienting attention, the experiment in Chapter 7 will try to answer the question on whether fine-grained cues to prominence, usually associated with background (i.e., the post-focal pitch accent occurring in questions of the Bari variety, see 2.3.3) can play a role in the orienting of attention. The next section will provide a clearer picture of this question, as it will deepen the understanding of the effects on prominence of the interplay between information structure and prosodic prominence.

3.8.1 Interplay of information structure and accentuation on semantic processing: N400 amplitude as a function of focus and pitch accent

As seen in the previous subsection, attentional orienting mechanisms guide processing and permit relevant information to be processed more deeply compared to non-relevant information. The present subsection will concentrate on the consequences of attention orienting on semantic processing. In particular it will underline the role of information structure and its prosodic marking in orienting attention and consequently manipulate the processing of semantic information.

A crucial observation for understanding the function of attention (re)orienting is that linguistic input is only partially analysed (cf. “good enough processing”; Ferreira, Bailey & Ferraro, 2002). This partial analysis is a strategy adopted to cope with the overload of information that our sensory systems are exposed to. As a consequence, under certain circumstances semantic anomalies and inconsistencies are not noticed. This effect has been called “semantic illusion” (referred to also as “Moses illusion”) and has been first discovered in an experiment by Erickson and Mattson (1981). In this experiment participants’ task was to read sentences containing a subtle anomaly with respect to world knowledge such as the following: *How many animals of each kind did Moses take on the*

Ark? In the experiment 48% of the readers failed to notice the anomaly contained in the sentence, meaning that it was Noah and not Moses to take the animals on the ark. This is a semantic illusion due to the semantic relation between Noah and Moses. This illusion is an example of how readers may adopt a “good enough” processing strategy to interpret a message (Ferreira, Bailey, & Ferraro, 2002; Sanford, 2002), namely that for purposes of cognitive efficiency, readers tend to extract the information that is needed for the current communicative situation while overlooking some parts of the input. This means that the depth of processing, a concept often used in the research on language comprehension (see Sanford & Graesser, 2006) and related to the amount of attention recruited to process a particular word or constituent, is not the same across the entire language input. In this context, anomalies in the semantics of a word may only be processed in a shallow manner. In Sanford and Grasser’s terminology, shallow processing is described as involving an incomplete representation of linguistic input, that could be good enough for the task.

A number of studies support the strong bond between linguistic focus and attention in showing that information structure modulates the semantic illusion. Bredart and Modolo (1998) conducted an experiment on French, asking participants to answer whether a statement was correct or incorrect. They compared statements such as *Moses put two animals of each kind on the ark* with the construction *It was Moses who put two animals of each kind on the ark,* which focus the name *Moses* through a cleft structure. They found the detection rate to be higher in the cleft sentences, indicating higher attention corresponding to focus. A more direct index for the underling cognitive process was the study of Wang, Hagoort, & Yang (2009). They investigated the modulation of information structure in on-line processing through an ERP study on Mandarin Chinese written sentences. In the experiment, the authors used *wh*-question-answer pairs to elicit focus either on the critical word or on another word in the sentence (critical word out of focus). The semantic appropriateness of the critical word to the context was further manipulated: the question *Who bought the vegetables for cooking today?* was followed either by the answer *Today Xiao Ming bought eggplant to cook* (with the word *eggplant* appropriate to the context) or by the answer *Today Xiao Ming bought beef to cook* (with the word *beef* inappropriate to the context; cf. Wang et al., 2009:55). The authors measured the N400 effect, indicated by the difference between the N400 elicited by the inappropriate and the N400 elicited by the appropriate condition. They found that while the inappropriateness elicited a large N400 effect for the focused words, it evoked a very reduced N400 effect relative to the non-focused words. This indicates a decrease in processing resources allocated to non-focused information.

Interestingly, non-focused information in the experiment came after the focused one, suggesting that the strategy adopted during language processing might be to skip the extra information after having encountered the requested one. This could be due to the expectations in finding either a full anaphora or a pronoun, both of which do not require to be attended to for understanding the message. This interpretation is related to the limitation of cognitive resources, which are, therefore, directed to the new information. Moreover, Wang et al. (2009) found that the N400 elicited by appropriately focused information was less negative than the one of the non-focused information, which suggests that the context induced an expectation regarding the position in which the new information would appear, helping semantic processing of the critical word.

In the auditory domain, a neurophysiological study on Dutch (Wang et al., 2011) showed that when a semantically incongruent word in an answer occurs in a focus position and is marked by an accent, it elicits larger N400 effects than semantically incongruent words in focal or post-focal positions that are unaccented. These results support the previous findings that context and accentuation interact. Thus, accented focused words are processed more deeply compared to conditions where focus and accentuation mismatched, or when new information is not prosodically marked. Accordingly, listeners allocate their attentional resources to the focus of the utterance with the consequence of being less aware of anomalies that occur in post-focal position. Therefore, it seems that an earlier focal accent draws attention away from deaccented words in the post-focal domain, leading to shallower processing of this domain.

However, findings of Wang et al. (2011) might not be a direct consequence of accent, but rather of focus. In fact, accented and deaccented words differed not only because of the presence or absence of an accent, but also in their information structure: the preceding context created a distinction between focused and non-focused information. Evidence that focus can on its own increase attention prevents a thorough understanding of the attentional mechanisms guided by accentuation only. Given this, the top-down and bottom-up processes could not be disentangled properly (Cutler & Fodor, 1979, where cross-splicing made the focused words prosodically identical to the non-focused; Birch &

Rayner, 1997; Wang et al., 2009; Ward & Sturt, 2007 where the prosodic information was not provided).

This subsection has reported on studies showing that focussed information attracts attention and prevents information to be shallowly processed. The studies here presented, despite referring also to the accentuation and the deaccentuation of the investigated semantic information, have not disentangled the effects of top-down from bottom-up processes. In particular, they have not shown the effects of prosodic cues on the processing of incongruences. This issue will be discussed in the following subsection and will be of interest for the experiment in Chapter 7.

3.8.2 Influence of prosodic prominence (bottom-up cues) on depth of semantic processing

Studies presented in section 3.8.1 were concerned with the depth of processing of semantic information in relation to information structure, in particular with the depth of processing of focus. However, they did not account for the depth of semantic processing influenced solely by bottom-up cues. The ERP studies mentioned until now argued that the role played by accentuation in spoken language comprehension is fundamental in influencing the ease by which the current discourse information is processed. However, the specific mechanisms by which accentuation, independently of focus, affects spoken language processing in real time have not yet been discussed. These mechanisms still remain a matter of debate for research, though some studies have started to shed light upon them.

In the previously discussed studies by Sanford et al. (2006) and by Cutler (1976), the authors have argued for an account of accentuation which entails that accentuation alone can modulate listeners' selective attention during speech processing. These studies (together with other previously discussed ones) are, however, primarily taking into consideration the top-down aspects of processing rather than bottom-up aspects, having either utterances presented with context or expectations built up by the previous part of the contour in the utterance (see also Schumacher & Baumann 2010, 2012).

By contrast, Li and Ren (2012) examined whether acoustic prominence alone could modulate selective attention during on-line resolution of semantic mismatches, disentangling the role of information structure and the bottom-up cues. They took into consideration the distinction between different degrees of prominence. In their experiment, they analysed the neural correlates of auditory stimuli heard by native Chinese speakers. Their stimuli consisted of (Chinese) utterances in isolation in which the critical word had three different degrees of prominence and was either semantically congruent or incongruent within the utterance. Crucially, the utterances presented to participants did not provide a context generating previous expectations regarding focus structure and the F0 contour preceding the critical word was kept constant in all the conditions, thereby preventing (at least to some degree) expectations driven by the preceding part of the contour.

The different degrees of prominence of the critical word (called by the authors Accent, Great Accent and DeAccent which corresponded to emphasis, great emphasis and no emphasis respectively) were realised by recording the utterances in three different question-answer contexts. The questions were used only in the recording phase in order to elicit critical words with different acoustic features in the answers. The first context led to the marking of new information focus through an increase in prominence (Accent condition), which consisted in an expansion of the pitch range of the lexical tone and by the lengthening of the syllable duration (strategy for Chinese to realise focus). The second context was instead created from a misunderstanding of the first answer, which led to a repetition of the question and a subsequent repetition of the answer, in order to be more clearly understood by the listener (who did not understand the first time; GreatAccent condition). The last condition was the one in which the critical word was attenuated and was created by the exact repetition in the answer of the word in the question (lexically and referentially given; DeAccent condition). Syllables of the critical words increased in duration, pitch maximum and pitch expansion from the attenuated condition to the relatively prominent condition (Accent) and from the relatively prominent condition to the most prominent condition (GreatAccent). The authors found an N400 effect in both the prominent conditions (Accent and GreatAccent), with the more prominent condition (GreatAccent) revealing a broader distribution of the effect. Moreover, the more prominent condition elicited a larger N400 effect over the frontal-central electrodes. By contrast, no significant N400 effect was found for attenuation. These results suggest that an increase in prominence increments the depth of information processing, subsequently leading to an increase in the allocation of attention towards more prominent information. On the contrary, non-prominent information is subject to shallow processing.

In support of these results are the findings by Kristensen, Wang, Petersson and Hagoort (2012), who showed that pitch accents (in Dutch) recruit parts of the general attentional network in the brain. Their experiment investigated the processing of the Moses illusion as modulated by the accentuation patterns in utterances in isolation. The authors compared the areas of the brain activated by the processing of the linguistic auditory stimuli with the ones activated during a spatial attention task (conducted with non-linguistic auditory stimuli). Therefore, they first localised the attention network and subsequently investigated whether pitch accents would also activate it. Critical words were presented as either accented or non-accented and could either be congruent or incongruent within the utterance. The authors found larger activation of the areas corresponding to the non-linguistic auditory spatial attention network when the words were realised with pitch accents compared to when they were deaccented. Thus, they have proposed an account in which prosodic marking of new information recruits the attentional network in the brain, in order to increase the processing of the constituent that is associated with this marking. They have argued that this strategy serves to prevent neglecting the most relevant parts of the linguistic input, in line with the findings regarding the interplay between accentuation and information structure, which have been discussed in the previous section.

According to the attention account, when information is more prominent, more attention is drawn towards it, allowing deeper processing in comparison to less prominent information. Further evidence on deeper processing of more prominent information is the higher ease with which words marked by more prominent accents are retained in memory and more accurately recalled in comparison to ones marked by less prominent accents (see Fraundorf, Watson & Benjamin, 2010 that compared the effect of H* accents and L+H* accents on memory). Generally, it is important to observe that not only pitch movements facilitate deeper processing, but also characteristics of accented syllables, such as increased duration and greater articulatory precision, factors that make the syllable acoustically clearer and, hence, easier to process. However, in the EEG experiment of the present thesis (Chapter 7), the post-focal compressed rise in pitch will be of particular interest. The next subsection will, therefore, present the effects of rises in pitch in the orienting of attention.

3.8.3 Rises are special in attention orientation

F0 rises are central to the understanding of the role of (bottom-up) cues to prominence and attentional orienting. As underlined by Hsu et al. (2015), pitch rises compared to steady F0 or to pitch lowering occupy a privileged role in spoken communication. Speakers use sudden rises to arouse listener attention, for example in contrastive accents, in beginnings of units of discourse and, even though not in all languages and varieties, in questions seeking a response. Pitch rises are also more likely to be used in the expression of intense emotion, fulfilling the function of attracting and maintaining listener's attention. For example, paralinguistically, across languages and cultures, happiness, anger, fear and surprise tend to be expressed by sustained elevated F0, usually involving an initial rise (see e.g., Pell, Monetta, Paulmann & Kotz, 2009; Schröder, 2001).

In the study of Hsu et al. (2015) the P300 component showed sensitivity to rises in F0, whereas there was no effect of attention orientation for F0 falls, suggesting that attention modulation mechanisms are triggered more when listeners perceive sudden rises in spoken F0, compatible with the fact that human pitch discrimination is more sensitive to F0 raising than lowering: the threshold for detecting pitch rises is lower than for detecting pitch lowering (Kishon-Rabin, Roth, Van Dijk, Yinon, & Amir, 2004). These results, together with Li and Ren's (2012) and Fraundorf et al.' (2010) ones, lead to the conclusion that more prominent accents, characterized by steeper rises, contribute greatly to discourse comprehension through the detection of importance of information, which crucially ought not to be missed. However, these mechanisms can be language-specific, since different languages can make different uses of the degrees of prominence for linguistic functions and these can influence the perception of acoustic cues. The next and last section of this chapter will discuss language-specific differences, with a particular interest in Italian.

3.9 Language-specific differences in the perception of prominence

3.9.1 The role of native language in prominence perception

Language-specific expectations in the perception of phrasal prosodic prominence have not been extensively explored yet and a clear picture is not available. While the role of transfer from native language prominence patterns to the non-native language is rather well attested in production (see Dahmen, 2013; Mennen, 2004; Nava, 2010; Swerts &

Zerbian, 2010; van Maastricht, Swerts & Krahmer, 2013 among others), how the native language affects perception is still a matter of debate, since the field of non-native prosody perception is relatively new. Since one of the objectives of the present thesis is to compare native with non-native perception of different degrees of prominence, this section will summarise the literature referring to this topic.

Some research has been done in the domain of word level stress perception. Among others, the study by Lehiste and Fox (1992) has shown differences in the sensitivity towards acoustic cues in native speakers of different languages (see also Ortega-Llebaria, Gu, & Fan, 2013 for an experiment involving English and Spanish native speakers and for a brief review on the literature; the interested reader is also referred to Peperkamp, Vendelin, & Dupoux, 2010; Wang, 2008). However, this chapter is concerned with phrasal prominence, which also needs to be seen in relation to information structure. In fact, as previously discussed, when rating phrasal prosodic prominence, participants are always aware of the mapping to functions, which can influence the judgment of prominence (top-down inferences). Therefore, it is important to understand the relation between prosodic form and function in the specific language analysed.

In the understanding of the role of the first language (L1) prosodic patterns in the perception of phrasal prominence, the view of the probabilistic association approach can be particularly interesting. Listeners' expectations are based upon the probabilistic distribution in their native language and these expectations can then be wrongly or not precisely enough mapped onto functions in the non-native language. If two languages mark functions by different prosodic strategies the perceptual use of intonation information from one language to the other can be different. This would lead to think that listeners might have difficulty in processing prosody in their non-native language, and in fact evidence supports this view.

For instance, Akker and Cutler (2003) found that native speakers of Dutch, while showing an effect of increased speed in phoneme recognition in accented words in comparison to non-accented words when performing the task in their native language, show this effect to a lesser extent when performing the task in English (despite being highly proficient in the language). This is also somehow surprising because English and Dutch are similar in the marking of prominence (and focus in particular). Indeed, both groups of participants (native and non-native) produced similar patterns in their respective native language and non-native speakers also showed to some extent the predicted effect of accent. This means that they can effectively make use of the marking of prominence in their second language, even though they are not able to map the intonation information onto another level of processing (phoneme recognition) as rapidly as in their native language.

In addition, Baker (2010) investigated the understanding of pitch accent placement in English by Mandarin and Korean speakers, arguing that both groups of non-native listeners had difficulties in distinguishing whether the prosody used for an utterance was correct in respect to the context (context eliciting broad and narrow focus). These results were the consequence of the differences in the use of pitch between the three languages both in terms of forms of lexical and phrasal prominence and in terms of the linguistic meaning of the contours. Further, the author found that the accuracy in performing the task improved with the increase of English proficiency. Similarly, Cruz-Ferreira (1984) demonstrated that native English learners of Portuguese and native speakers of Portuguese who learned English were similarly ineffective in reaching native-like accuracy in matching pairs of utterances, differing only in intonation, with their intended meaning.

Further support for these findings is the study by Swerts and Vroomen (2015), concerning the perception of prominence of two groups of listeners in their native language when the two languages have a different prosodic structure and a different strategy to mark focus and information status. The authors investigated whether native speakers of Dutch, a language in which the information status of elements inside a noun phrase is indicated by intonation, and native speakers of (Canadian) French, a language which does not make use of intonation to mark information status inside noun phrases[13], differed in the processing of noun phrases presenting different accent distributions. Results indicate that French listeners did not make use of intonation information to have an advantage on the processing of the utterances, whereas Dutch did. According to the authors, this was a consequence of the typological difference in the prosody of the two languages. Furthermore, van Maastricht, Swerts and Krahmer (2014) investigated the perception of non-native speech (Spanish native speakers speaking Dutch) by native speakers (of

[13] The description the authors make for French is similar to the one that Swerts et al. (2002) make for Italian, see 2.3.2.

Dutch), finding difficulties in the processing of non-native speech. These difficulties related to the different distribution of accents in the marking of new and given information in non-native speakers' speech. These results are in line with the difficulty in the processing of incorrect prosodic marking of information structure, as discussed in a previous section (3.6.2).

On the other hand, when asked to mark prosodic prominence only (without referring to its relation to functions), the influence of the native language on prominence perception seems to decrease, even if the small amount of literature on the topic does not provide a clear picture. For instance, Pintér, Mizuguchi and Tateishi (2014) have found that prominence scores of natives (English) and non-natives (Japanese learners of English) show a surprisingly higher correlation than expected and that the reliance on acoustic cues in both cases is similar. Yet, non-native speakers were not successful in reaching native-like performance, because it can be assumed that to native speakers a wider range of prominence encoding strategies is available. By contrast, Yamane, Yoshimura and Fujimori (2016) have contradicted these findings, arguing that the Japanese's perception of prosodic prominence in English approximates that of native speakers. Differently, Huang (2004) has showed that (American) English native speakers, whose task was to discriminate tones in Mandarin, considered the onset height of the pitch, while native listeners considered the contour as a whole.

Grabe, Rosner, García-Albea & Zhou (2003) have investigated whether a group of native English speakers compared to a group of Spanish native speakers and Mandarin native speakers would perceive similarities and differences in different pitch contour shapes in the same way. The three groups distinguished rising from falling contours, but showed differences in the distinction of contours within the group of rising and falling accents. The authors have concluded that common auditory mechanisms are primarily important for the basic perceptual distinction (common interpretation of bottom-up cues), while the experience in a native language can then build on these mechanisms and yield different outputs (some degree of cross-language specificity; see also Shport, 2015). Eriksson, Grabe and Traunmüller (2002) have supported this argument, showing that learners of a language perceive prominence in a similar way in respect to native listeners, but the perception of prominence changes when there is conflict between the expected prominence and the actual production of prominence. In this study participants were required to distinguish prominence at a more fine-grained level, since the ratings were provided on a slider with continuous values, which allowed them to indicate gradient differences between the syllables rated.

The present section has underlined the lack of a clear picture on the prominence perception of non-native prosodic prominence. The scarce literature on the topic seems to reveal that while the native language has an influence in the mapping between prosodic realisation with functions, it has none, or at least to a lesser extent, in the perception of only the degree of prominence of an entity. However, on this latter point, results are mixed, some showing that the native language experience plays a role in the interpretation of the bottom-up cues. The experiment in Chapter 6 will further the understanding on this matter, testing the hypothesis that the different distribution of prominence and the different probabilistic mapping between prominence and functions in the native language can influence the perception of the degree of prominence. In addition, results of the last study here discussed (Eriksson et al., 2002) suggest that more gradient distinctions on prominence perception could be generally more informative in the studies on non-native perception of prominence. Therefore, studies employing a continuous scale to rate prominence could uncover differences that might not be uncovered by standard RPT methods in which participants distinguish only the prominent syllables from the non-prominent ones (as already suggested for acoustic studies in 3.2.1). The use of such a continuous scale could be particularly interesting in the non-native perception of prominence in post-focal position in Italian. In addition, it could also be more informative regarding native perception. The next section (3.9.2) will deepen the discussion on native prominence perception in Italian.

3.9.2 Perception of prominence in Italian

Chapter 2 has already underlined the differences in the phrasal prominence production in Italian in comparison to German (and West-Germanic languages). In particular, the differences between these languages arise in the prosodic marking of information structure and in the distribution of prominence in the post-focal position. Specifically, Italian (especially some varieties, see 2.3.2) is more variable in the allocation of prominence in the post-focal position and in the prosodic marking of elements' information status. Taking into consideration bottom-up cues to prominence and the expectation-based processes, these differences in production can be assumed to play a role in the perceptual domain. This could particularly hold true under the *inference under uncertainty* hypothesis (see e.g., Clayard et al., 2008; Kleinschmid et al., 2015; Norris et al., 2003). If the building of expectations is realised through probabilistic knowledge (3.4), expectations of prominence in the post-focal position could be different in the two languages (Italian and German), since the probabilistic association between prominence and post-focal position, and between givenness (lexical and referential) and deaccentuation might differ.

In comparison to the literature available for West-Germanic languages, few studies have been conducted on the perception of phrasal prominence and on the mapping between intonation and information status (or information structure) in Italian. An exception is the study by Krahmer and Swerts (2006) which has argued that Italian participants, contrary to Dutch participants, are not able to identify the context in which a noun phrase is realised. However, this study was conducted on data collected in a previous production experiment (Swerts at al., 2002; see also 2.3.2), in which speakers, despite the different information status of the elements in the noun phrase, did not make a distinction in the F0 contours among conditions (different conditions were all produced with a flat hat shape contour). Although this might be the standard production for Italian (with no marking of the information status within the noun phrase, see discussion in 2.3.2), the result of this experiment only shows that the indistinguishable contours from the production study also fail to be distinguished in perception, namely that there are no other cues that allow listeners to trace back the previous context.

Similarly, not much work has been conducted to investigate whether Italian listeners correctly map the prosodic realisation to the focal structure intended by the speaker (as for instance the studies discussed in section 3.3, regarding West-Germanic languages). An exception are the works by D'Imperio (on the Neapolitan variety of standard Italian; 1997, 2001). In one of her studies, D'Imperio (1997) has shown that Italian (Neapolitan) listeners make use of prosodic information to reliably distinguish the context eliciting (uttered) statements in broad focus from the one with a late narrow informational focus. The ambiguity given by the occurrence of the nuclear accent in final position for both focal structures is overcome by the perceptual distinction of the different accent types marking the two utterances. Although the ambiguity still remains for late nuclear accent questions (narrow focus structure not successfully distinct from broad focus structure), this study provides evidence for the success of Italian listeners in discriminating speakers' intentions through prosody, at least in the distinction between contrastive information and all-new information.

However, for Italian there are no studies that systematically investigate the degrees of perceived prominence of different accent types and different degrees of prominence. In addition, there is only one study (D'Imperio, 2001) that has investigated the perception of post-nuclear accents. More precisely, it has concerned the perceived prominence of a compressed rising accent occurring after the nuclear accent in questions uttered in the variety of standard Italian spoken in Naples. In this study, listeners were asked to mark the word that appeared to be most "important" between the two or three words that could make up the utterances. Interestingly, D'Imperio found that the word bearing a post-focal compressed accent was not judged by listeners as the most important element of the sentence. The constituent perceived as the most prominent was the one associated with the fully-fledged accent, corresponding to the focus of the sentence. As the author said, this result suggests the need to redefine the notion of nuclear accent as being at the same time the perceptually most prominent accent and the last accent occurring in the Intonation Phrase (see 1.3.1.2 and see the revised definition of nuclear accent described as the last fully-fledged accent occurring in the Intonation Phrase provided in Grice et al., 2005). Surprisingly, these results have shown that Italian is to some extent similar to German (Baumann, 2014; Baumann & Röhr, 2015; Ayers, 1996), in that post-focal constituents are regarded as having a significantly lower degree of prominence than the ones bearing a contrastive accent.

Although it is important to point out that D'Imperio's (2001) experiment helped in the building of a revised definition of nuclear accent, there is the need to underline that D'Imperio's experiment was limited to a discrete evaluation of prominence. Listeners were in fact asked to perceptually judge only the most prominent word in the utterance. By contrast, the perception of prominence in the post-focal position might be better described by a more fine-grained distinction of the degrees of prominence. This might reveal that the word bearing the post-focal compressed accent, although not the perceptually most prominent one in the utterance, could still be perceived as high in prominence in relation to other words before the focus or to words not bearing a compressed accent.

Concerning the on-line processing of prosodic prominence, the ERP studies presented in the previous sections have all been conducted on West-Germanic languages, with the exception of one on French (different from West-Germanic languages in the realisation of information structure). The latter yielded results compatible with the ones on West-Germanic languages in regard to the mismatching of prosodic input and context. However, this showed the effect of the incorrect positioning of the accent on the non-focused word or the absence in the focussed word after an information seeking question. By contrast, a study looking at the relation between given and new information with prosodic prominence (Swerts & Vroomen, 2015) has suggested that the prosodic information signalling the given/new element inside a noun phrase is not used by (Canadian) French listeners in real time (see 3.9.1). Therefore, the lesser use of prosodic information to mark information status of the elements seems to have an influence on processing and the processing load associated with utterance comprehension. If Italian does not systematically rely on the prosodic marking of information status the hypothesis could be that its behaviour would reflect the one of French. However, no studies are currently available. Similarly, no ERP study concerning the processing of phrasal prosodic prominence has yet been conducted in Italian.

Nonetheless, an ERP study on the processing of German utterances by Italian learners of German (Zappoli, Vespignani, Baumann, Grice & Schumacher, 2018) is available. This study was concerned with the processing of German utterances by a group of native speakers of German and a group of native speakers of Italian learning German as their L2. The stimuli had the same structure (but different lexical items) as the ones in Schumacher and Baumann (2010), investigating the prosodic marking of different degrees of information status. Results of the study showed that the three different prosodic conditions of the stimuli (falling accent, rising accent and deaccentuation) on the accessible information elicited a three-way-modulation of the N400 in the L1 group, whereas in the L2 group the modulation of the N400 component was only binary, with no effect for deaccentuation. The authors have interpreted this difference between the groups of participants as indicating the absence for the L2 group of a mismatch for deaccentuation, which for the learners, in the authors' view, "is not yet integrated in the processes that deal with information status and affect the N400" (Zappoli et al., 2018:1). They have argued that the expectations derived from the L1 play a role in the processing of the prosodic mismatch, given that prosody and information structure seem to be differently integrated in Italian compared to German (see 2.3.2). However, more research is needed on Italian and on the on-line processing of L2 prosody to reach an understanding of how and why deaccentuation is (or could be) differently processed by Italian listeners. In fact, Italian might differ from West-Germanic languages as a result of the difference in the possibility of the two languages of (probabilistically) associating the prosodic realisation with functions and with information status (see also 1.2.4.2). The experiment in Chapter 6 will test whether the two languages differ, aiming to further the understanding of non-native prominence perception and the role of the probabilistic mapping between prosody and function in prominence perception.

3.10 Summary

This chapter has presented an overview on both behavioural and on-line perception of prosodic prominence, particularly focusing on the features of prominence as an attention-orienting device. It has been shown that phrasal prosodic prominence facilitates the comprehension of an utterance, as it contributes to rapid identification of the semantically most central part of the message. Listeners can in fact make use of prosodic information to discriminate the focus type signalled by the speaker and this information helps the real-time incremental processing of the utterance. In fact, both by the building up of expectations through the incremental processing of the information conveyed by the F0 contour and by the redirection of attention through the bottom-up information, intonation contributes to reduce the cognitive load that the presence of several concomitant stimuli entails, helping to select the most relevant ones. Several factors, such as (language-specific) probabilistic knowledge of the most probable context for specific acoustic features and the global unfolding of the F0 contour contribute to the expectations of where to allocate attention during the processing of speech (top-down processes relative to selective attention), while the presence of (less expected) prosodic prominence at the local level reorients attention (bottom-up processing). This interplay between top-down and bottom-up processes is very important in investigating the perception of prominence. However, despite the large amount of research that has been presented in this chapter, the exact way in which these two mechanisms interact with each other still needs to be completely understood and researched. A better understanding of the role of prominence in top-down and bottom-up processes is needed. In particular, a topic that needs to be more broadly investigated is the influence of language-specific characteristics in prominence expectations, in selective attention and in its reorientation. As previously underlined, an interesting case for a better understanding of the expectation and stimulus-driven processing of prosodic prominence is the case of Italian, which presents post-focal compressed accents. Indeed, these accents could both prevent the post-focal region from being shallowly processed and play a role in building expectations regarding the degree of prosodic prominence in this region. Moreover, the research on Italian is interesting in itself, since literature on the perception of phrasal prosodic prominence is still largely missing for this language.

Chapter 4

Testing prosodic prominence

4.1 Summary of the previous chapters

In the previous chapters, the acoustic characteristics of prosodic prominence and their auditory perception have been delineated. Chapter 1 illustrated that a bundle of different phonetic parameters and their interaction denote different degrees (or levels) of prominence. It also emphasised that the discrete categorisation of the F0 contour and of the phenomena connected to prosodic prominence need to be complemented with the measurement of continuous parameters, and that these two aspects need to be thought of as part of a unique system. The chapter further described three continuous phonetic parameters (synchrony, scaling and PEM) that recent literature has identified as effective for the description of intonation and prominence. These measures are connected to periodic energy and, given that periodic energy is directly related to how pitch is perceived, these measures offer an advantage in characterising the intelligibility of the intonation contour compared to other measures. Thus, synchrony, scaling and PEM, are going to be used throughout the experiments presented in the following chapters (Chapter 5, 6 and 7).

In Chapter 2, the probabilistic nature of the relation between different degrees of prominence and information structure (and information status) was outlined. In particular, a reported difference among Italian and German in the probabilistic mapping between degrees of givenness and prosodic realisation was discussed. Indeed, results from studies on German and West-Germanic languages in general and studies on Italian seem to indicate that the percentage of co-occurrence of given items with deaccentuation is lower for Italian. Furthermore, Chapter 2 discussed the reportedly higher tendency of Italian to modulate word order to place prominent elements at the end of the utterance (or of the phrase) compared to West-Germanic languages. In addition, the chapter presented the difference in the distribution of prominence in the post-focal position between two varieties of Italian, the variety spoken in Udine and the variety spoken in Bari, and between the variety of Italian spoken in Bari and German. These differences consist of a higher probability for the Bari variety to present acoustic characteristics connected to prominence in the post-focal position.

Lastly, Chapter 3 focused on prominence perception, reporting on studies which indicate that not only signal-based cues to prominence have an impact on the degree of prominence perceived by listeners, but also that expectation-based inferences play a role. These expectation-based inferences can be language-specific and can arise from the probabilistic distribution of prominence in the language and from the probabilistic mapping between form and function. In this chapter, the influence of these two types of inferences on online processing was discussed. In particular the impact of expectations coming from the information structure on where the important element will occur in the sentence and the integration of these expectations with the incoming signal were presented. In this chapter studies on online processing of prosodic information, principally conducted using the ERP method, were discussed. They reveal that prosody is integrated in real time in processing and that it functions as an attention orienting device: expectations built from prosodic information direct the top-down allocation of prominence resources, while the signal progressively processed by the sensory system acts as a reorienting device, attracting attention in a bottom-up fashion. Additionally, studies investigating the effect of prosody on the semantic processing were described. Among these, studies showing that focus and its prosodic marking ease the processing of semantic information were especially interesting for the present work. These studies particularly point out that the depth of processing of semantic information is increased by the orienting of attentional resources towards the focussed constituents. By contrast, as a result of these attention orienting mechanisms caused by increased prominence (both acoustic and structural), background, deaccented information is subject to shallow processing.

The delineation of this theoretical background makes clear the general need to deepen some aspects concerning the production and the perception of prosodic prominence. The first question that arises concerns how acoustic cues other than F0 movement are distributed in the utterance and how their interaction contributes to prosodic prominence, especially in the post-focal domain. This question can be formulated as follows: can different degrees of prominence in the post-focal position be signalled by the interaction of different acoustic cues which are not related to pitch excursions, and can these degrees of prominence be perceived by listeners? A second question is related to the role that the probabilistic mapping between form and function and the probabilistic distribution of prominence cues within the utterance play in the perception of prominence and can be formulated as follows: do the inferences that arise from the probabilistic association of a certain form to a function and the inferences derived from the probabilistic association of a certain position in the utterance to certain acoustic cues impact the listener's perception of prominence? This issue is connected to the role of prominence in influencing the expectation-based and signal-based inferences in the perception of prominence. To deepen the understanding of the mechanisms in play in prominence perception, it is crucial to investigate its contribution in guiding top-down (expectation-based) and bottom-up (signal based) processing. Specifically, a question that can further the comprehension of prominence is the role of fine-grained cues in building and modulating top-down processing. These general questions will be addressed in the next chapters through the investigation of more specific research questions, which will be introduced in the next subsections and which concern two varieties of Italian.

4.2 Introduction to the following chapters

In the following three chapters (Chapter 5, 6, and 7), three studies are presented, which concern the prominence relations in the marking of different information structures in two varieties of Italian, the role of top-down inferences on the perception of prosodic prominence and the role of top-down and bottom-up prosodic prominence cues in the orienting of attention. Chapter 5 reports on a production study on the Udine variety of Italian and will serve as the basis for the following perception experiment in Chapter 6, which will also involve the variety of Italian spoken in Bari. Chapter 7, in turn, will be based on the results of the perception experiment in Chapter 6 and test the influence of bottom-up and top-down factors on the real-time processing of prominence.

More specifically, the aim of the study reported in Chapter 5 is primarily to analyse the production of the post-focal position in one variety of Italian in order to assess its degree of prominence in relation to two other focal positions (narrow focus and broad focus). The purpose is to establish, for the variety considered, the prominence relations among elements in utterances with different information structure. The rationale for conducting this production experiment is the presence of studies both reporting the possibility of having a high degree of prominence in post-focal position (Bocci & Avesani, 2011) and reporting that, in some cases, in Italian information status and information structure are not marked with prominence. Given these reported findings the present goal is to assess whether words occurring in different focal structures would be realised with different degrees of prominence. The degree of prominence of a word and its prominence relations within the utterance are here analysed through the modulation of F0 and Periodic Energy Mass (PEM; see 1.3.3). The clarification of the prominence relations between the words in an utterance will help to define listeners' expectations regarding the degree of prominence of words occurring in different focal structures. This will guide the interpretation of results in the perception study in Chapter 6. Indeed, the acoustic characteristics of the stimuli found in the production experiment will be related to the prominence ratings in Chapter 6. In addition, the recordings made for the study in Chapter 5 will serve as stimuli for the study in Chapter 6.

In the study reported in Chapter 6, the perception of prominence by native speakers of two varieties of Italian and by a group of native speakers of German learners of Italian will be investigated. The rationale for the prominence rating experiment is twofold. Firstly, language-specific expectations can play a role in prominence perception. Thus, the different probabilistic distribution of prominence registered between the two varieties of Italian and between the Bari variety and German could be hypothesised to have an effect on expectations of prominence in the post-focal position, which in turn can have an effect on the level of perceived prominence in this position. Secondly, compared to German, Italian (therefore also the two varieties here investigated) is considered to have a different mapping between information structure and information status and its prosodic marking. This probabilistically different association can also play a role in the building of expectations and in prominence perception. Therefore, the aim of the study is to test whether these two different characteristics of the two Italian varieties on the one hand and of Italian and German on the other, will have an effect on the perceived prominence. Firstly, the perception of prominence in the two varieties of Italian is expected to differ because of the different expectations on the degree of prominence in the two varieties. Similarly, the perception of Germans and Bari participants is expected to differ, because of the probabilistic different distribution of prominence in the post-focal position between the two languages, because of the lack of knowledge by Germans of the distribution of prominences in Bari Italian and, possibly, because of the different probabilistic mapping between given elements and attenuation.

Finally, the study reported in the last chapter (Chapter 7) investigates the online perception of prominence in the variety of Italian spoken in Bari. It concerns the reorienting of attention in the post-focal position as driven by prominence. In this experiment the variety of Italian spoken in Bari was chosen because of the presence of post-focal tones in questions. The comparison between the processing of low and flat contours found in statements' post-focal position and the pitch accent found in questions will help to disentangle expectation-based and signal-based processing.

Chapter 5

Production study: prosodic marking of information structure in Udine Italian

5.1 Introduction

This chapter reports the results of an exploratory experiment on the degree of prominence of the post-focal position. The aim of this experiment is to understand the prominence profile of different focal domains in one variety of Italian. This investigation will serve as a basis for the rating study in Chapter 6, which will investigate the perception of the prominence profiles found in the present experiment (part of the utterances of the study in the present chapter will be used as part of the stimuli for the study in Chapter 6). In particular, the interest of both the study in this chapter and in Chapter 6 lies on the post-focal domain. The variety that is here investigated is the variety spoken in Udine.

The present experiment, inspired by Mücke and Grice's (2014) experiment on German, attempts to examine the effects of focus-background structures on the prosodic highlighting of words produced in one variety of Italian. A study on the comparison between unstressed syllables, stressed syllables and accented syllables (in a contrastively focussed constituent) in Italian, investigating the Tuscan variety has already been conducted by Avesani, Vayra and Zmarich (2007). Results of this study have shown that accentuation leads both to laryngeal and supralaryngeal modifications, the latter characterised by longer, larger and faster lip kinematics, compatible with the enhancement of intrinsic sonority (cf. Avesani et al., 2007:984). However, in this study, comparisons investigated are only between extremely distant degrees of prominence (i.e. contrastive constituents vs. constituents out of focus). By contrast, the present experiment examines more subtle differences in focus structure (constituents out of focus, constituents in broad focus and constituents in narrow contrastive focus[14]) to understand whether a systematic increase in duration and energy characterises these subtle distinctions. The modifications will be investigated both between constituents in focus (broad vs. narrow contrastive focus) and between constituents in focus compared with constituents out of focus (constituents in narrow contrastive focus vs. constituents in post-focal position and constituents in broad focus vs. constituents in post-focal position).

[14] Broad focus is considered to be less prominent than contrastive focus since the former involves an unspecified alternative set, while the latter involves an explicit alternative set (see 2.2.1).

The variety of Italian that is the object of investigation is the one spoken in Udine, which is not well described in the literature. The reason for choosing this variety is twofold. Firstly, this variety should realise the post-focal position of sentence-length utterances with a flat and low contour, as reported for other varieties and as attested for questions (see 2.3.3). This flat and low contour presumably co-occurs also with a reduced duration and reduced energy of the stressed syllable (see e.g., Roessig & Mücke, 2019). However, the literature on Italian suggests that cues to prominence in the post-focal region might be present even without excursion in pitch. Therefore, the investigation of the degree of prominence in post-focal position might reveal that a certain degree of prominence is conveyed by the interaction of phonetic characteristics in the post-focal position. Indeed, Bocci and Avesani (2011) have found that the metrically strong position in which a word occurs is signalled through the combination of increased energy and duration. The investigation of the degree of prominence in the three focal conditions is, therefore, considered necessary, in particular to relate the acoustic characteristics of the production of the different focal conditions to perception. The degree of prominence of words occurring in different focal structures as perceived by listeners of this variety will be analysed in Chapter 6 and some of the utterances collected in the present experiment will serve as stimuli in this experiment.

Secondly, the investigation will provide insights into prominence relations in the production of information structure in this variety. To my knowledge, studies involving information structure in this variety have not been conducted (with the exception of Roseano et al., 2015, but note that this study reports on the Friulian language spoken in the region, not on the regional variety of Italian[15]). Moreover, an innovative aspect of the present study is the employment of continuous measures (scaling, synchrony and PEM) to determine the degree of prominence connected to F0 modulation, duration and energy (see 1.2.5 and 1.3.3).

The current study consists of a reading task in which the focal structure of sentences is manipulated by means of contexts in the form of question–answer pairs. Two sets of data were collected. A first set of utterances from the Udine variety revealed to be very

[15] Note that in their study, D'Agostin and Romano (2007) together with data on Friulian also presented data on the variety of Italian spoken in Udine (specifically in Codroipo, the same analysed in Chapter 5), but the authors focus on the prosodic realisation of the declarative and interrogative modality, not on information structure. Moreover, the focus of the study is mainly on Friulian and the results for the Udine variety are not clear.

interesting from the point of view of the analysis of inter- and intra-individual variability. However, the question arose whether the high amount of variability could impede the characterisation of prominence relations in the three different focal structures and make the utterances recorded not suitable for a subsequent perception experiment. Indeed, stimuli for the first dataset were originally designed not only to be used later for the prominence rating task in Chapter 6, but also to have a clear idea on what their prosodic realisation was in order to further use them as a basis to create stimuli for a perception study using EEG. Since the stimuli for this EEG experiment needed to comprise semantic incongruence with the context, participants of this set were required to utter long utterances, comprising not only of the answer to the question, but also of a repetition of part of the previous question. Even though all the sentences that the participants were required to utter were congruent with the context, the material proved to be problematic, due to the high artificial construction of the design. Speakers were in fact presented with utterances they would very rarely encounter in spontaneous speech, which caused difficulties for the participants in trying to simulate a real conversation.

Therefore, a second set of stimuli was collected, which was more controlled and more suitable for a perception experiment. This second set of stimuli was realised by manipulating the length of the utterance: while in the first set the first part of the contextualising questions was repeated in the answer, in the second set this repetition was elided. The aim was to explore whether the difference in length affects the subjects' ability to differentiate between focus structures. The question addressed was the following: Do task requirements – as reflected by the makeup of the question-answer pairs – play a role in the planning of the focus structure a speaker has to produce? The results section reports the analysis of the differences between the two datasets.

Even though this study is mainly exploratory in nature some hypotheses concerning the different focus structures and, in particular, the post-focal position can be formulated. These predictions were tested: (i) that a clear intonation distinction in terms of presence or absence of a pitch movement distinguishes broad and narrow contrastive focus from post-focal position, with the post-focal position lacking pitch movement and the broad and narrow focus conditions presenting pitch movement; (ii) that a gradual change in the acoustic continuous parameters measured is present, showing an increase from post-focal position, through broad focus, to narrow contrastive focus. However, in light of the results of Mücke and Grice (2014), the distinctions between post-focal and broad focus might not be as clear and systematic as the distinctions between broad and narrow contrastive focus. This possibility is also supported by the results reported by Bocci and Avesani (2011), who have found that in (Florentine) Italian, words in a structurally prominent position can be realised with similar values of duration and energy of a word occurring in broad focus and not in a structurally prominent position. Thus, the present experiment will help to shed further light on the acoustic characteristics of the post-focal position in Italian.

The parameters that are here measured and that undergo gradual changes are constituted by (i) scaling and synchrony, reflecting the amount and the direction of movement across and within the syllables and where an increase of values corresponds to wider movement; (ii) PEM, which is related to sonority and the enhancements of which reflects enhancement in the highlighting of prosodic information (see Beckman, Edwards & Fletcher, 1992 and see sonority expansion in 1.3.2).

5.2 First set of recordings (long dataset)

5.2.1 Materials

The reading material consisted of pairs of questions and answers. For the experiment the answers, and, in particular the word defined as the target were of interest. Contextualising questions were used to elicit three different focal conditions: target word occurring in broad focus (BF; note that it is not the focus exponent), in contrastive narrow focus (NF) or in post-focal position (PF, i.e., as part of the background). An example of the stimuli in the three renditions is reported in (19)-(21). The complete list of stimuli is available in Appendix A1. The target is written in bold and the narrow focus in capital letters. Whilst the major interest in the prosodic realisation was primarily directed to the target, the prosodic analysis considered also two other critical words to better qualify the prosodic realisation of the whole utterance: the main verb and the noun in the prepositional phrase (PP). Therefore, the words considered were the three content words (henceforth, critical words) that are underlined in the examples. In all the stimuli the target word was a noun (direct object of the verb) and was followed by a PP. The target words were always trisyllabic with the stress on the second syllable, containing either the vowel /i/ or /a/. The stimuli used in the experiment forced the speakers to repeat a part of the question followed by the second part of the sentence consisting of the verb *bisogna* (one needs to), followed by the verb that relates to the question, a noun, which is considered as the target word and the final prepositional phrase. All the PPs except one were comprised of a preposition followed by a noun (as in (19)-(21); in the exception instead of the noun there was an adverb). The noun in the PPs was considered a critical word (in the only exception the adverb was considered as the critical word).

(19) TARGET in BF

Q: Cosa bisogna fare quando si va in gita?
PRO IMPRS INF CNJ IMPRS PREP N
What need do when go on trip

What does one need to do when going on a trip?

A: Quando si va in gita,
CNJ IMPRS PREP N
When go on trip

When going on a trip,

[bisogna <u>preparare$_{verb}$</u> un **<u>panino$_{TARGET}$</u>** per la <u>merenda$_{noun\ in\ PP}$</u>]$_{FOCUS}$
IMPRS INF DET N PREP DET N
need prepare a sandwich for the snack

one needs to <u>prepare</u> a <u>sandwich</u> as a <u>snack.</u>

(20) TARGET in NF

Q: Quando si va in gita, bisogna preparare un panino o un caffè per la merenda?
CNJ IMPRS PREP N IMPRS. INF DET N CNJ DET N PREP DET N
When. go on trip need prepare a sandwich or a coffee for the snack

When going on a trip, does one need to prepare a sandwich or a coffee as a snack?

A: Quando si va in gita,

When going on a trip,

bisogna <u>preparare$_{verb}$</u> [UN **<u>PANINO$_{TARGET}$</u>**]$_{FOCUS}$ per la <u>merenda$_{noun\ in\ PP}$</u>.

one needs to <u>prepare</u> a <u>sandwich</u> as a <u>snack.</u>

(21) TARGET in PF

Q: Quando si va in gita, bisogna preparare o comprare un panino per la merenda?
CNJ IMPRS PREP N IMPRS INF CNJ INF DET N PREP DET N
When go on trip need prepare or buy a sandwich for the snack

When going on a trip, does one need to prepare or buy a sandwich as a snack?

A: Quando si va in gita,

When going on a trip,

bisogna [PREPARARE$_{\text{verb}}$]$_{\text{FOCUS}}$ un **panino**$_{\textbf{TARGET}}$ per la merenda$_{\text{noun in PP}}$

one needs to prepare a sandwich as a snack.

The word identified as target was considered the most interesting because of its position before a prepositional phrase. Indeed, considering the mapping between syntactic structure and prosodic structure, an intermediate phrase boundary would occur after the target (see Nespor & Vogel, 1986; Selkirk, 1995, but also Bocci & Avesani, 2011). The target was, therefore, supposed to be always realised in a different intermediate phrase than the prepositional phrase and be its metrical head. For its metrically prominent position the target was considered to be of most interest for the analysis of its prominence degree. In fact, the target in post-focal position should present some cues to prominence even if realised with flat and low contour. Its prominence degree should however be less than the one of the target occurring in broad focus, given that it also occurs in prominent position and should be featured by a pitch accent (see for example Gili-Fivela et al., 2015).

5.2.2 Participants

For the experiment 14 native speakers of the Italian variety spoken in Udine (specifically all born and grown up in the surroundings of the district of Codroipo) were tested (10 female, 4 male). Participants were between 19 to 26 years old (mean age 21.86, *SD* 2.38). All speakers had been continuously exposed to the respective variety of Italian, used it for everyday conversation and had a similar educational level (either university students or people who recently graduated). None of them self-reported any speech or auditory impairment. In addition, none of the participants had a background in phonetics or prosody. All participants gave written informed consent.

5.2.3 Procedure

The speakers were digitally recorded at a sampling rate of 44 kHz with a 16-bit resolution. The material was presented on slides in a PowerPoint presentation in a pseudo-randomised order. Each slide contained the question eliciting the focal condition and the answer to the question. Subjects were paired and instructed to first read both the question and the answer silently. When they felt they had understood the meaning conveyed by the question-answer pair, one of them had to utter the question aloud, while the other had to answer. The participants were asked to correct themselves if they thought that their production was incorrect or unnatural. Participants were asked to come with a friend, so each participant knew the fellow participant and felt comfortable with him/her. The reason for this choice was to create a comfortable situation in which the informants could express themselves as naturally as possible. The motivation to test pairs of participants was to best elicit their variety and to best elicit focal structures, both of which emerge more easily in a situation of dialogue rather than in the case of a single participant uttering a sentence, because read dialogues are prosodically closer to spontaneous speech (Niebuhr, Bergherr, Huth, Lill, & Neuschulz, 2010; see also Niebuhr, & Michaud, 2015). Each recording session was preceded by a training session of five question–answer pairs. This block did not enter the analysis. Question-answer pairs were randomized to avoid repetitions in sequence: the different conditions of the same item were separated by at least 10 slides. The number of question-answer pairs recorded for each participant was 60 (20 items × 3 focus structure).

Before the main experiment participants took part in a Map Task (Anderson et al., 1991) eliciting semi-spontaneous questions and answers. The purpose of the Map Task was to let the participants get used to the modality of the experiment, that is to become accustomed to having their voices recorded. The Map Task used was part of the ones created for the CLIPS corpus (CLIPS, 2004). By contrast, for the main experiment, controlled stimuli were used (described in 5.2.1). Controlled stimuli were used for different reasons. Firstly, the use of controlled stimuli in a controlled setting allowed for a strict monitoring of the contribution of stress and accent and how they interact with each other. Secondly, the fact that critical words occur in the same position within an utterance allowed for a more homogeneous comparison among the conditions. Thirdly, this production experiment was designed to be the basis for the rating study (Chapter 6), in which controlled stimuli were necessary for the purpose of establishing the perception of prominence in the post-focal domain.

5.2.4 Analysis

Utterances were analysed by two trained transcribers using Praat (Boersma & Weenink, 2020). The transcribers worked together on a consensus annotation of the data in terms of ToBI labels, using their auditory impressions and the F0 contour in Praat. The cases of disagreement were discussed until a consensus on the transcription was found. In all cases, the solution was found rapidly. Since a phonological description entails a systematic examination of several phenomena (as for example tonal alignment, see 1.2.4.1), the analysis presented here needs to be considered provisional and functional to identify the main differences in the production of the different focal structures. Nonetheless, the present analysis was conducted mainly referring to the existing description of Friulian, a language spoken in the same region, made by Roseano et al. (2015). When the description proposed in the previous study appeared consistent with the present analysis, the description has been mantained.

The quantitative analysis of the contours was conducted through the parameters of synchrony and scaling (see 1.2.5). To estimate the strength of the perceived prominence for each syllable Periodic Energy Mass (PEM) was calculated (see 1.3.3). These values were obtained using Praat's (Boersma & Weenink, 2020) and R's (R Core Team, 2019) scripts made available by Albert and Cangemi (2020). More in detail, values of synchrony corresponded to the distance in time between the Center of Gravity (CoG) and Center of Mass (CoM) within each periodic energy fluctuation (i.e., syllables. CoG minus CoM). Scaling was measured as the difference between the values of F0 at the CoM of successive syllables. The last measure, PEM, referred to the area under the periodic energy curve for each fluctuation of the periodic energy (i.e. to each syllable). Throughout the present thesis PEM will be intended as relative PEM and will refer to a value relative to other values within the same utterance. This relative value is calculated first by measuring the periodic energy mass for the whole analysed utterance and dividing it by the number of periodic energy fluctuations, thus obtaining an average value (y); second, by dividing the periodic energy mass of the single syllable (x) by the previously calculated average value (PEM = x/y). This measure is informative, since values around 1 are of average strength (the average value will always be around 1, given that it is a division between the overall strength in the utterance and the number of syllables of the utterance). By contrast, the more PEM values of a syllable are distant from 1, the more this syllable is either reduced (under 1) or enhanced (above 1).

To analyse the difference in the acoustic features of the stimuli for each word position and for each condition, linear mixed effects models of the relationship between focal condition, critical word position and prominence were run. Firstly, F0 dynamic was considered a cue to prominence: mixed effect models were performed, with POSITION of the word (verb, target and noun in PP) and CONDITION (broad focus, narrow focus, post-focal position) as independent variables, and values of synchrony and scaling for the stressed syllables as dependent variables, considered separately. Random intercepts for ITEM and random intercepts and slopes for SUBJECT were considered as random effects. Secondly, relative PEM for each stressed syllable of each word position was also considered a cue to prominence, and regarded as the dependent variable of mixed effects analyses. For these analyses, the factors CONDITION and POSITION were considered as independent variables. Random intercepts for ITEM and random intercepts and random slopes for SUBJECT were considered as random effects.

For the analysis, R (R Core Team, 2019) and *lme4* (Bates, Maechler & Bolker, 2012) were used. The *afex* package (Singmann, Bolker, Westfall & Aust, 2018) was used to obtain p-values.

5.2.5 Results of the ToBI analysis

Results of pitch accents are displayed in Figure 25. This figure shows the distribution of the ToBI labels for each focus condition and each position of the word in the utterance. Results are pooled over speakers. The figure shows that there is some variability in the use of pitch accent types across the focal conditions. Nonetheless, it also shows preferences for certain pitch accent types per condition and position of the word. The focus exponents of the utterances containing contrastive focus, i.e., the verb in PF and the target in NF are realised with a high degree of variability. The majority of them were realised with a rising accent (L+H*, 45.79% of the cases for verb in PF; 43.32% of the cases for target in NF) and in some rare cases with a (rising-)falling accent (H*+L, 6.23% of the cases for verb in PF; 1.44% of the cases for target in NF). Both these accent types are attested for different varieties of Italian for marking narrow focus (Gili-Fivela et al., 2015). In particular H*+L accents are attested in the dialect spoken in the region of Udine for nuclear contours of epistemically biased statements (Roseano et al., 2015). In addition, some of the cases of words in contrastive narrow focus were realised with a H* accent (21.98% of the cases for verb in PF; 15.16% of the cases for target in NF), though not attested for contrastive-corrective narrow focus (Gili-Fivela et al., 2015). In a few cases NF was realised without a pitch movement (1.47% for the verb in PF; 2.88% for the target in NF).

The production of the target words occurring in BF and PF were also variable. However, preferred contours were observed for these cases as well. Targets in BF were mostly realised with a H+L* accent (69.93%), mostly attested for different varieties of Italian for words occurring in this type of focal structure (Gili-Fivela et al., 2015; Roseano et al., 2015). Other cases were realised with shallow rises (H*, 23.19%), while very few cases were realised with prominent rising accents (L+H*, the 3.26%) and without pitch movement (3.62%). The target in PF showed the same occurrence of contours found in BF, but with different proportions, since the majority of the realisations did not show a pitch movement (50.92%).

The noun in PP, was mostly realised in all the conditions with a falling contour (H+L*, 89.89% of the case for BF; 70.25% of the cases for NF; 58.3% of the cases for PF). In addition, in PF, a high percentage was also realised with a low and flat contour (40.59% of the cases), which is the shape that is mostly attested for elements in post-focal position of declaratives (see 2.3.3). The other accents in the noun in PP were very rarely realised (a total of 8.3% for H* accents and a total of 4.69% for L+H* accents).

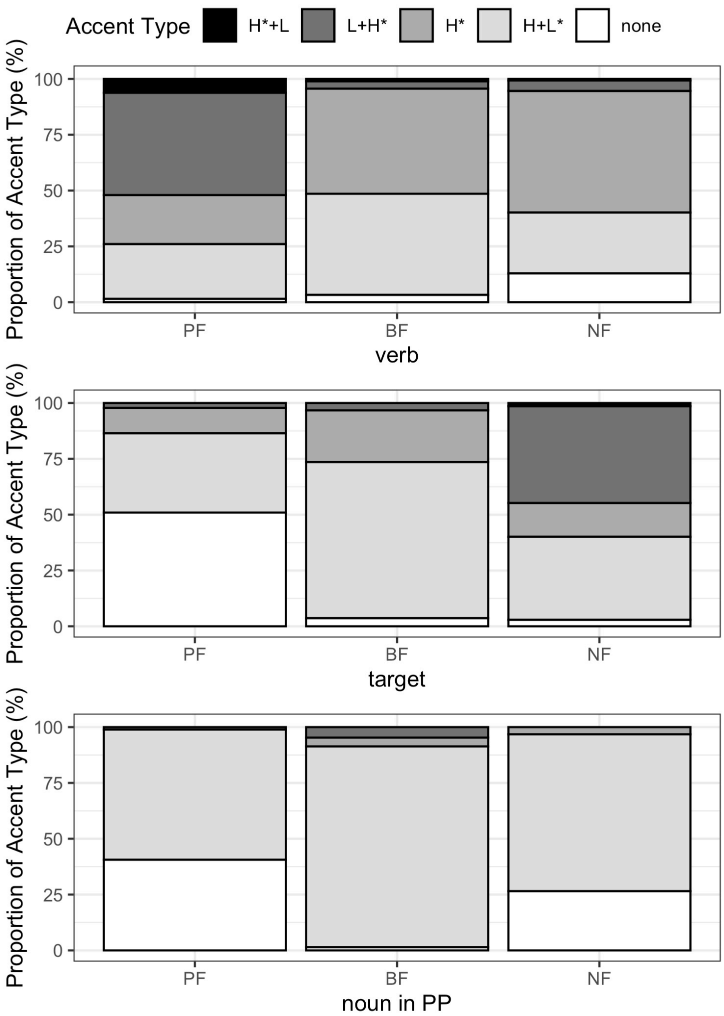


Figure 25. Distribution of pitch accent types across conditions and word position. Data are pooled across speakers.

Looking at the productions of individual speakers (Figure 26), a high inter-individual variability can be noticed. Not all the speakers used all accent types to distinguish the conditions. For example, four speakers (F01, F02, F07, M04) hardly made use of the wide pitch range rising (or rising-falling) accent (L+H* or H*+L) to mark contrastive narrow focus (F01: 25% of the cases on the narrowly focused verb in PF and 20% of the cases in the narrowly focused target in NF; F02: 5.56% of the cases on the narrowly focused verb in PF and 5.26 % of the cases in the narrowly focused target in NF; F07: 21.05% of the cases on the narrowly focused verb in PF and 0% of the cases in the narrowly focused target in NF; M04: 0% of the cases on the narrowly focused verb in PF and 5% of the cases in the narrowly focused target in NF). In particular, speaker M02 used it to some extend in the target word in PF (31.58% of the cases), which occurs in post-focal position, and never to mark the target word in NF. For these speakers, in the majority of the cases the categorical analysis did not show a distinction in the focal structure. On the contrary, speaker F03, F04, F05 F06, F08, F09, F10 and M03, used the rising accent (L+H*) or the (rising-)falling accent (H*+L) to mark the focal exponent of the utterance. However, with the exception of speaker F04, F05 and F06, the variability was very high within participants. The distinction between BF and the other two conditions was also in some of the cases coded by a high intermediate boundary tone (H-) after the target (in the 59.73% of the cases).

In the next section the continuous parameters relating to F0 and periodic energy are investigated, in order to see whether the categorical analysis missed some regularities in the data and can reveal that speakers make distinctions between categories using modulations of continuous parameters and their interactions, which a categorical analysis cannot grasp.

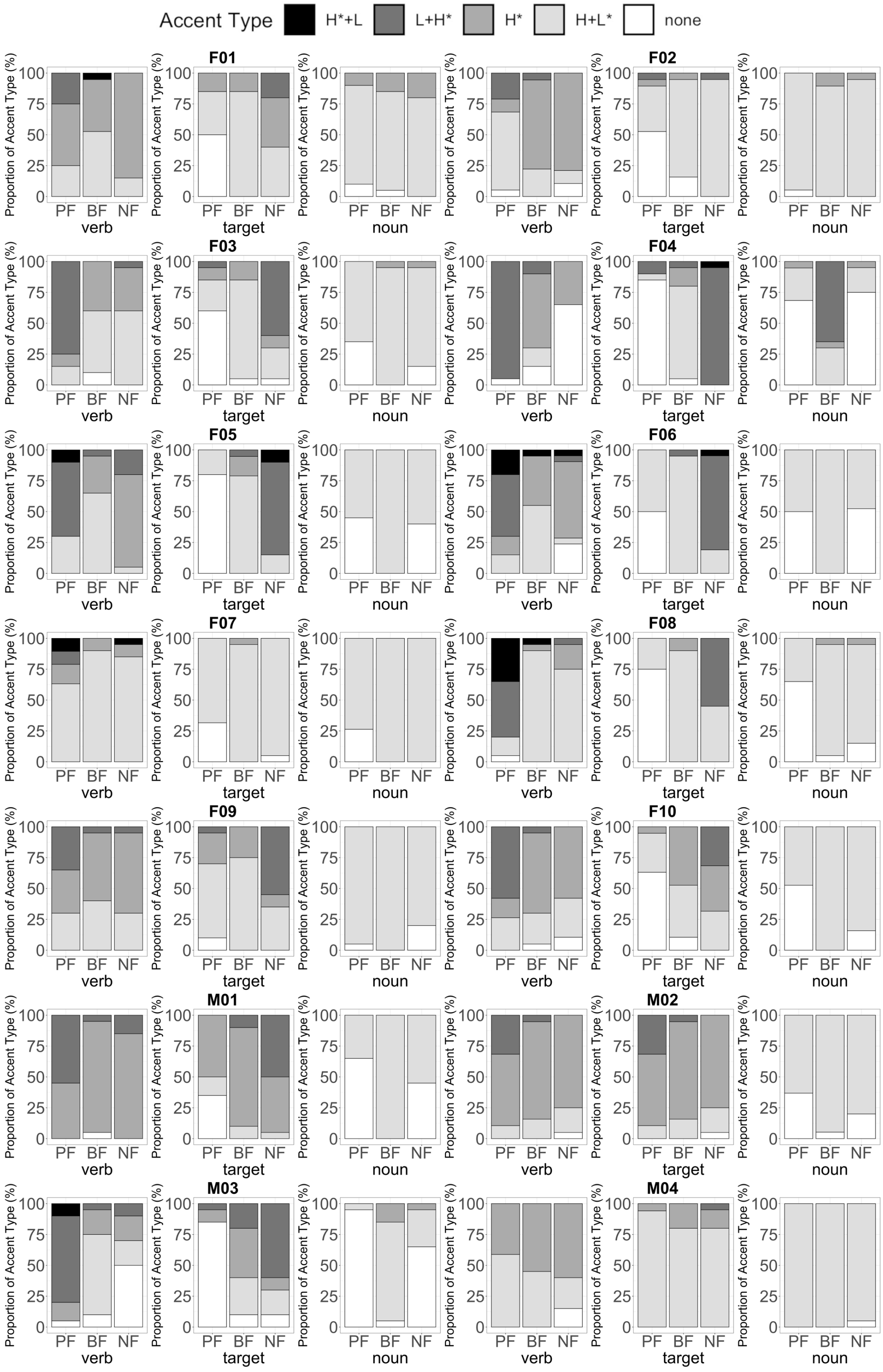


Figure 26. Distribution of pitch accent types across conditions and word position for each speaker separately.

5.2.6 Global contours

The superimposed contours in Figure 27 (female speakers) and Figure 28 (male speakers) show intra- and inter- individual variation in the three focus conditions. For speakers F02, F03, F04, F05 and F06 a clear tendency towards a distinction of the conditions is registered. The differences can be at best noticed in the target, which has the tendency to be produced by these speakers with a different alignment of the movement in BF compared to NF and with a flat and low contour in PF. By contrast, the verb in PF shows a high amount of movement, which is not shown in NF and is shown only to a lesser extend in BF. For the other speakers the tendency in differentiating the conditions is there, but cannot be easily identified by looking at the contours. In addition, for all the speakers a high amount of intra-individual variability can be noticed.

Table 6 reports F0 means and standard deviation (in parentheses) for each condition and for each position of the word at the target height, corresponding to the starred tone of the ToBI transcription. For post-focal position the reference for measurements was placed at the half of the total duration of the stressed syllable.

condition	**F0 verb [Hz]**	**F0 target [Hz]**	**F0 noun in PP [Hz]**
PF	**Male** 112.73 (23.6) **Female** 204.41 (31.02)	**Male** 94.76 (14.14) **Female** 178.04 (18.87)	**Male** 88.44 (14.45) **Female** 178.16 (29.24)
BF	**Male** 102.79 (15.71) **Female** 192.65 (21.67)	**Male** 103.13 (23.19) **Female** 191.52 (21.86)	**Male** 84.92 (18.47) **Female** 177.27 (31.37)
NF	**Male** 103.55 (25.23) **Female** 195.0 (18.94)	**Male** 101.37 (20.74) **Female** 202.97 (33.96)	**Male** 83.84 (15.49) **Female** 172.17 (26.36)

Table 6. F0 mean and standard deviation (in parentheses) at the reference point for each word position and prosodic condition of the stimuli. Values for male and female are shown separately.

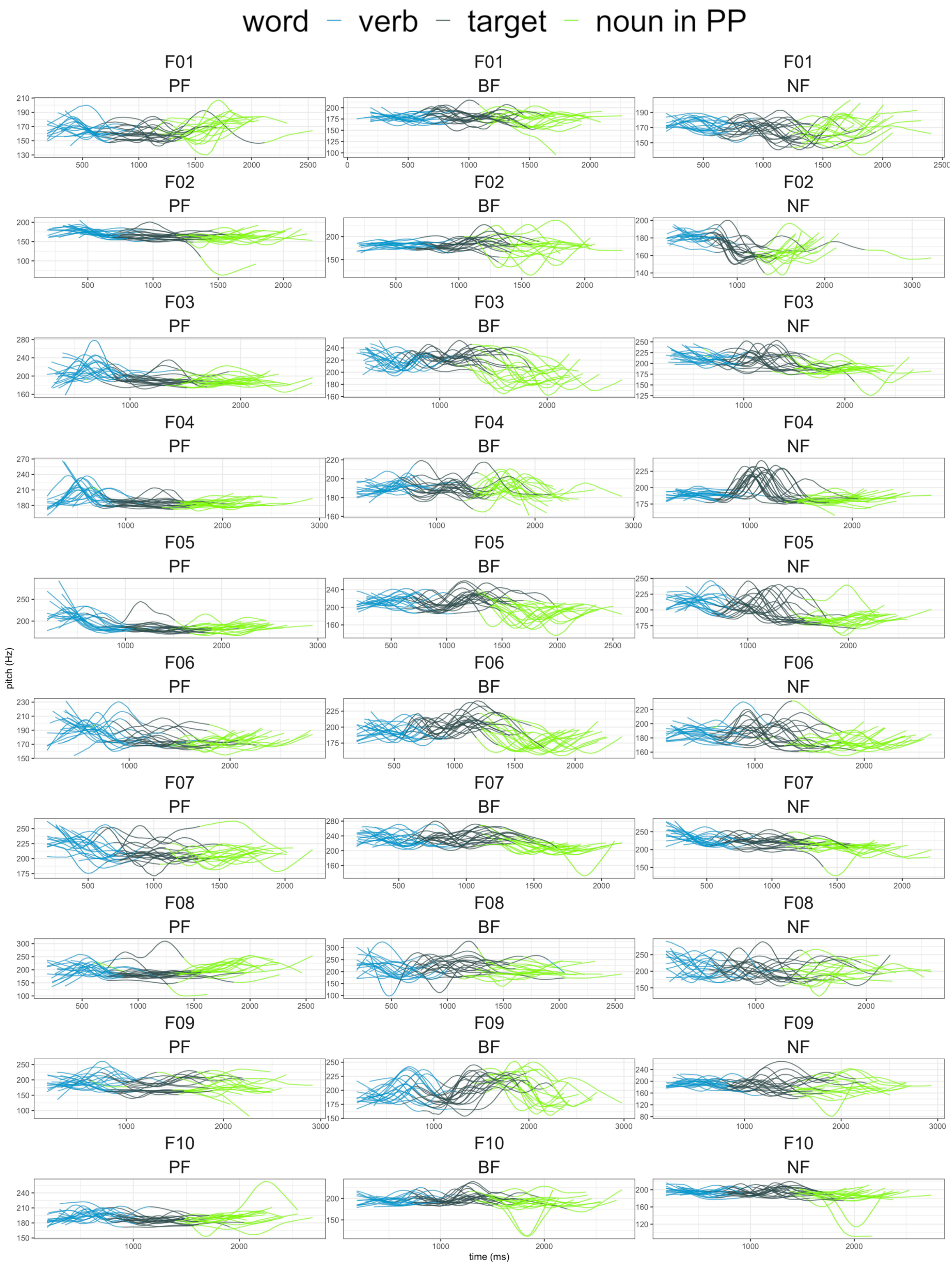


Figure 27. Superimposed contours for each intended focal condition (from left to right: post-focal position, broad focus and narrow focus) for each female speaker separately. Different colours of the contours indicate the critical words (verb, target and noun in PP).

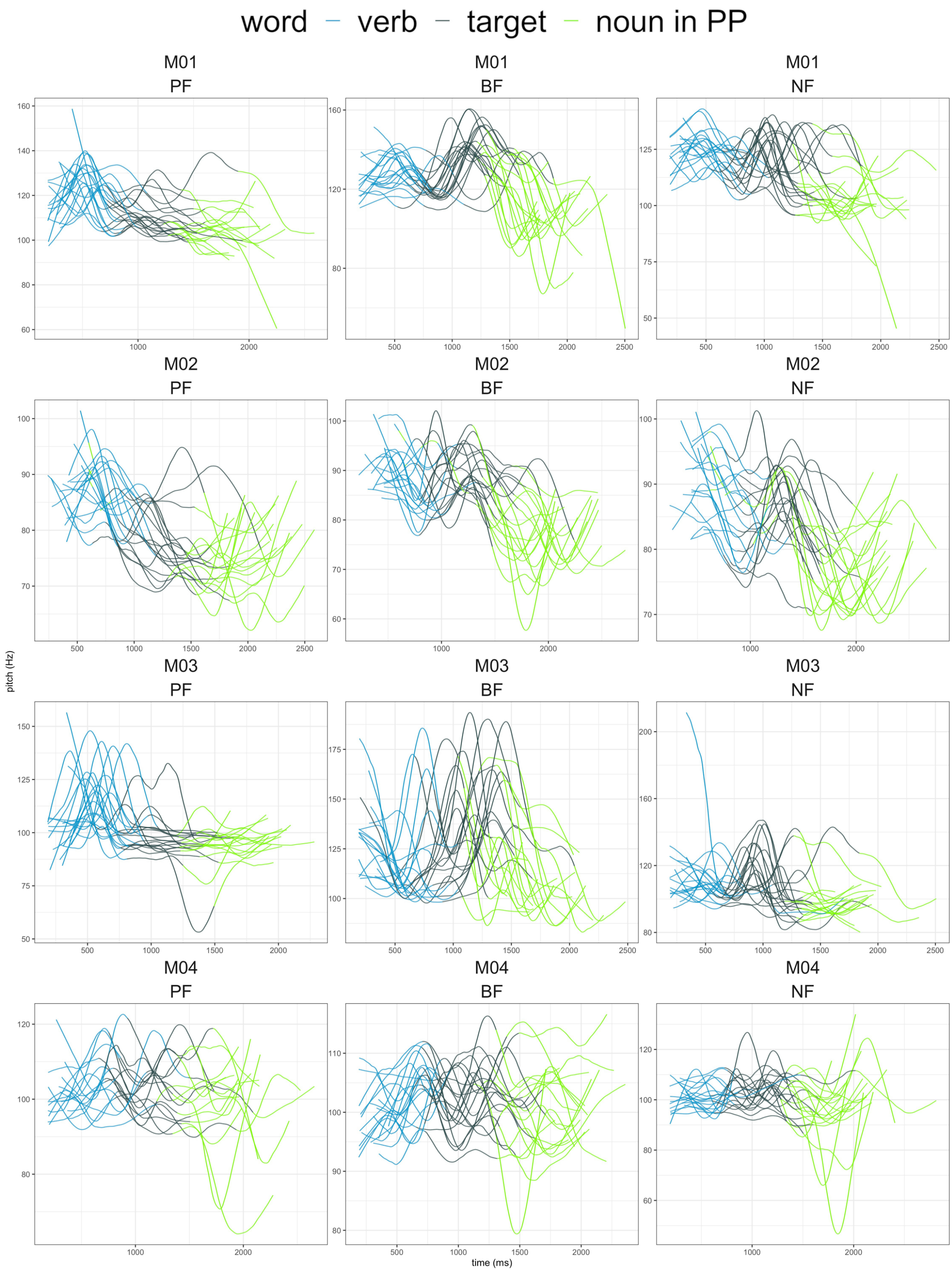


Figure 28. Superimposed contours for each intended focal condition (from left to right: post-focal position, broad focus and narrow focus) for each male speaker separately. Different colours of the contours indicate the critical words (verb, target and noun in PP).

In the following section, the contours will be quantitatively described and measured through the parameters of synchrony and scaling. Periodic Energy Mass will also be presented. These three measures will help describe the prosodic realisation of each word position in different focal conditions, as well as their degree of prominence.

5.2.7 Synchrony, scaling and Periodic Energy Mass

To descriptively explore the continuous parameters that are considered in this section, violin plots will be employed. Violin plots are based on the kernel density estimations and, therefore, show the probability density of the data at different values: the more bulging parts of the graphs correspond to higher likelihood of a value.

Figure 29 provides a closer look at continuous measures (F0 and periodic energy and the related measures of synchrony and scaling) pooled per speakers, which shows the measures of synchrony and scaling for the whole contours and confirms the tendency towards a distinction of the conditions. The figure only refers to the case in which verb and target have the stress on the second syllables and noun in PP on the first, and is only exemplary for the values found in these cases. The other possible combinations can be retrieved at https://osf.io/5m8hw/. Note that the first syllable of scaling is not present in the graphs. This is because scaling only refers to previous values. Therefore, it is not meaningful to report the scaling for a first syllable that does not have preceding values to which it could be compared.

The values reported in parentheses in the following description of the data and in Table 7 refer to the stressed syllables of all the utterances (also of verb and noun in PP that do not present the stress on the second syllable; the target always has the stress on the penultimate syllable). These measures interpreted together show a tendency towards a distinction in the modulation of F0 in different conditions.

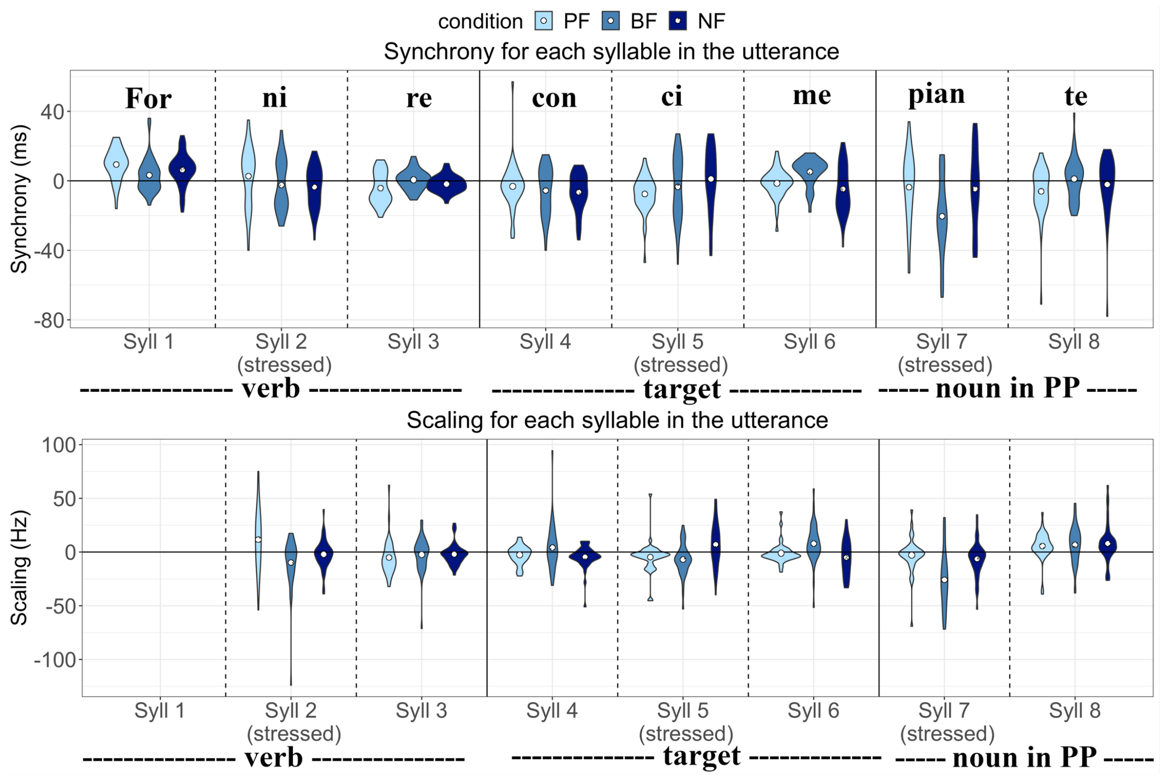


Figure 29. Synchrony and scaling of each syllable of the critical words. The stress is on the second syllable of the verb and on first syllable of the noun. White dots on the violin plots indicate the mean value. Each box in the graphs corresponds to a syllable in the utterance and presents values relative to this syllable for each condition (from left, PF, BF and NF). Note that the values of scaling for the first syllable are not presented because they are not meaningful. Solid vertical lines correspond to the boundaries of each critical word (from the left, verb, target and noun in PP), while dotted vertical lines indicate the boundaries of the syllables. Labels on the top provide an example of how each syllable is displayed and contains the three syllables of the critical words of the example utterance *Fornire il concime alle piante* (Provide fertilizer to plants).

Regarding synchrony there is not much difference among conditions and word position. By contrast, differences can be seen in the scaling values: stressed syllables of the verb show the highest values of scaling for PF, stressed syllables of the target show the highest values for NF, while the distribution of values for the target in PF is more condensed around zero. By contrast, BF shows a wider distribution of values. In the noun in PP values for PF and NF are similar, even though NF shows more negative values, while BF shows a trend towards negative values.

What can be noticed is that the verb in PF presents a rising pitch, indicated by positive values of scaling and mostly positive values of synchrony on the stressed syllable. The two critical words that follow the verb are realised with a flat and low contour, indicated by synchrony and scaling values near zero on all the syllables (Syll 4: synchrony mean -1.98, *SD* 9.74; scaling mean -1.09 *SD* 9.41; Syll 5 synchrony mean -4.72, *SD* 9.96;

scaling mean -2.98, *SD* 9.6; Syll 6 mean 0, *SD* 7.26). BF presents a bimodal distribution of synchrony values on the stressed syllable of the verb, which does to some extent overlap to the one of PF, but shows a higher proportion of negative values (mean -1.78, *SD* 13.75). While values of synchrony for the verb in BF do not differ from the ones in PF, scaling values show a shift towards more negative values in BF compared to PF. Negative values of the stressed syllable of the verb in BF indicate the presence of a majority of falling pitch in this condition. The following syllable of the verb (Syll 3), synchrony and scaling for PF present a tendency towards more negative values than BF (PF synchrony mean -4.2, *SD* 8.57; scaling -6.8, *SD* 13.99; BF synchrony mean 0.3, *SD* 6.58; scaling 0.71, *SD* 14.31), indicating for PF a falling movement from the previous syllables (Syll 2). In NF both values of synchrony and scaling for all syllables of the verb present values of synchrony and scaling around zero, which together indicate compressed movement on the verb.

Values of synchrony and scaling for all the syllables of the target in PF, indicate that the contour is flat and low. In fact, although values of synchrony show some variability within the syllables, they concentrate around zero (Syll 4 mean = -1.98, *SD* = 9.74; Syll 5 mean = -4.72 *SD* = 9.96; Syll 6 mean = 0, *SD* = 7.26). In addition, the distribution of scaling values around zero (Syll 4 mean = -1.09, *SD* = 9.41; Syll 5 mean = -2.98 *SD* = 9.6; Syll 6 mean = 3.67, *SD* = 11.61) indicates that no substantial pitch change occurs on the target word. The same flat and low contour is registered also for the noun in PP occurring in PF (Syll 7). By contrast, the target in BF is characterised by a bimodal distribution of synchrony values on the stressed syllable, indicating the presence of either rising or falling pitch. On the following syllable (Syll 6), the pitch is slightly rising (synchrony: mean = 6.47, *SD* = 6.14; scaling: mean = 10.37, *SD* = 11.78), indicating the phrase accent reported in 5.2.5. The stressed syllable of the noun in PP is characterised by a falling contour, indicated by negative values of both scaling and synchrony in the stressed syllable (synchrony: mean = -11.43, *SD* 20.82; scaling mean = -17.17, *SD* 21.79). The target word in NF shows a tendency towards positive values of scaling and synchrony for the stressed syllable (Syll 4), indicating the presence of a rising pitch. The stressed syllable of the noun in PP shows a slightly falling pitch on the stressed syllable and a tendency towards no movement on the last syllable (Syll 7, synchrony: mean = 3.15, *SD* = 17.98, scaling: mean = -6.82, SD = 15.66; Syll 8, synchrony: mean = 0.86, *SD* = 11.87, scaling: mean = 6.76, *SD* = 16.87).

condition	verb	target	noun in PP
PF	Synchrony: -0.69 (17.21) Scaling: 6.37 (23.28)	Synchrony: -4.72 (9.96) Scaling: -2.98 (9.6)	Synchrony: 1.8 (18.13) Scaling: -1.7 (15.22)
BF	Synchrony: -1.78 (13.75) Scaling: -6.41 (16.92)	Synchrony: 0.95 (15.55) Scaling: -2.55 (14.81)	Synchrony: -11.43 (20.82) Scaling: -17.17 (21.79)
NF	Synchrony: -2.71 (10.89) Scaling: -3.49 (13.68)	Synchrony: 1.32 (15.65) Scaling: 8.61 (20.20)	Synchrony: -3.15 (17.98) Scaling: -6.82 (15.66)

Table 7. Mean and standers deviation (in parentheses) of synchrony and scaling values for the stressed syllables of the critical words in all conditions for all the stimuli.

Statistical models carried out for the values of synchrony and scaling of the stressed syllables confirmed the tendencies reported in the description of the graphs. Results of the models run for each position of the word on the differences between conditions are listed in Table 8, where the intercept of the model is BF.

	verb	target	noun in PP
PF	Synch: β = 0.6 ± 2.19, p = 0.79 **Scaling: β = 11.93 ± 4.91, p = 0.03**	**Synch: β = -5.37 ± 2.23, p = 0.03** Scaling: β = -0.21 ± 1.43, p = 0.89	**Synch: β = 12.69 ± 1.88, p < 0.0001** **Scaling: β = 15.18 ± 3.53, p = 0.001**
NF	Synch: β = -0.49 ± 1.43, p = 0. 74 Scaling: β = 2.45 ± 1.71, p = 1.75	Synch: β = 0.37 ± 2.5, p = 0.89 **Scaling: β = 11.01 ± 3.18, p = 0.004**	**Synch: β = 7.93 ± 1.83, p < 0.0001** **Scaling: β = 9.76 ± 3.07, p = 0.01**

Table 8. Results of the mixed model analyses of the differences between the stressed syllables' values in each condition in each word position. Values are relative to the intercept, corresponding to the relative critical word (verb, target and noun in PP) in broad focus. Significant results are displayed in bold.

Table 9, Table 10 and Table 11 report the results for the difference of the stressed syllable of each critical word within each condition. In post-focal position (Table 9), the verb has higher movement in comparison to the target both within the stressed syllables and across the stressed syllable and the preceding one (increase in values of synchrony and scaling). The noun in PP shows a reduced movement within the stressed syllable compared to the stressed syllable of the target, but not different movement across the stressed syllable and the preceding one.

In broad focus (Table 10), the verb is distinguished from the target through an increase of the range in the falling movement across the stressed syllable and the preceding one (lowering of scaling values). The noun in PP is different from the target in terms of an increase in the range of the falling both within the stressed syllable (lowering in synchrony) and across syllables (lowering in scaling).

In narrow focus (Table 11), the stressed syllable of the verb shows reduced values of scaling from the preceding syllable compared to the target, confirming that the movement, if present, is rather reduced in the verb. A higher reduction of movement registered both by values of synchrony and scaling occurs on the noun in PP.

	PF
verb	**Synch: β = 5.63 ± 1.3, p < 0.0001**; **Scaling: β = 9.22, ± 1.39, p < 0.0001**
noun in PP	**Synch: β = 6.23 ± 1.3, p = 0.0001**; Scaling: β = 1.44 ± 1.32, p = 0.27

Table 9. Results of the mixed model analyses of the differences between each word position in post-focal position. Values are relative to the intercept, corresponding to the target in the relative condition.

	BF
verb	Synch: β = - 0.34 ± 1.42, p = 0.81; Scaling: β = - 3.37 ± 1.51, p = 0.03
noun in PP	**Synch: β = -11.93 ± 1.41, p < 0.0001**; **Scaling: β = -14.18 ± 1.43, p < 0.0001**

Table 10. Results of the mixed model analyses of the differences between each word position in broad focus. Values are relative to the intercept, corresponding to the target in the relative condition

	NF
verb	Synch: β = - 1.21 ± 1.27, p = 0.34; **Scaling: β = - 12.02 ± 1.37, p < 0.0001**
noun in PP	**Synch: β = -4.36 ± 1.27, p = 0.001; Scaling: β = -15.44 ± 1.3, p < 0.0001**

Table 11. Results of the mixed model analyses of the differences between each word position in narrow focus. Values are relative to the intercept, corresponding to the target in the relative condition.

Looking at the individual differences, the distinction of NF from the other two conditions becomes more evident. Figure 30 and Figure 31 show values for the target word (values of synchrony and scaling respectively). Values of synchrony and scaling show a difference between NF and BF that can be expressed in one of the three syllables of the target. Speakers differed in conveying the different focal conditions using more within syllable modulations of F0 or across syllables F0 modulations. For example, speaker M02 did not distinguish NF from the other two conditions with modulations across the syllables (values of scaling around zero for all the conditions), but did so within the stressed syllable.

By contrast, the distributions of synchrony values in PF and BF are more difficult to bring apart. However, a general tendency of BF towards lower values than PF in the stressed syllable was present in speaker F01, F03, F06, F08 and M02. Other speakers, M01, M03, M04, F09, F10 show more positive values for BF condition on the stressed syllable of the target. F05 shows more negative values on the first syllable, while F07 shows to have both lower and higher values on the stressed syllable for BF in comparison to PF, indicating a higher degree of movement for BF.

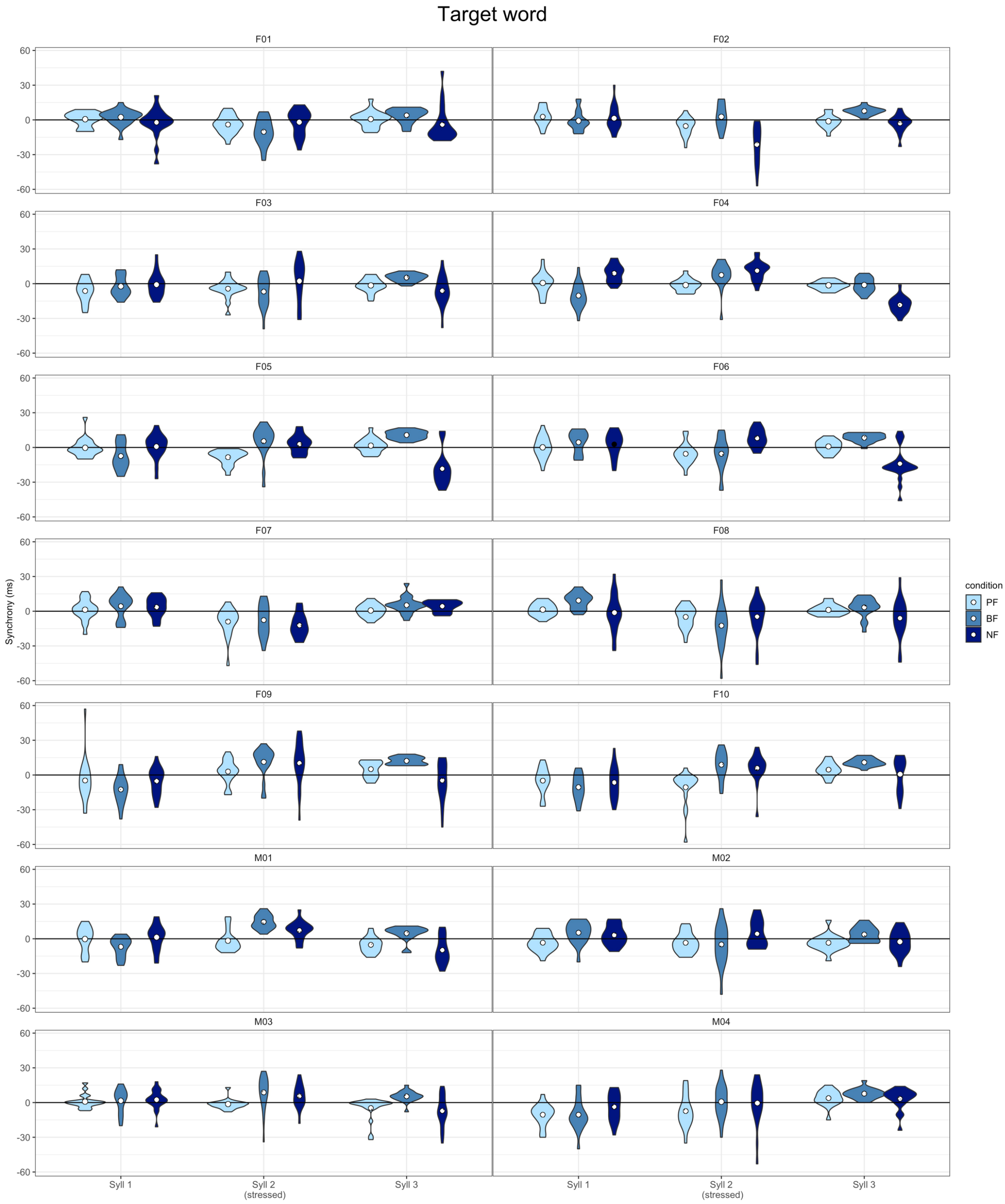


Figure 30. Synchrony of each syllable of the target word for each condition for each speaker (one speaker per panel). White dots indicate mean values.

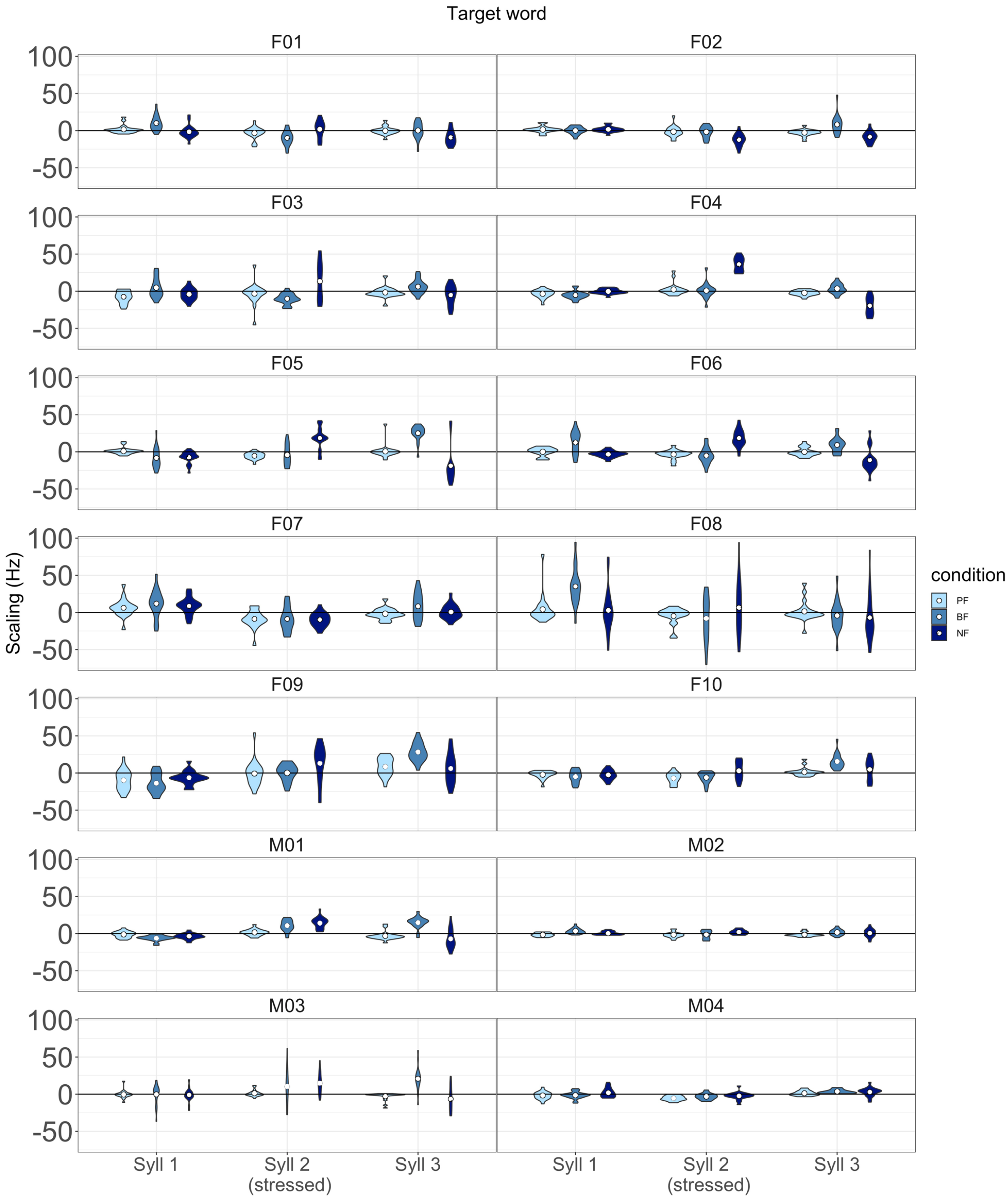


Figure 31. Scaling of each syllable of the target word for each condition for each speaker (one speaker per panel). White dots indicate mean values.

So far, the results of the data show that prominence as coded by the modulation of the F0 indicate that:

- the verb in PF is made prominent by a greater excursion in pitch in comparison to the other two conditions, which do not differ in this position

- the target in NF is made prominent using the same strategy adopted for the verb in PF. While the target in NF and BF presents a pitch movement, this movement is not present in PF.
- Modulations in the F0 mark the noun in PP as prominent when occurring in BF and not when occurring in PF and NF (i.e., in post-focal position).

Another dimension in which the distinction in prominence among conditions can be coded is Periodic Energy Mass (PEM), shown in Figure 32. Results of the statistical model are reported in Table 12. As expected, the data reveals that values of PEM for the verb are higher in PF compared to the other two conditions. For the target, stressed syllables in NF shows a tendency towards higher values than in BF. Interestingly, the stressed syllable of the target does not show different PEM values between BF and PF. Moreover, the noun in PP shows unexpectedly high values of PEM. Indeed, results of the models also registered higher values of PEM for the noun in PP than for the target (see Table 13). For the noun in PP a difference emerged among conditions, with stressed syllables in PF and NF unexpectedly showing higher values than in BF. This can be due to the fact that the noun in PP is realised as a separate intermediate phrase. Moreover, this indicates that prominence cues are present that are not connected to pitch movement. The difference in the PEM values in the noun in PP is actually not noticeable from Figure 32. This is due to the high degree of variability between participants, which is shown in Figure 34.

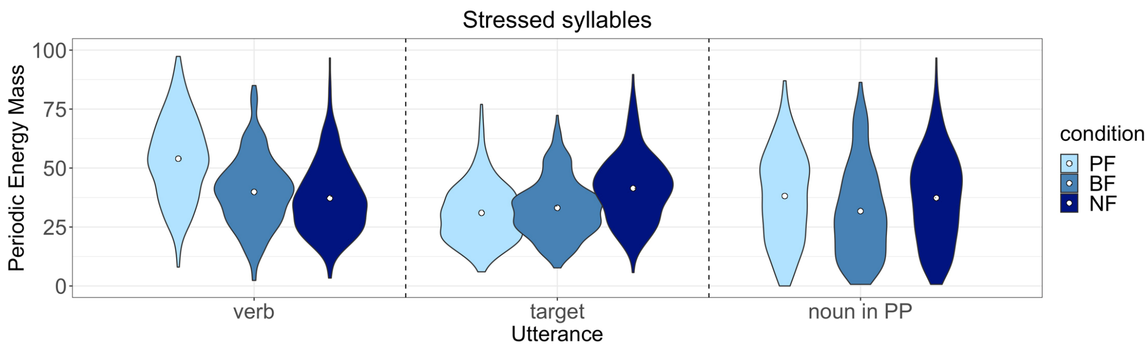


Figure 32. PEM for the stressed syllables of each critical word (from left to right, verb, target and noun in PP) and each condition (from left to right, post-focal position, broad focus and narrow focus). White dots indicate mean values.

	verb	target	noun in PP
PF	**β = 0.45 ± 0.07, p < 0.0001**	β = -0.06 ± 0.03, p = 0.06	**β = 0.19 ± 0.07, p = 0.02**
NF	β = -0.06 ± 0.03, p = 0.05	**β = 0.25 ± 0.04, p < 0.0001**	**β = 0.17 ± 0.06, p = 0.01**

Table 12. Results of the mixed model analyses of the differences between the stressed syllables' values in each condition in each word position. Values are relative to the intercept, corresponding to the relative critical word (verb, target and noun in PP) in broad focus. Significant results are displayed in bold.

	PF	BF	NF
verb	**β = 0.72 ± 0.04, p < 0.0001**	**β = 0.20 ± 0.04, p < 0.0001**	**β = -0.12 ± 0.04, p = 0.003**
Noun in PP	**β = 0.21 ± 0.04, p < 0.0001**	β = -0.04 ± 0.04, p = 0.24	**β = -0.12 ± 0.04, p = 0.002**

Table 13. Results of the mixed model analyses of the differences between each word position within each condition, from the left: broad focus, narrow focus and post-focal. Values are relative to the intercept, corresponding to the target in the relative conditions. Significant results are displayed in bold.

For these measurements the major interest lies in the target. Indeed, as shown by previous measures, the target in PF does not bear a pitch movement. The dimension of PEM is therefore useful to determine the degree of prominence in this position. Crucially, considering only the stressed syllables might result in missing important information of the movement and the corresponding energy within the words. Figure 33 enables to take a closer look at the target and to see the PEM values also for the first and the third syllables of the target (Syll 1 and Syll 3). Results demonstrate that the stressed syllable for NF has the highest values in comparison to the preceding and following syllables (Syll 1 and Syll 3), while for PF and BF values remain the same across the whole word. Values of the first syllable are higher in BF than in PF ($\beta = -0.1 \pm 0.03$, $p = 0.01$), while there is no difference between BF and NF ($\beta = -0.03 \pm 0.03$, $p = 0.27$). This is in line with the fact that the movement in the first syllable is generally high in BF and NF, while it is not present in the first syllable of the target in PF. The third syllable shows the same values in BF and NF, which are both higher than in PF ($\beta = -0.13 \pm 0.03$, $p < 0.0001$), in line with the presence of pitch movement in the former conditions.

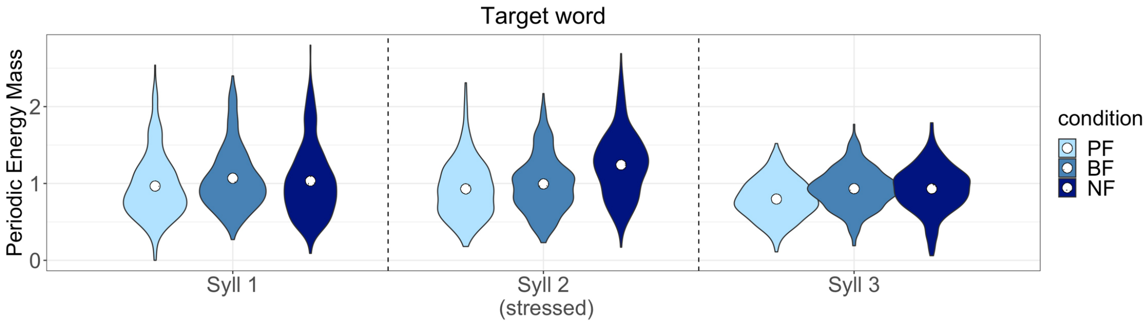


Figure 33. PEM for each syllable in the target (from left to right, Syll 1, Syll 2 and Syll 3) and for each condition (from the left to right, PF, BF. NF). Vertical dotted lines indicate syllable boundaries. The stressed syllable is indicated by the label stressed in parentheses. With dots indicate mean values.

Looking at the individual differences (Figure 34), most of the speakers appear to make a distinction between NF and the other two conditions in the stressed syllable. However, speaker F07, F10, M02 and M04 do not show this distinction, with NF showing a distribution of values close to BF. The lack of difference in the PEM of the target between BF and PF that was registered for the data pooled across participants is found in almost all the speakers: F01, F02, F03, F04, F08, F09, F10, M01, M02, M03 and M04. In the noun in PP the situation is more variable, with some of the speakers not showing differences between the conditions, and some showing higher values for PF than BF and just F04 showing higher values for BF than PF.

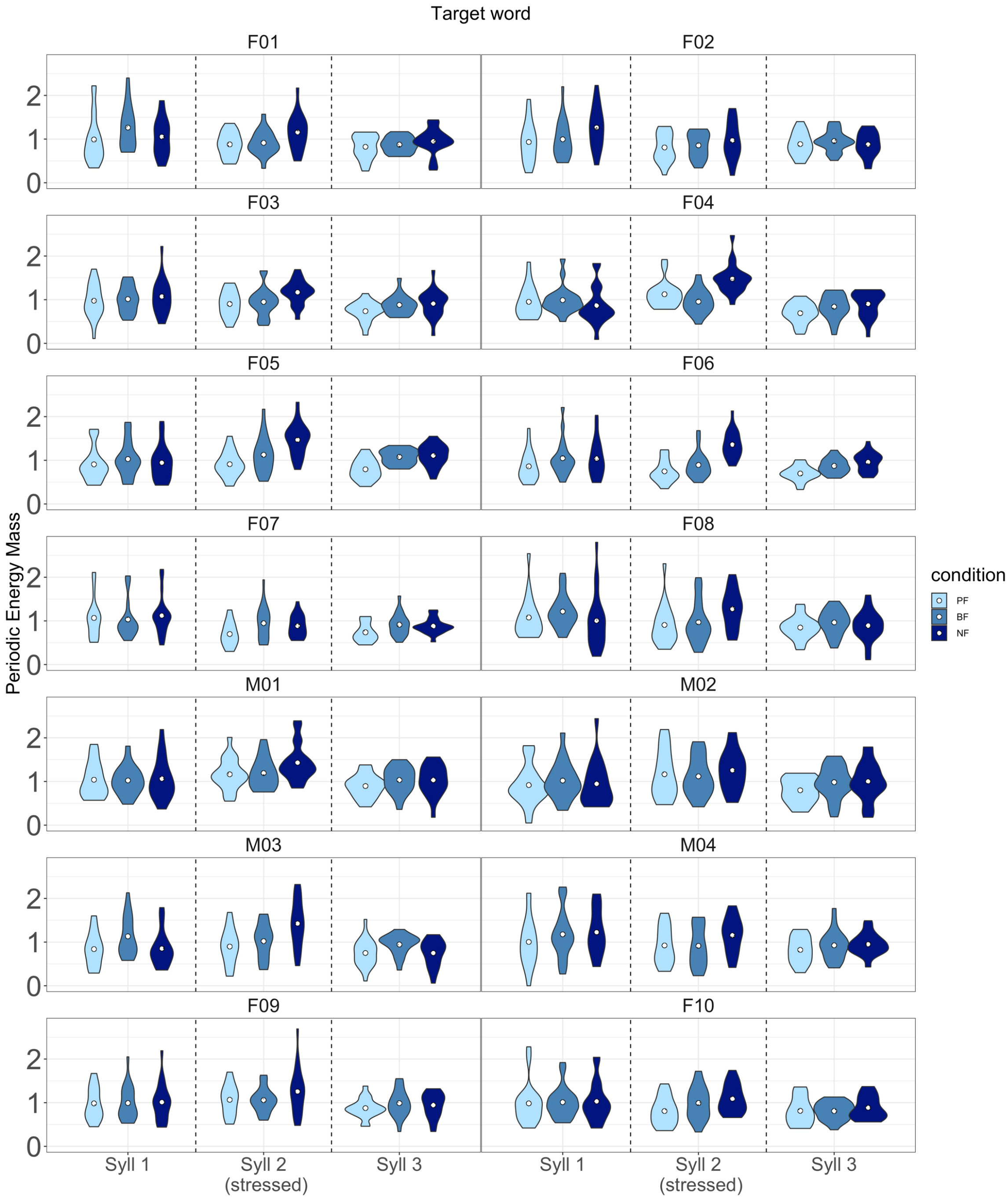


Figure 34. PEM for each syllable in the target (from left to right, Syll 1, Syll 2, Syll 3) and for each condition (from left to right, PF, BF, NF). Each panel shows productions for one speaker. The stressed syllable is indicated by the label stressed in parentheses. White dots indicate mean values. Vertical dotted lines indicate syllable boundaries.

Figure 35 shows the values for the domain of the whole word. Looking at this domain is useful since, considering values of the stressed syllables only, might result in the loss of information regarding the energy that might correspond to parts of pitch movement that start or end outside the domain of the stressed syllable. The models run on the differences between conditions within each word position registered a difference in both the target and the verb between PF and BF. This difference is not present for the noun in PP. In PF the verb shows the highest values of PEM compared to the target and the noun in PP, while both in BF and NF the target shows higher values than the verb and the noun in PP. Table 14 and Table 15 report the results of the models.

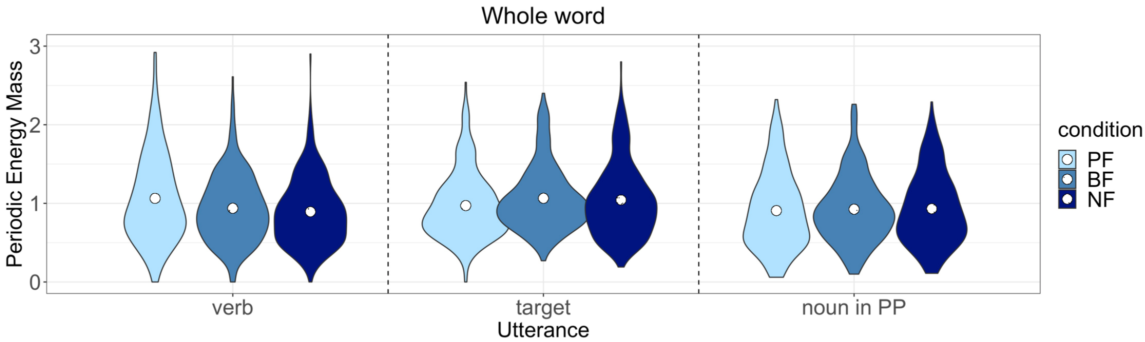


Figure 35. PEM for the domain of the whole word for each word position. From left to right panel: verb, target and noun in PP. Within each panel, from the left: PF, BF, NF. White dots indicate mean values.

	verb	**target**	**noun in PP**
PF	**β = 14.31 ± 3.69, p = 0.0001**	**β = -9.69 ± 3.71, p = 0.01**	β = -2.14 ± 4.45, p = 0.57
NF	β = -4.55 ± 3.66, p = 0.21	β = -2.45 ± 3.69, p = 0.51	β = 0.37 ± 3.72, p = 0.92

Table 14. Results of the mixed model analyses of the differences between the entire word's values in each condition in each word position. Values are relative to the intercept, corresponding to the relative critical word (verb, target and noun in PP) in broad focus. Significant results are displayed in bold.

	PF	**BF**	**NF**
verb	**β = 11.25 ± 4.16, p = 0.01**	**β = -12.74 ± 3.34, p < 0.01**	β = -14.87 ± 3.6, p < 0.0001
noun in PP	β = - 6.35 ± 4.19, p = 0.13	β = -13.92 ± 3.36, p < 0.001	β = -11.07 ± 3.63, p < 0.01

Table 15. Results of the mixed model analyses of the differences between each word position within each condition, from the left: broad focus, narrow focus and post-focal. Values are relative to the intercept, corresponding to the noun in PP in the relative conditions. Significant results are displayed in bold.

5.2.8 Discussion and conclusions (long dataset)

This experiment looked at the production of sentences uttered in three focal structures by speakers of the variety of Italian spoken in Udine. Results of the ToBI analysis registered a considerable variability. Nonetheless, general tendencies could also be identified: the majority of narrow focus exponents were realised with a steep rising accent (L+H*), the majority of words in the broad focus domain were realised with a falling accent (H+L*) and the majority of words in post-focal position without movement. These general trends in the prosodic marking of the focal structure are in line with what has been reported by the literature (see e.g., Gili-Fivela et al., 2015). The variability registered consisted in the presence of a minority of different accent types in each focal structure. For example, the presence of shallow rising accents (H*) concomitant with narrow contrastive focus or the presence of shallow-rising accents and falling accents in post-focal position. Results reveal that this variability is due to great inter-individual variability in the production of the different focus conditions, with some speakers making clear differentiations between the cases of broad focus, narrow focus on the target and narrow focus on the verb (i.e. target in post-focal position) and other speakers differentiating to a lesser degree between conditions. Moreover, a great intra-individual variability was found.

The experiment further investigated whether the differentiation between conditions could be better captured with the use of continuous parameters registering subtler modulations of the F0 and of the prominence degree than the ones accounted for by the discrete labelling. Therefore, F0 and measures related to periodic energy were computed. Results indicate that all the speakers had the tendency to distinguish between the conditions, with the target in narrow focus presenting the highest number of cues to prominence in comparison to the other two conditions (rising pitch and the highest values of PEM). The target in broad focus and post-focal position showed for the overall data a difference in the F0 modulation realised either at the stressed or at the preceding or following syllable. Differences were also registered in PEM of the syllables preceding and following the stressed syllable, but not on the stressed syllable. This tendency was present for the majority of the speakers. However, when considering PEM for the domain of the whole word, differences between conditions were statistically significant.

The high variability within and between participants in PEM and the lack of a difference in some cases between broad focus and the post-focal condition might indicate that speakers were not reliable in conveying the intended meaning and, hence, the particular prominence relations. One reason for this might be the unnatural makeup of the long answers. Therefore, a second set of data was collected with different stimuli. This second experiment is presented in the next section (5.3). The main change in the make-up of the sentences for the second set of recordings was to make them shorter compared to the first set, in order to allow them to be more similar to the ones that would occur in spontaneous conversation.

5.3 Second set of recordings (short dataset)

The objective of the present set of recordings was to investigate the production of prominence relations using stimuli, whose naturalness was improved compared to the ones used in the previous recordings. The improvement consisted in shortening the length of the sentences used as stimuli. This change in stimuli was expected to yield results more closely approximating the ones of spontaneous speech. In addition, a shorter answer was considered to be more suitable for the planning of the focal structure required by the context and appropriate to test whether part of the intra-speaker variability could be due only to the task requirements (see Niehbur & Michaud, 2015). By trying to lower the variability in the production of the utterances, the aim was to have further, possibly more reliable data on the production of the prominence relations and to have further evidence of the distribution of prominence in the post-focal position in this variety. This would further help in the interpretation of results in the perception study in Chapter 6.

The next subsections will provide a description of the experiment design and of the results. Furthermore, a comparison between the two datasets will be provided.

5.3.1 Materials

Stimuli were created by shortening the answers that speakers had to utter in the previous set (first set of recordings, long dataset, 5.2). While in the long dataset, speakers were required to repeat part of the question when answering, in the present set they were not. Thus, the productions resulted in shorter utterances. As in the first set, questions were used to elicit the different focal conditions, with the target word realised in broad focus (BF), narrow focus (NF) and post-focal position (PF). A comparison of the two sets of stimuli is exemplified in Table 16 (see Appendix A2 for the complete list of stimuli[16]), where the underlined part of the sentence (*When you go on a trip*) is the one that is present in the first set and absent in the second. The answers in the second set were considered to be more natural than the ones containing the repetition (first set). This hypothesis was also confirmed by the results of a questionnaire conducted with 45 native speakers of Italian (of the varieties spoken in Udine and in Bari). In this questionnaire, participants were presented with the question in Table 16, and with the two possible

[16] Note that the stimuli for the short dataset were almost all changed, with the exception of three items, in order to both avoid participants to recall sentences from the previous set of recordings and to better control the number of syllables of the verb and the noun in PP (verb either 3 or 4 syllables, noun in PP either 2 or 3 syllables).

answers (one from the long dataset, the other from the short dataset, see Table 16). From these two answers, they were required to select the one that they would use in a dialogue with a friend. 38 out of 45 participants chose the short answer.

Question	
Cosa bisogna fare quando si va in gita? *What does one need to do when going on a trip?*	
Answers	
Set 1 (long dataset)	**Set 2 (short dataset)**
Quando si va in gita, bisogna preparare un **panino** per la merenda. *When going on a trip,* *one needs to prepare a* ***sandwich*** *as a snack*	Bisogna preparare un **panino** per la merenda. *One needs to prepare a* ***sandwich*** *as a snack*

Table 16. Differences in the answers used in the two sets (on the left the first set of recordings, i.e., long dataset, on the right the second set of recordings, i.e., short dataset; questions were the same for both datasets).

5.3.2 Participants

Speakers that were recorded in the first set (long dataset) were also recorded for the second set (see 5.2.2). All participants gave written informed consent.

5.3.3 Procedure

To allow participants to forget about the first session of recordings, the present recordings were performed four months later. Similarly to the first session, in the present session speakers were digitally recorded at a sampling rate of 44 kHz with a 16-bit resolution and the question-answer pairs were presented on slides in a PowerPoint presentation in a pseudo-randomised order. Also in this case, slide questions elicited three different focal conditions in which the target word occurred in: broad focus (BF), narrow focus (NF) and in post-focal position (PF). The questions were uttered by the experimenter (native speaker of the same variety), in order to have a more homogeneous intonation with which participants were confronted in the questions. Therefore, another difference to the first set consists in the fact that participants only had to utter answers, never questions. The rest of the procedure was the same as in the first set: participants were asked to first read both question and answer silently, and to correct themselves if they thought that their production was incorrect or strange in any sense. Each recording session was preceded by a training session of five question–answer pairs, which did not enter the analysis; question-answer pairs were randomised to avoid sequence effects: the different conditions of an item were separated by at least 10 slides.

5.3.4 Analysis

The analyses conducted on this dataset were the same as described in 5.2.4. In addition, analyses of the comparison of the two datasets were conducted. To analyse the differences between groups for each parameter (synchrony, scaling and PEM) the Levene Test of homogeneity of variance was performed. To perform this test the function leveneTest() from the *car* package (Fox & Weisberg, 2019) was used.

5.3.5 Results of the ToBI analysis

Results of the pitch accent types are displayed in Figure 36. This figure shows the distribution of the ToBI labels for each focus condition and each position of the word in the utterance. Results are pooled over speakers. The figure shows the presence of some variability in the use of pitch accent types across the focal conditions. Nonetheless, it also shows preferences for certain pitch accent types for condition and position of the word. From the inspection of the graph, the variability present in this dataset appears to be less than in the previous dataset (long sentences, see 5.2.5). The focus exponents of the utterance, i.e., the verb in PF and the target in NF are mostly realised with a rising (L+H*) or with a (rising-)falling (H*+L) accent, typical of contrastive focus (76.52 % of the cases in verb in PF, 82.2 % of the times in target in NF). The use of H* and H+L* accents is also present, but to a considerably less extent (23.49% of the cases in the verb in PF, 17.67 % of the cases in target in NF).

In PF, the target was in most of the cases realised with a flat contour (77.74% of the cases, a higher percentage than in the previous set), while in the other cases were present either falling accents (H+L*, the 15.47% of the cases) or, in lesser amount, shallow rises (H*, 6.42% of the cases). Only one occurrence was produced with a rising pitch accent. The target occurring in BF, is mostly realised with a falling contour (H+L*, 56.6% of the cases). The rest of the cases present shallow rises (H*, in the 29.43% of the cases) and more steep rises (L+H*, in the 5.66% of the cases), while a very small amount seems not to be marked by a movement in pitch (8.3% of the cases). In the present dataset, as in the previous dataset, in the majority of the cases the target ended with a rise (H-, the 88.07% of the cases).

The noun in PP, in BF is mostly realised with a falling contour (H+L*, 89.06% of the cases), in NF and PF with no movement (71.81% of the cases in NF, 72.24% of the cases in PF), which is mostly attested for elements in post-focal position in long statements (see 2.3.3).

As observed for the previous dataset, the distribution presents inter-speaker variability. Figure 37 shows the pitch accents' distribution separately for each speaker. Looking at the production of individual speakers it can be noticed that all speakers use all accent types to distinguish the conditions, even if in a different proportion. Speaker F04, F06 and M03 mark the focal exponent of the utterance in PF and NF with a L+H* accent 100% of the cases, while verb and target in BF are marked by these speakers by either a falling accent (H+L*, for F04 the 68.42% of the cases, for F06 the 73.68% of the cases, for M03 the 50.88% of the cases) or by a slightly rising accent (H*, for F04 the 22.81%, for F06 the 21.05% of the cases, for M03 31.58% of the cases). It should be noted that for speaker M03 a certain percentage of targets in BF is realised with L+H* accent (36.84% of the cases). Speakers F01, F02, F03, F05, F08, F09, F10 and M01 use mostly rising L+H* or (rising-)falling H*+L accents to mark the focal exponent in NF and PF, even if the percentage does not correspond to 100% as in the case of the previously mentioned speakers. Unlike all speakers so far listed, speaker F07 uses the falling contour in the focal exponent of NF and PF to a high percentage (the 36.84% of the cases in target in NF, the 57.9% of the cases in verb in PF), while speaker M01 and speaker M02 use rising L+H* pitch in the target when it is realised in PF (M01 the 5.26% of the cases, M02 the 38.89% of the cases).

The inspection of the graphs of the second set (Figure 36 and 37) and the first set (Figure 24 and 25) reveals that the variability in the accent types is higher in the first set than in the second.

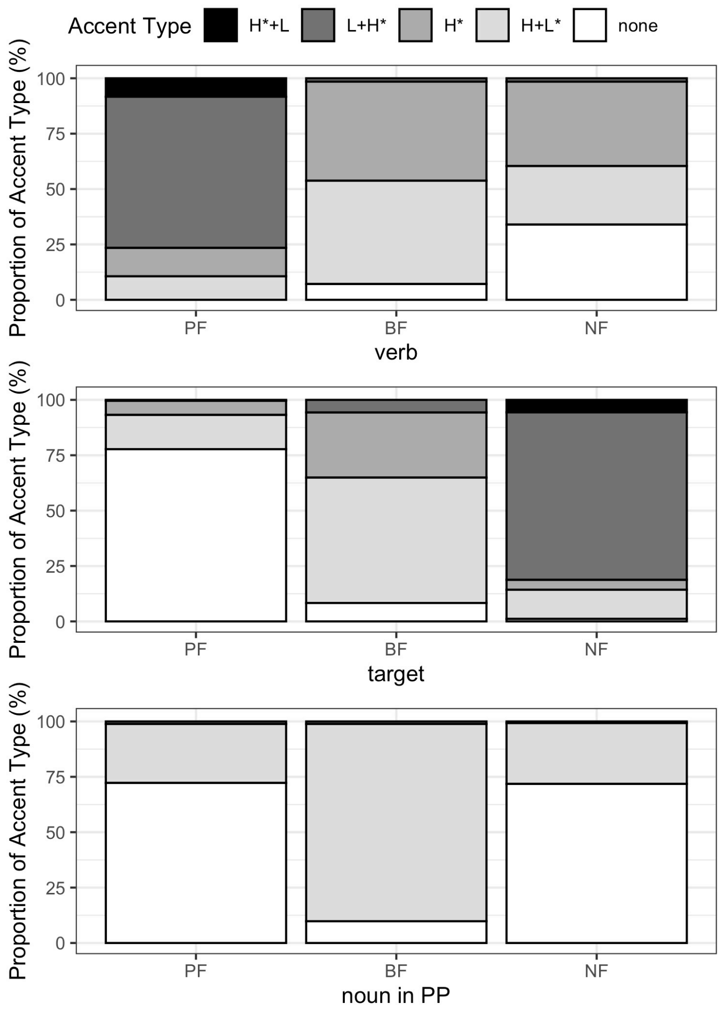


Figure 36. Pitch accent types' distribution. Data are pooled across speakers.

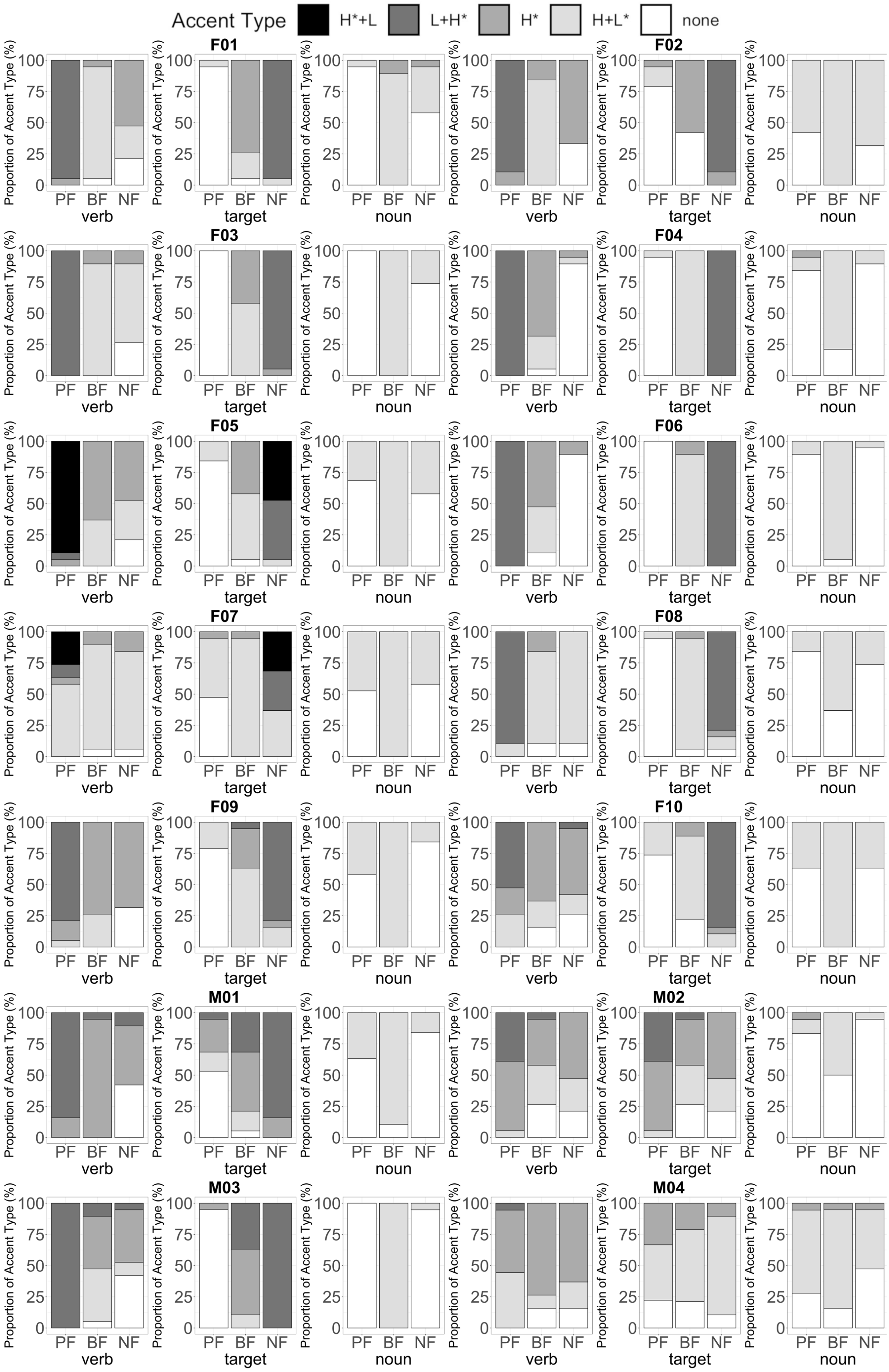


Figure 37. Distribution of pitch accent types across conditions and word position for each speaker separately.

In the next section we explore the continuous measures of synchrony, scaling and PEM to detect quantifiable difference in the realisation of prominence in the different conditions and word positions. Moreover, these measures will help to quantify the differences between the two sets.

5.3.6 Global contours

The superimposed contours in Figure 38 (female speakers) and Figure 39 (male speakers) show intra- and inter-individual variation in the three focus conditions, but a clear tendency to distinguish the conditions can be noted in almost all the speakers, with the exception of speaker M04, for whom the differences are difficult to notice by just looking at the contours.

Table 17 reports F0 means and standard deviations (in brackets) for each condition and for each position of the word at the target height, corresponding to the starred tone of the ToBI transcription. For post-focal position the reference for measurements was placed at the middle of the total duration of the stressed syllable.

condition	**F0 verb** **[Hz]**	**F0 target** **[Hz]**	**F0 noun in PP** **[Hz]**
PF	Male 116.62 (34.88) Female 214.43 (28.63)	Male 91.69 (9.97) Female 170.31 (14.82)	Male 87.45 (12.48) Female 168.75 (16.97)
BF	Male 101.64 (5.81) Female 191.41 (17.56)	Male 94.88 (9.42) Female 187.24 (20.14)	Male 91.11 (20.1) Female 165.47 (23.64)
NF	Male 102.49 (10.7) Female 192.93 (14.6)	Male 102.41 (20.44) Female 208.49 (24.99)	Male 83.84 (88.02) Female 165.38 (23.64)

Table 17. F0 mean and standard deviation (in brackets) at the reference point for each word position and prosodic condition of the stimuli. Values for male and female are shown separately.

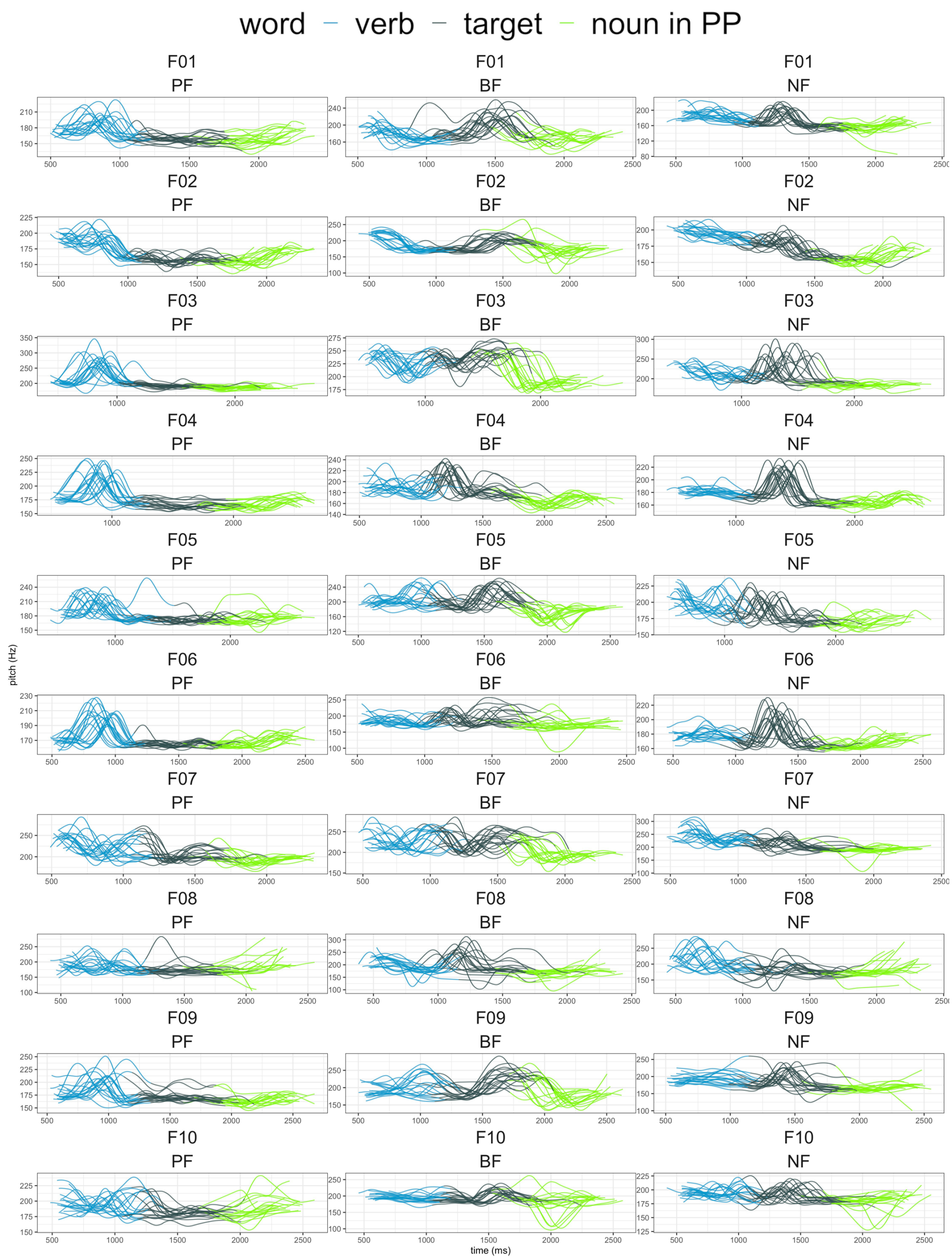


Figure 38. Superimposed contours for each intended focal condition (from left to right: post-focal position, broad focus and narrow focus) for each female speaker separately. Different colours of the contours indicate the critical words (verb, target and noun in PP).

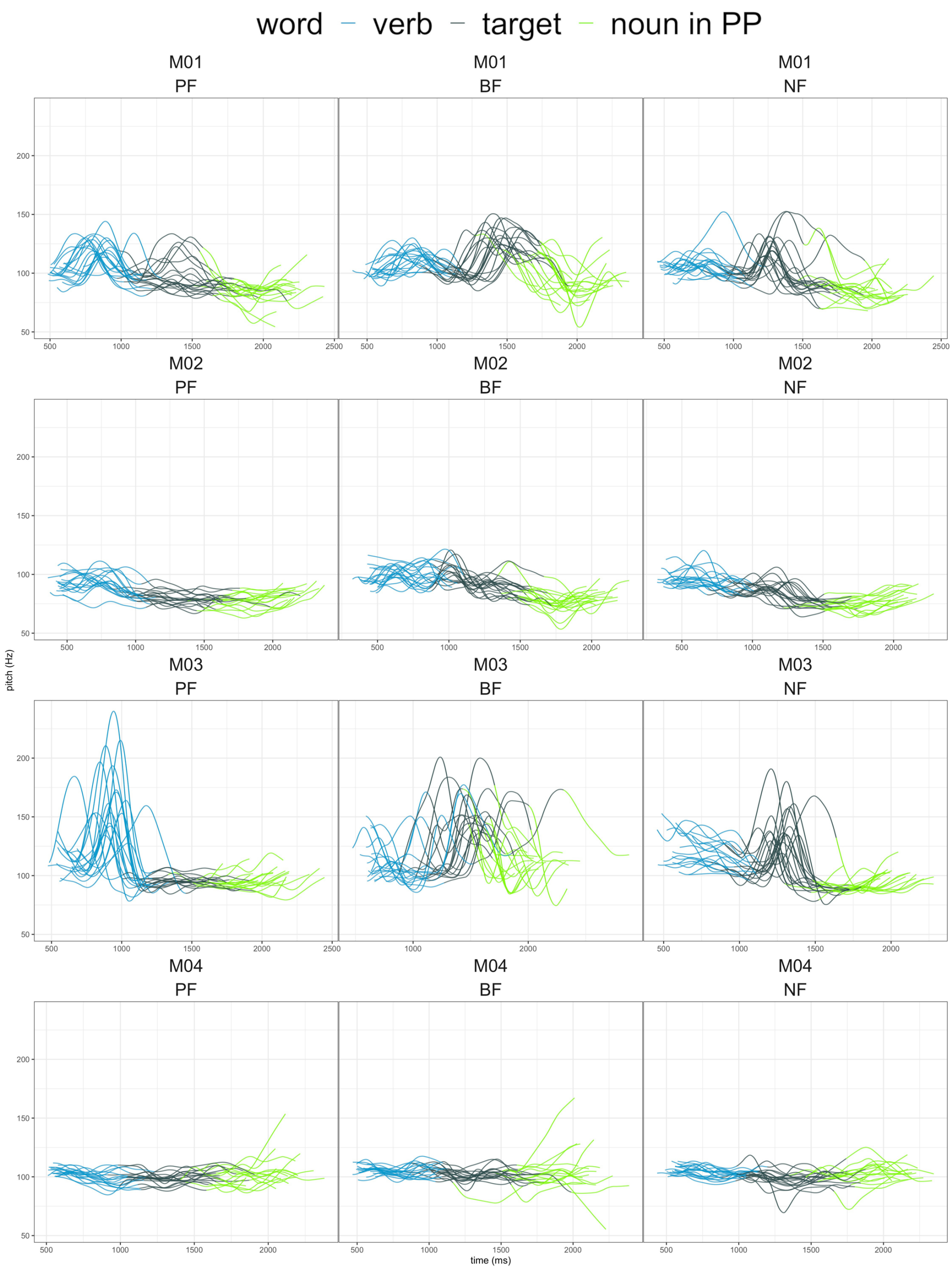


Figure 39. Superimposed contours for each intended focal condition (from left to right: post-focal position, broad focus and narrow focus) for each male speaker separately. Different colours of the contours indicate the critical words (verb, target and noun in PP).

5.3.7 Synchrony, scaling and Periodic Energy Mass

Looking at the measures of synchrony and scaling for the whole contour (Figure 40), a tendency towards a distinction of the conditions can be seen. Compared to the first set, in the current set differences between conditions in the modulation of F0 within syllables seem to be more pronounced. Table 18 reports means and standard deviations of the values of synchrony and scaling for the stressed syllables of all the critical words in all the conditions. Stressed syllables of the verb show a tendency towards positive values of synchrony for PF, negative values for BF and a distribution of values around zero for NF. Stressed syllables of the target reflect the highest positive values for NF, while the distribution of values for the target in PF is more condensed around zero and BF shows a bimodal distribution, with some positive and some negative values. In addition, values of synchrony of the last syllable of the target (Syll 6) also present a high differentiation among conditions. Finally, in the stressed syllable of the noun in PP, values for PF and NF are similar, which is expected given that both these words in these conditions occur in post-focal position. By contrast, BF shows a trend towards negative values.

Differences among conditions can be noticed also in scaling. Stressed syllables of the verb have positive values for PF, and negative values or values around zero for BF and NF. Stressed syllables of the target show positive values for NF, while the distribution of values for the target in PF is more condensed around zero, and BF reflects negative values. The last syllable of the target (Syll 6) presents the same pattern for scaling that is also found in synchrony. In the stressed syllable of the noun in PP, values for PF and NF are similar and condensed around zero, while BF shows negative values.

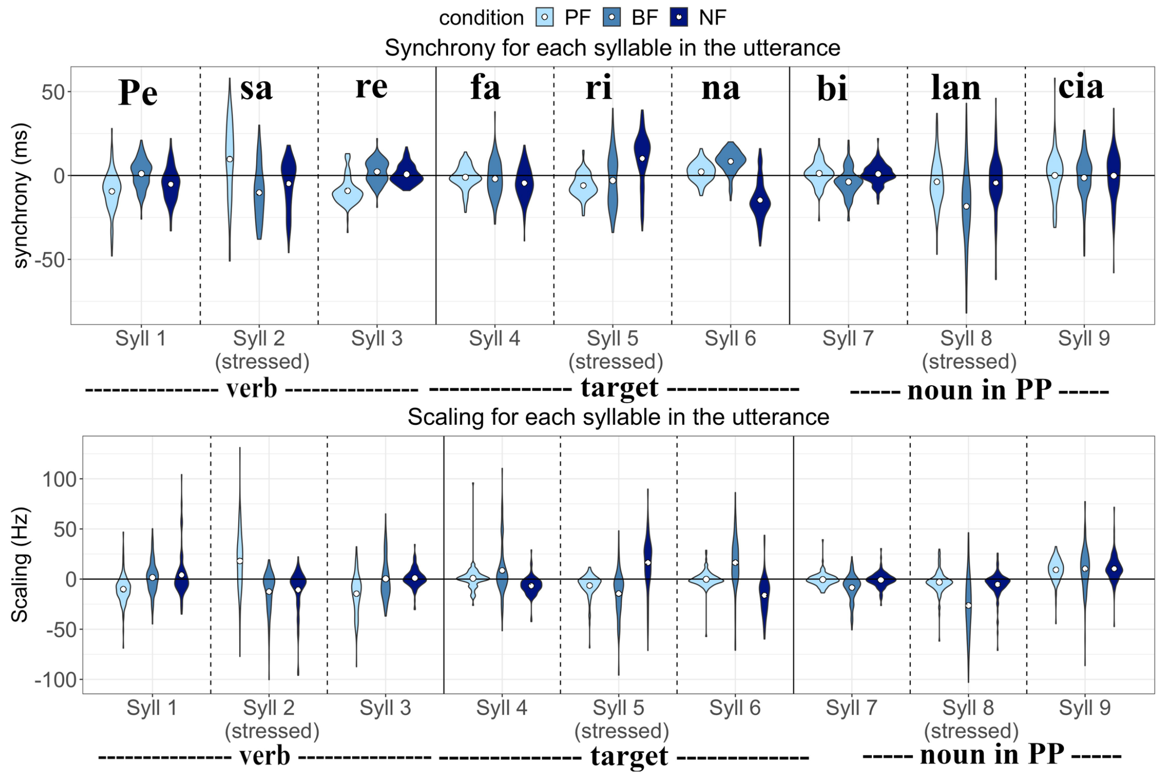


Figure 40. Synchrony and scaling of each syllable of the critical words. Stress on the second syllable of the verb and on second syllable of the noun. Black dots on the violin plots indicate the mean value. Each box in the graphs corresponds to a syllable in the utterance and presents values relative to this syllable for each condition (from left, PF, BF and NF). Note that the values of scaling for the first syllable are presented because they refer to the preceding syllable (*gna* of the word *bisogna*) which was available in the presented data. Solid vertical lines correspond to the boundaries of each critical word (from the left, verb, target and noun in PP), while dotted vertical lines indicate the boundaries of the syllables. Labels on the top provide an example of how each syllable is displayed and contains the three syllables of the critical words of the example utterance *Bisogna pesare la farina con la bilancia* (One should weigh the flour with the scales).

condition	**verb**	**target**	**noun in PP**
PF	**Synchrony:** 11.41 (19.78) **Scaling:** 21.29 (26.96)	**Synchrony:** -6.39 (7.18) **Scaling:** -7.34 (11.51)	**Synchrony:** -0.01 (17.64) **Scaling:** -5.11 (12.19)
BF	**Synchrony:** -9.29 (17.42) **Scaling:** -11.83 (18.57)	**Synchrony:** -2.6 (16.77) **Scaling:** -14.63 (20.39)	**Synchrony**: -20.32 (24.42) **Scaling:** -35.83 (30.20)
NF	**Synchrony:** -4.17 (16.08) **Scaling:** -8.18 (19.25)	**Synchrony:** 13.96 (14.73) **Scaling:** 18.37 (24.83)	**Synchrony**: -2.36 (18.36) **Scaling:** -6.82 (14.83)

Table 18. Mean and standers deviation (in brackets) of synchrony and scaling values for the stressed syllables of the critical words in all conditions for all the stimuli.

Results of the models run for each position of the word on the differences between conditions are reported in Table 19, where the intercept of the model is BF. These results show that in the present set there is clear differentiation across conditions. In more detail, results indicate that the stressed syllable of the verb in PF presents higher values of synchrony compared to the verb in BF (the first set did not reach significance). The verb in NF is, again, not differently realised compared to the verb in BF. By contrast, in this set the target in PF is distinguished from the target in BF by means of scaling rather than synchrony, a trend opposite to the first set. In NF, the target shows a difference to the target in BF in both values (synchrony and scaling; in the previous set, there was only a contrast in scaling), while the noun in PP also registered differences in all the values among all conditions.

	verb	**target**	**noun in PP**
PF	**Synch: β = 19.51 ± 5.47, p = 0.004** **Scal: β = 30.92 ± 6.17, p < 0.001**	**Synch:** β = -2.24 ± 3.07, p = 0.48 **Scal: β = 7.84 ± 2.68, p = 0.01**	**Synch: β = 17.92 ± 3.26, p < 0.001** **Scal: β = 26.59 ± 4.45, p < 0.001**
NF	**Synch:** β = 5 ± 3.06, p = 0.13 **Scal:** β = 3.13 ± 2.63, p = 0.26	**Synch: β = 16.03 ± 4.21, p < 0.01** **Scal: β = 31.19 ± 5.1, p < 0.0001**	**Synch: β = 15.51 ± 3.34, p < 0.001** **Scal: β = 25.13 ± 4.36, p < 0.001**

Table 19. Results of the mixed model analyses of the differences between the stressed syllables' values of synchrony (synch) and scaling (scal) in each condition in each word position. Values are relative to the intercept, corresponding to the relative critical word (verb, target and noun in PP) in broad focus. Significant results are displayed in bold.

Table 20, Table 21 and Table 22 report the results for the difference of the stressed syllable of each critical word within each condition. As in the first set, in post-focal position (Table 20) the verb has higher values of synchrony and scaling compared to the target and again the noun in PP shows a reduced movement within the stressed syllable compared to the stressed syllable of the target, but not across the stressed syllable and the preceding one. In broad focus (Table 21), the verb is distinguished from the target only in synchrony, while the noun in PP is distinguished in both dimensions.

In narrow focus (Table 22), the stressed syllable of the verb shows reduced values of synchrony and scaling from the preceding syllable compared to the target and a similar reduction is registered also for the noun in PP.

	PF
verb	**Synch: β = 15.51 ± 1.26, p < 0.0001**; **Scal: β = 26.46, ± 1.42, p < 0.0001**
noun in PP	**Synch: β = 6.85 ± 1.26, p < 0.0001**; **Scal:** β = 2.17 ± 1.42, p = 0.13

Table 20. Results of the mixed model analyses of the differences between each word position in post-focal position. Values are relative to the intercept, corresponding to the target in the relative condition. Significant results are displayed in bold.

	BF
verb	**Synch: β = - 6.38 ± 1.57, p < 0.0001**; **Scal:** β = 3.06 ± 1.81, p = 0.09
noun in PP	**Synch: β = -13.34 ± 1.57, p < 0.0001**; **Scal: β = -16.57 ± 1.8, p < 0.0001**

Table 21. Results of the mixed model analyses of the differences between each word position in broad focus. Values are relative to the intercept, corresponding to the target in the relative condition. Significant results are displayed in bold.

	NF
verb	**Synch: β = - 17.4 ± 1.33, p < 0.0001**; **Scal: β = - 24.96 ± 1.59, p < 0.0001**
noun in PP	**Synch: β = -13.89 ± 1.33, p < 0.0001**; **Scal: β = -22.59 ± 1.6, p < 0.0001**

Table 22. Results of the mixed model analyses of the differences between each word position in narrow focus. Values are relative to the intercept, corresponding to the target in the relative condition. Significant results are displayed in bold.

Looking at the individual differences (Figure 41 and Figure 42), the tendency towards the distinction of the conditions is present in every speaker. The continuous values underline a slight differentiation of the conditions even in speaker M04, which was not registered in the categorisation in accent types and in the graph showing the contours. The target in PF is always realised with synchrony and scaling values around zero, showing an almost complete lack of movement.

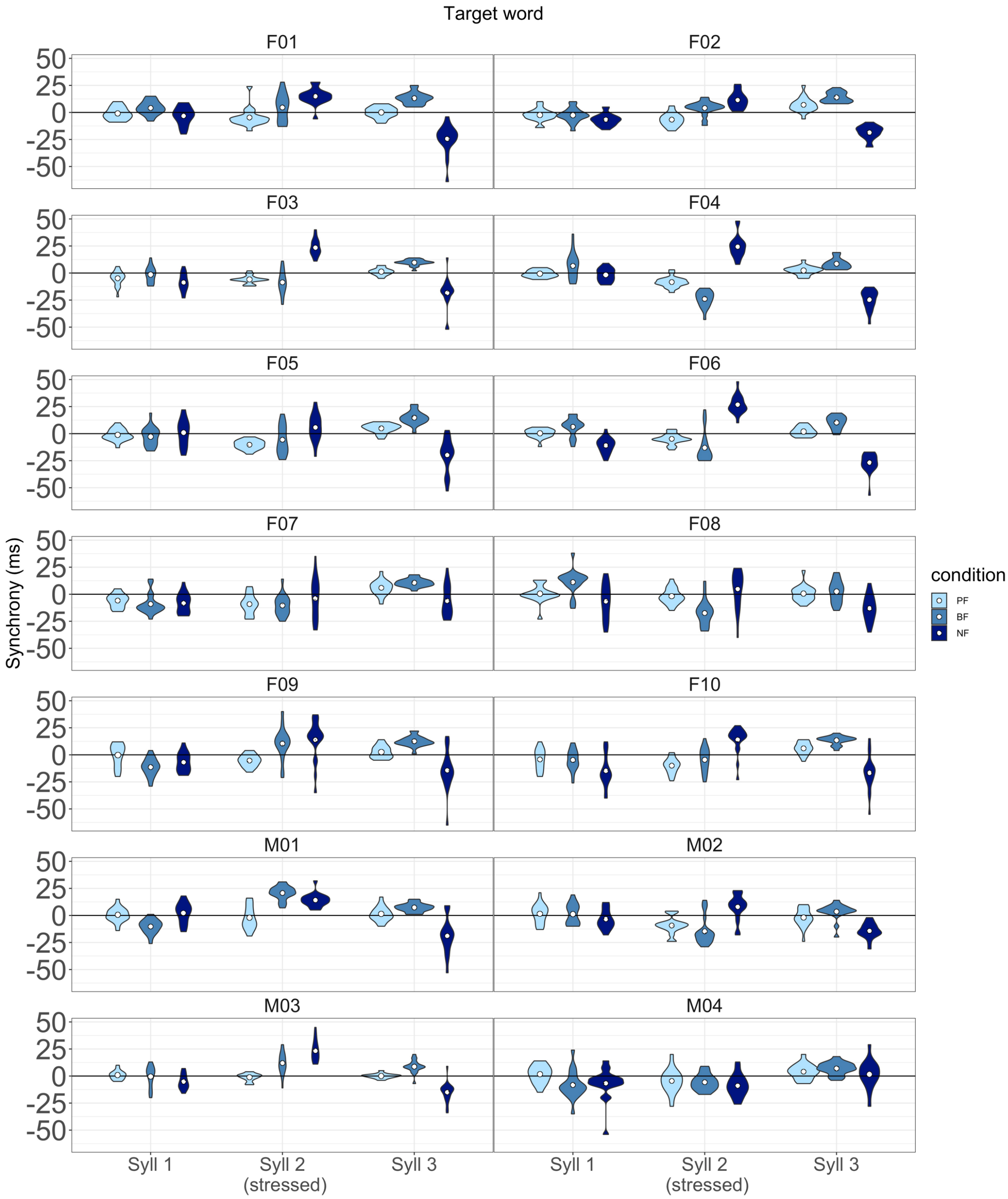


Figure 41. Synchrony of each syllable of the target word for each condition for each speaker (one speaker per panel). Stressed syllable is indicated by the label stressed in parentheses. White dots indicate mean values.

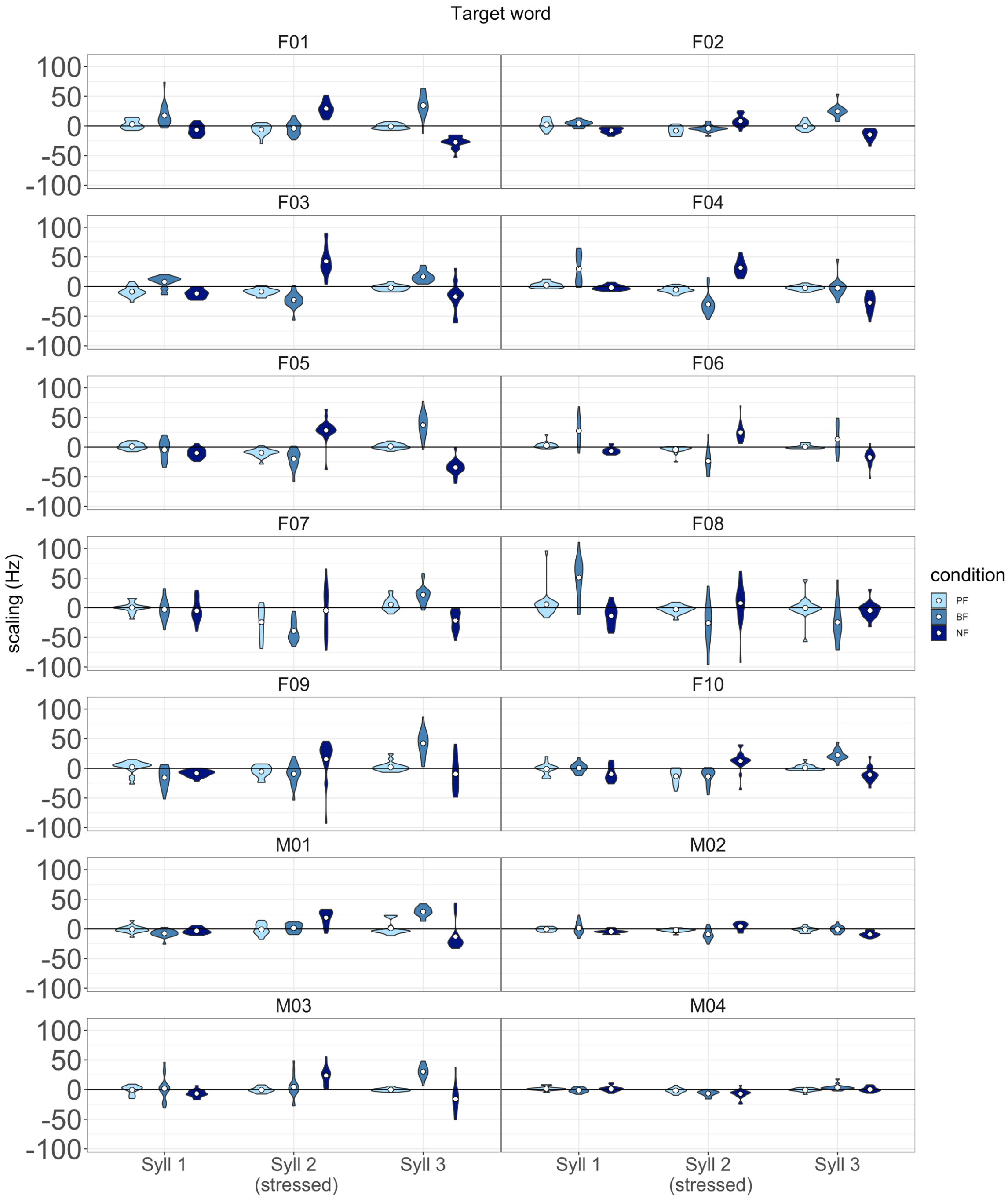


Figure 42. Scaling of each syllable of the target word for each condition for each speaker (one speaker per panel). Stressed syllable is indicated by the label stressed in parentheses. White dots indicate mean values.

Figure 43 shows values of PEM for the stressed syllables of each critical word. Results of the statistical model are reported in Table 23.

Results (Table 23) show higher values of PEM for the verb in PF compared to the verb occurring in the other two conditions. For the target, stressed syllables in NF show higher values than in BF, which in turn shows higher values than in PF. Moreover, the majority of values for the noun in PP occurring in PF and NF conditions are higher than expected. In fact, the distribution of values in Figure 43 indicates that a consistent amount of values is close to, or higher than the average (either around 1 or higher, see 5.2.4). By contrast, values for the noun in PP are expected to be lower than the average for NF and PF, given that they occur in post-focal position. Nonetheless, in the other word positions, the tendency to distinguish between conditions by PEM is present: in PF the verb has the highest value, in NF the target has the highest value and in BF values are similar for each stressed syllable of each critical word in the utterance.

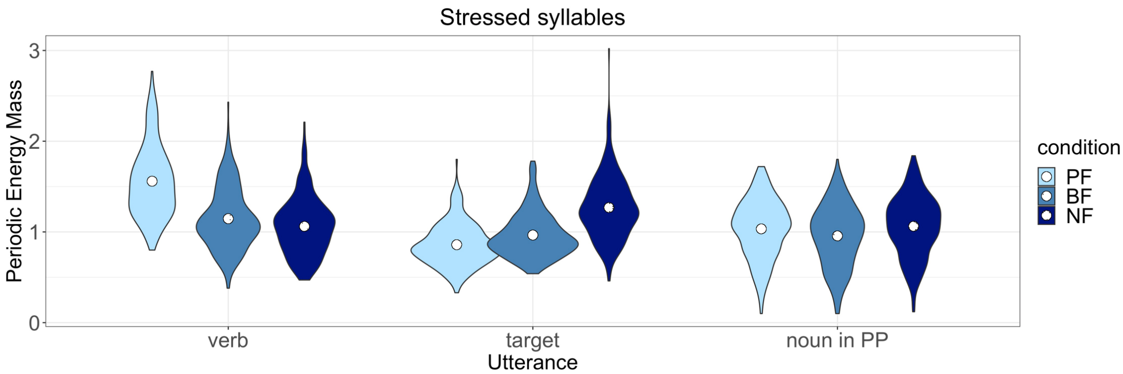


Figure 43. PEM values of the stressed syllables for each critical word in each condition. White dots indicate mean values.

	verb	target	noun in PP
PF	**β = 0.41 ± 0.03, p < 0.0001**	**β = -0.10 ± 0.02, p = 0.001**	β = 0.08 ± 0.04, p = 0.051
NF	**β = -0.09 ± 0.03, p < 0.01**	**β = 0.3 ± -0.04, p < 0.0001**	**β = 0.10 ± 0.04, p = 0.01**

Table 23. Results of the mixed model analyses of the differences between the stressed syllables' values in each condition in each word position. Values are relative to the intercept, corresponding to the relative critical word (verb, target and noun in PP) in broad focus.

Furthermore, for BF and NF, the first and third syllable of the target (see Figure 44, Syll 1 and Syll 3 respectively) are not distinguished for condition, while the stressed syllables are (Syll 2; $\beta = 30.37 \pm 2.11$, $p < 0.0001$). By contrast, PF shows differences in the conditions for all the three syllables, registering the lowest values (comparison with BF: Syll 1: $\beta = - 10.61 \pm 2.06$, $p < 0.0001$; Syll 2: $\beta = - 10.45 \pm 2.06$, $p < 0.0001$; Syll 3: $\beta = - 12.26 \pm 2.06$, $p < 0.0001$).

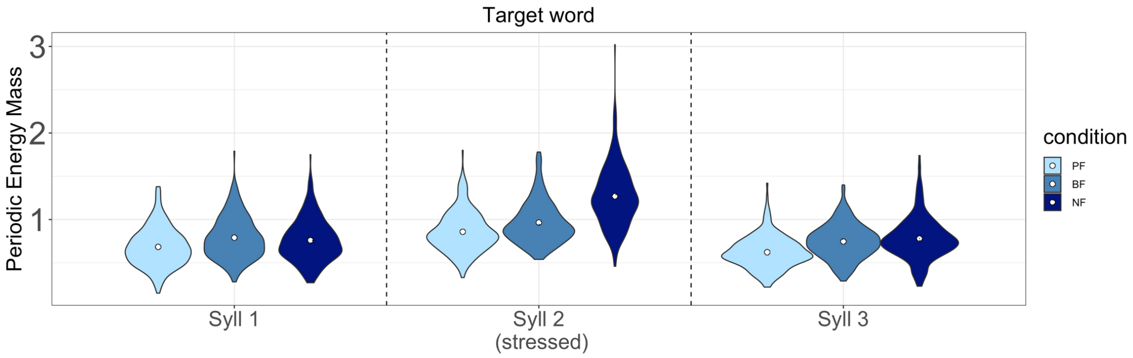


Figure 44. PEM for each syllable in the target word. Stressed syllable is indicated by the label stressed in brackets.

The same overall tendency presented in Figure 44, is found in almost every speaker (Figure 45), with the exception of speaker F04, F03, M01 and M02. For these speakers, PEM values of the stressed syllable of the target words occurring in broad focus and post-focal position do not consistently differ. However, among these four speakers, three (F03, M01 and M02) show a tendency towards an increase of PEM values either in the first (M02, F03) or in the last (M01) syllable of the target occurring in broad focus compared to the target occurring in post-focal position, thereby showing a distinction in the conditions.

With regard to the domain of the entire word (Figure 46), the differences in the PEM values of the target between PF and BF remain significant (as for the domain of the syllable reported above), while there is no difference between BF and NF (contrary to the domain of the syllable reported above). In addition, the verb in PF shows the highest values of PEM compared with the target and the noun in PP, which in turn has higher PEM values compared to the target. Also in BF and NF the target presents lower values than the noun in PP, while the verb in BF shows higher values and the verb in NF does not show a difference. Table 24 and Table 25 report the results of the statistical analyses.

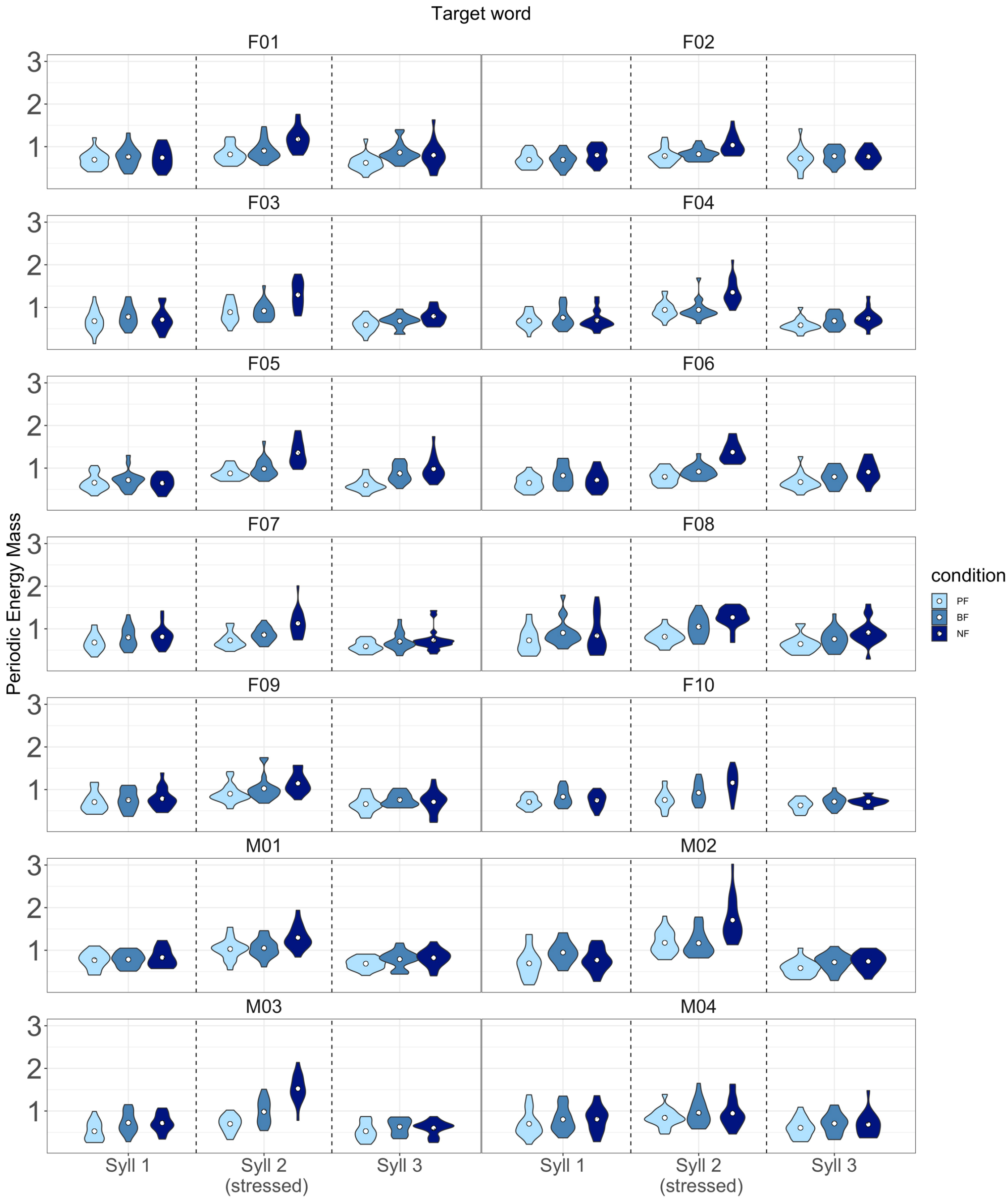


Figure 45. PEM for each syllable in the target (from left to right, Syll 1, Syll 2, Syll 3) and for each condition (from left to right, PF, BF, NF). Each panel shows productions for one speaker. The stressed syllable is indicated by the label stressed in parentheses. Black dots indicate mean values. Vertical dotted lines indicate syllable boundaries.

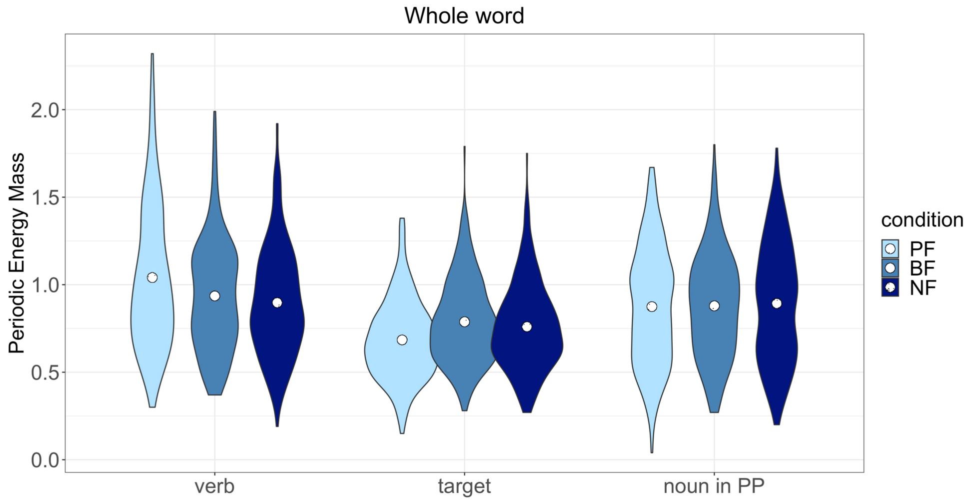


Figure 46. PEM for the domain of the whole word for each word position. From left to right panel: verb, target and noun in PP. Within each panel, from the left: PF, BF, NF. Black dots indicate mean values.

	verb	**target**	**noun in PP**
PF	**β = 0.11 ± 0.03, p < 0.0001**	**β = -0.1 ± 0.03, p < 0.0001**	β = -0.01 ± 0.03, p = 0.83
NF	β = -0.04 ± 0.03, p = 0.14	β = -0.03 ± 0.03, p = 0.29	β = 0.01 ± 0.03, p = 0.65

Table 24. Results of the mixed model analyses of the differences between the entire word' values in each condition in each word position. Values are relative to the intercept (BF), corresponding to the relative critical word (verb, target and noun in PP) in broad focus. Significant results are displayed in bold.

	PF	**BF**	**NF**
verb	**β = 0.02 ± 0.02, p < 0.0001**	**β = 0.06 ± 0.02, p = 0.02**	β = 0.01 ± 0.02, p = 0.85
target	**β = - 0.2 ± 0.02, p < 0.0001**	**β = -0.09 ± 0.02, p < 0.001**	**β = -0.13 ± 0.02, p < 0.001**

Table 25. Results of the mixed model analyses of the differences between each word position within each condition, from the left: broad focus, narrow focus and post-focal. Values are relative to the intercept, corresponding to the noun in PP in the relative conditions. Significant results are displayed in bold.

5.3.8 Discussion and conclusions (short dataset)

The categorical analysis of the short dataset confirms the lack of one-to-one mapping between form and function, as already shown by the results of the long dataset and by previous studies (e.g., Grice et al., 2017). Nonetheless, a preference for marking focal exponents with certain accent types has been observed. The focal exponents indicating a correction are realised similarly to the previous dataset. The majority of these occurrences were realised either with a rising accent or with a (rising-)falling accent (L+H* or H*+L). Similarly to the previous dataset, the majority of the targets occurring in broad focus were realised with a falling accent (H+L*), while targets in post-focal position were realised with a flat and low contour. The percentage of these accent types referring to each focal condition was higher in the short dataset than in the long dataset. However, part of the occurrences of certain focus types were realised with different accent types. For example, a considerable number of narrow focus exponents was also realised with falling accents, words in broad focus were also realised with shallow rises and the post-focal position was also realised with falling or shallow rising accents.

The systematic investigation of the prosodic marking of information structure in the Udine variety conducted in this chapter, has shown that also in this variety, speakers have the tendency to mark different domains and the sizes of focus with different accent types. The general patterns identified and described in the previous paragraph are the same reported for some other varieties of Italian analysed in Gili-Fivela et al. (2015). In particular: (i) in all varieties investigated by these authors, words occurring in broad focus statements are realised with a falling pitch accent (H+L*), as the one found here for the Udine variety; (ii) the marking of the focal exponents of contrastive-corrective narrow focus with a steep rising pitch accent (L+H*) typical of a large number of varieties (Milan, Turin, Florence, Siena, Lucca, Naples, Salerno, see Gili-Fivela et al., 2015:160) is also typical of the Udine variety. In addition, the patterns identified in the present study for the Udine variety of Italian show some similarities to the ones identified by Roseano et al. (2015) for the Friulian language. For example, the focal exponent of contrastive statements in the study of Roseano et al. was reported to be a steep rising accent (L+H*), and the focal exponent of broad focus statement is reported to be a falling accent (H+L*). Both these prosodic marks are found in the present experiment. A difference in the results of Roseano et al. compared to the results of the present experiment is detectable in the pre-nuclear accents of broad focus statements, which in the former study are reported to be late rising accents (L*+H), where in the present study they are mostly identified as falling accents (H+L*). Interestingly, the study on Friulian, reported the presence of H*+L accents to indicate epistemically biased statements (e.g., expressing doubt or obviousness). Some rare occurrences of this accent type are found also in the present experiment. In this case this accent type may have been used by speakers to express some degree of obviousness relative to the answers they needed to utter.

Continuous parameters measured in the data helped to identify the similarities in the tendencies that every speaker adopted to distinguish between conditions. Indeed, speakers differentiated between conditions both through F0 modulations and through PEM modulations. As for the main question concerning the degree of prominence of post-focal position, it has already been shown in the previous dataset (long dataset) that despite some variability, in the Udine variety of Italian the post-focal region is realised without substantial movement in pitch. The analysis of the present dataset (short dataset) confirms this result. Moreover, this second dataset shows that, in addition to the flat and low contour, stressed syllables in the post-focal position are also characterised by low duration and low energy. Indeed, participants made a three-way distinction among words occurring in the three different focal positions: prominence patterns are distinguished both through the modulation of F0 on the critical words and of PEM values. Results therefore indicate the presence of systematic modifications in duration and energy to signal fine-grained differences in the degree of prominence, with broad focus being realised with an increase in prosodic strengthening compared to post-focal position and with a lower level of prosodic strengthening compared to narrow contrastive focus.

The increase in prominence from the constituents occurring in post-focal position to the constituents occurring in broad focus, corresponded to a more extreme distribution of synchrony and scaling values (reaching either more negative or positive values). The increase in prominence between constituents occurring in broad focus and in contrastive focus, corresponded to an enhanced probability of finding positive and more extreme values for synchrony and scaling. In addition, parallel to the stepwise increase in the structural prominence of the constituents (from constituents occurring in post-focal position, in broad focus and in contrastive focus), all speakers showed an increase of PEM values. Taken together the results of the categorical and continuous analysis confirm the predictions that an intonation distinction between broad focus and post-focal position involves both a categorical distinction in terms of presence or absence of pitch movement, as well as gradual changes in the acoustic continuous parameters. The acoustic continuous parameters, together with the categorical change in pitch accent type were also gradually modified to signal the differences between broad and contrastive focus.

The present results provide additional evidence coming from a variety of Italian, to the data that have been reported by Mücke and Grice (2014) and by Roessig and Mücke (2019) for German. These studies have shown that prosodic prominence is expressed not only by the presence or absence of pitch movement to signal the change between constituents in-focus and constituents out-of-focus, but is expressed also by gradual changes within the in-focus group. Results of the present experiment show that also in Italian, for words occurring in the same position of the metrical structure, changes connected to the sonority expansion strategy (increase in duration and in energy), are controlled to express different degrees of prominence. In addition, unlike the results reported by Mücke and Grice (2014) the present results indicate that the tendency towards an increase of energy and duration of the stressed syllables in broad focus compared to post-focal position, is systematic (being present in all speakers with only one exception), despite the target in broad-focus bearing a prenuclear and not a nuclear accent.

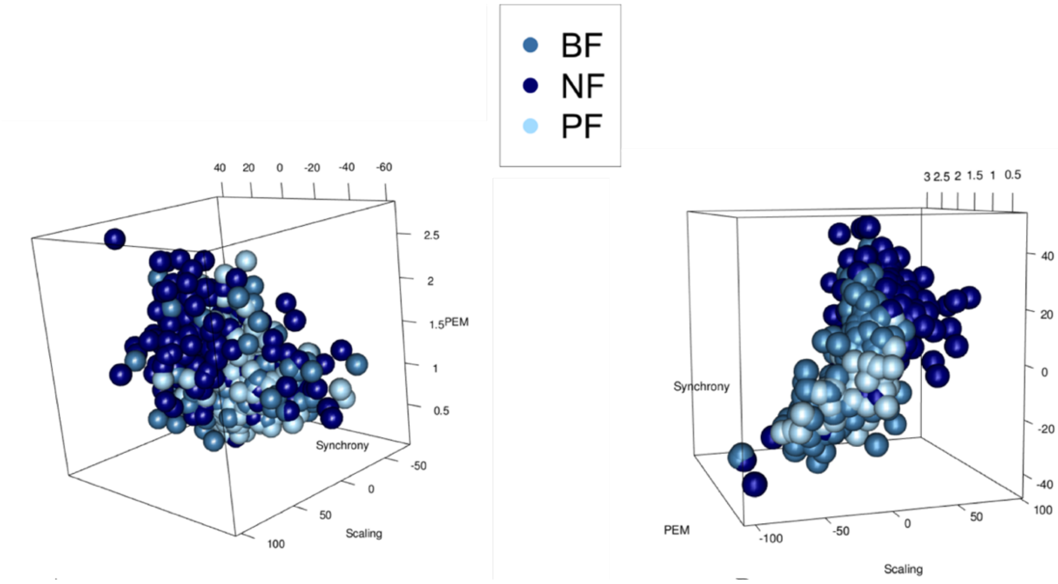
It should be noted that these results do not contrast the ones by Bocci and Avesani (2011). The present experiment shows that the distinction in the prominence degree between broad focus and post-focal position is present in the case in which both words are in the same position in the metrical structure but does not enable conclusions to be drawn regarding the case in which words are in different positions. Indeed, the current experiment has not compared the acoustic characteristics connected to prominence of words in the post-focal region occurring in a structurally prominent position with words in the post-focal region occurring in a non-structurally prominent position. This comparison has been investigated by Bocci and Avesani (2011) for the variety of Italian spoken in Tuscany. The researchers, through the use of right dislocation, created two comparable conditions in which the target word occurred: in a structurally prominent position (created by the fact that it occurred before a right dislocation) and in a non-structurally prominent position because the sentence did not present the right dislocation. The experiment reported in the present chapter suggests that it would be interesting to use PEM, which has not been used before, to investigate the same comparison to further our understanding of the acoustic implementations of different degrees of structural prominence. The lack of this comparison in the current experiment was principally due to the fact that this experiment was primarily conducted to create stimuli for the perception experiment (Chapter 6). Moreover, this was a first attempt to investigate the post-focal position in the variety of Italian spoken in Udine using the parameters connected to periodic energy. Therefore, a simpler design was considered to help to explore the issue at hand more easily. In addition, in designing the experiment, problems concerning the task were also considered: the presence of another condition may have fatigued the speakers and let them more easily produce stereotypical contours of read speech and more easily depart from the intention of conveying meaning. The investigation of this further comparison in the degrees of prominence might be fruitful for future research.

An additional direction for future studies in the analysis of prominence relations in (varieties of) Italian would be to collect a large amount of data from spontaneous or semi-spontaneous speech. Indeed, the present study is concerned only with read speech, which on the one hand presents the advantage to compare conditions in a more controlled setting, but on the other hand makes it difficult to generalise the results to spontaneous speech. The results taken from the comparison between the two datasets (long and short) and shown in the following subsection (5.3.9) provide further evidence to the already accumulated evidence that production studies on intonation are not only sensitive to individual strategies employed by participants, but that they are also very sensitive to task requirements (Niebuhr & Michaud, 2015).

5.3.9 Comparison between the two datasets

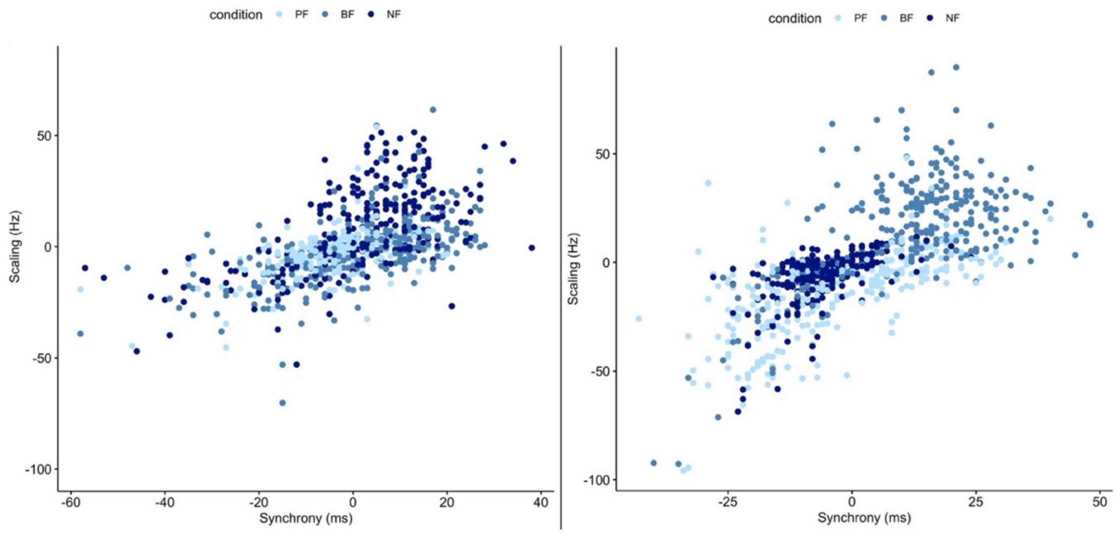
The comparison between the two datasets is illustrated by Figure 47 with respect to synchrony, scaling and PEM and by Figure 48 for synchrony and scaling. Figure 47 provides a useful representation of the interplay of the three parameters in conveying prominence and allows us to visualise their distribution in the space for the two sets. Figure 48 allows us to focus on the modulation of F0.

In Figure 47, the higher degree of aggregation in the same region of the cubes shown in the first dataset (long sentences) appears to confirm that participants in this experimental session distinguished prominence patterns to a lesser extent compared to speakers in the second dataset (short sentences). Results of the analysis of variance supported this observation, showing that prominence relations measured through PEM revealed to be less varied in the second dataset [$F(1) = 246.87$, $p < 0.0001$], indicating a better control in signalling prominence patterns. In addition, the analysis showed that synchrony has a higher variation in the first dataset (long sentences) compared to the second dataset [$F(1) = 68.45$, $p < 0.0001$], showing that speakers in the second set are more coherent in the F0 modulation within syllables. By contrast, scaling values presented a higher variance in the second dataset (short sentences) compared to the first dataset [$F(1) = 80.69$, $p < 0.0001$], indicating an overall reduced range in F0 for the first dataset. This suggests a reduction of expressiveness in the first dataset, due to the higher artificial design and the increased cognitive load it generates.



a. b.

Figure 47. 3D plot showing the three parameters of synchrony, scaling and PEM for first dataset (a.) and second dataset (b.).



a. b.

Figure 48. Scatter plot showing the two parameters of synchrony and scaling first dataset (a) and second dataset (b).

These results suggest that when the setting is less natural, because the task requires participants to produce a sentence they would not (or extremely rarely) produce in spontaneous speech, the inter-individual and intra-individual variability in the prosodic marking of the focal structures increases. Indeed, participants show a higher number of occurrences in which their planning of the focal structure was consistent with the expectations given the elicitation questions, suggesting that part of the variability was only due to the design of the experiment and to task requirements (Niebuhr & Michaud, 2015). Thus, part of the inter-individual variability can be attributed to the fact that some participants could better adapt to an unnatural scenario than other participants. Possibly, the planning of the focal structure was to some extent made more difficult by the uncommon structuring of the answer and by its length. This difficulty might have consequences for the conclusion drawn from the analysis of the utterances. This point is particularly important in the research on post-focal position, since many times the stimuli created to make speakers realise words post-focally which have to occur in prominent positions, are rather complex and can make it difficult for speakers to engage in emulating a real conversation setting.

To conclude, this experiment has shown that, despite the variability driven by the tasks,speakers of the Udine variety of Italian employ changes in F0, duration and energy to signal differences in information structure. In the next chapter (Chapter 6), an experiment will be presented which aims to understand whether the changes in production here described are perceived by listeners.

Chapter 6

Rating study: perception of prominence in the post-focal region in Udine and Bari Italian

6.1 Introduction

The previous study (production study) has shown that speakers of the variety of Italian spoken in Udine mark different focal structures of utterances with different prominence relations. The study reported in this chapter investigates whether prominence relations are perceived by listeners of the same variety. In particular, the degree of prominence of the post-focal position is of interest. The prominence perceived by native speakers of Udine, is then compared with the prominence perceived by native speakers of Bari, rating their native variety. In addition, the perception of prominence relations in these two varieties of Italian (Udine and Bari) by native speakers of German, learning Italian as their foreign language, is investigated.

The rationale for conducting this experiment is provided in Chapter 2 and Chapter 3, where the discussion on the relation between acoustic characteristics of prosodic prominence and their perception is outlined. As explained in these chapters, prominent elements are perceived to "stand out" in comparison to their neighbouring elements (Terken, 1991). The acoustic characteristics that make them stand out are not only F0 movements but also duration and intensity and the interaction between them (Arnold et al., 2013; Baumann & Winter, 2018; Campbell, 1995; D'Imperio, 1998; Mo, 2008; Turk & Sawusch, 1996; Wagner et al., 2015).

Perception of prominence is also expectation-based and the building up of expectations depends on several factor (see Chapter 3), such as the knowledge of the language and the probabilistic distribution of accents. In order to investigate expectation-based inferences, i.e. the role of a speaker's native language and the probabilistic mapping of prominence, the present study compares prominence ratings of the two groups of native speakers of Italian between each other and to those of the learners. Expectations on the degree of prominence of the post-focal region are forecasted to differ among groups, as a result of the inferences built on the probabilistic distribution of prominences within the utterance characteristic of the native language (or variety). Indeed, the two varieties of Italian differ in the probabilistic distribution of prominences, given the presence of acoustic cues to prominence in the post-focal region of question in the Bari variety. German patterns with the Udine variety concerning the distribution of prominences in the post-focal region. In addition, it has a different knowledge of the phonological system of Italian. This last characteristic of learners might cause differences also with the Udine variety, despite the same distribution of prominences (in both flat and low post-focal contour, irrespective of sentence modality).

In both varieties of Italian, the post-focal region of the utterances chosen is realised with a flat and low contour (see 2.3.3). This characteristic enables us to answer two questions regarding the role of non-pitch driven cues to prominence and the role of expectations in the rating of prominence. As mentioned above, these two varieties have been chosen because, while presenting a similar distribution of acoustic cues to prominence in the post-focal position of long statements, they differ in questions. The Bari variety (henceforth, BI) has the property of realising questions with a post-focal pitch accent (see Chapter 2), while in the Udine variety (henceforth, UI), the contour of questions is realised similarly to German. In German and in the Udine variety, a post-focal rising-falling pitch movement associated with the stressed syllable and signalling modality is not present: modality is signalled by a final rise associated with the phrase edge (see Chapter 2.3.3). The hypothesis is that participants of the BI group, because they are exposed to post-focal cues to prominence, have a higher expectation to encounter prominence in the post-focal region compared to participants of the UI group. As a consequence, the group of listeners of BI would rate post-focal material as prominent even when realised with a high degree of attenuation.

The decision to have statements and not questions as stimuli, the former realised in BI with a post-focal flat and low contour, the latter with a movement in pitch associated with a stressed syllable, was guided by the desire to investigate not only bottom-up perception of the acoustic cues to prominence, but also the role of expectations. Accordingly, the stimuli in post-focal position needed to have the least number of bottom-up cues.

To summarize, the goal of the current experiment was to understand to what extent the post-focal position is perceived as attenuated by the three groups and whether the different distribution of accents in BI impacts perception. In particular, the knowledge of the possible presence of a pitch accent in the post-focal position is more available to listeners coming from Bari. By contrast, it is not available to German learners, for whom the variety is likely unknown, as well as for UI participants, who are not commonly exposed to it. Therefore, BI participants are predicted to more highly expect prominences in the post-focal position compared to the other two groups. This is expected to reflect in BI participants rating post-focal position as more prominent than UI and German participants.

In addition, in German, the attenuation of post-focal material is usually reported to be realised in a more clear-cut fashion than in Italian. This might also play a role in the top-down inferences on prominence perception, leading not only to differ from the BI group, but also from the UI group. Germans might perceive post-focal position with a lower degree of prominence compared to UI participants. In the present experiment predictions concerning all groups will be investigated through the use of a rating task (see 6.4).

6.2 Method

6.2.1 Participants

The prominence ratings of three sets of listeners were compared with each other: 17 native speakers of the Udine variety of Italian (10 women, 7 men; age: mean = 27.52, *SD* = 5.16), 16 native speakers of the Bari variety of Italian (13 women and 3 men; age: mean = 26.47, *SD* = 3.93), and 18 German intermediate and upper-intermediate learners of Italian (14 women and 4 men; age: mean = 24.72, *SD* = 2.65). The learners rated the two varieties in separate sessions, with an interval of at least two weeks between each session. The order of presentation of the two varieties was controlled, so that the number of participants rating the Bari variety first and the number of participants rating the Udine variety first was balanced. The choice of two different levels of proficiency was made in order to test whether this variable affected the prominence ratings. German native speakers were selected from language classes corresponding to the two levels of proficiency. In addition, participants self-assessed to which level of proficiency they belonged. The proficiency level did not have an effect on the prominence ratings in any of the two varieties rated, as revealed by the mixed analysis. Consequently, the two groups of learners were not considered as separate groups[17].

[17] The mixed effect analysis that was run considered prominence ratings as dependent variable and PROFICIENCY, CONDITION, WORD POSITION and VARIETY as fixed effects. Random intercepts

Listeners of all the above-mentioned sets (Udine, Bari and German) were all either university students or recently graduated students. Moreover, among all sets, none of the participants had a background in phonetics or prosody. All participants gave written informed consent.

6.2.2 Material

The stimuli consisted of 30 utterances (10 utterances per 3 conditions) which were presented to the listeners without context. The stimuli for UI were selected among the utterances produced in the production experiment presented in the previous chapter (Chapter 5). For the BI group, stimuli were selected among those produced by 4 speakers and presented in 6.2.5.3. Utterances used in the perception experiment were produced in three different focal conditions in which the target word occurs as broad focus (BF), contrastive narrow focus (NF) or as part of the background, realised after a contrastive narrow focus on the verb preceding it, therefore in post-focal position (PF). An example of the stimuli in the three renditions is reported in (22)-(24), where the narrow focus is indicated in capital letters and the target word in bold (see Appendix A3 for the complete list of stimuli). Listeners were asked to evaluate the three content words (underlined in the example, henceforth critical words). The target word (a noun) was the direct object of the verb preceding it and was always followed by a prepositional phrase. As discussed in 5.2.1, the target was realised as the metrical head of its intermediate phrase and would, therefore, occupy a prominent position. This might have an influence on the level of prominence perceived.

(22) TARGET in broad focus

[Bisogna <u>pesare</u>$_{verb}$ la <u>farina</u>$_{TARGET}$ con la <u>bilancia</u>$_{noun\ in\ PP}$]$_{FOCUS}$

3 SG .INF DET N PREP DET N

should weigh the flour with the scales

One should weigh$_{verb}$ the flour$_{TARGET}$ with the scales$_{noun\ in\ PP}$

---

and slopes for SUBJECT and ITEM were entered into the model as random effects. As further evidence to the fact that proficiency does not influence the ratings of prominence in a design as the one presented here, a larger group of German learners of Italian was subjected to the investigation. This set was comprised of 155 German learners of Italian who were not selected for age and education (51 women and 104 men; age: mean 44.75, *SD* 17.24) and had four different levels of proficiency: beginner, intermediate, upper-intermediate and advanced. Also in this case, the mixed analysis preformed on the data revealed that proficiency did not play a role in the ratings.

(23) TARGET in narrow contrastive focus

Bisogna pesare$_{verb}$ [LA **FARINA**$_{TARGET}$]$_{FOCUS}$ con la bilancia$_{noun\ in\ PP}$

(24) TARGET in post-focal position

Bisogna [PESARE$_{verb}$]$_{FOCUS}$ la farina$_{TARGET}$ con la bilancia$_{noun\ in\ PP}$

6.2.3 Statistical analysis of the acoustic stimuli

Similar to the analysis conducted in the production study of Chapter 6, linear mixed effects models of the relationship between focal condition, critical word position and prominence were carried out. First, F0 dynamic was considered: mixed effect models were performed, with POSITION of the word (verb, target and noun in PP) and CONDITION (broad-focus, narrow focus, post-focal position) as independent variables, and values of synchrony and scaling for the stressed syllables as dependent variables, considered separately. Random intercepts for ITEM were considered as random effects. Secondly, relative periodic energy mass (PEM) was considered and regarded as the dependent variable of mixed effects analyses. For these analyses, the factors CONDITION and POSITION were considered as independent variables. Random intercepts for ITEM and random intercepts and random slopes for SUBJECT were considered as random effects.

6.2.4 Acoustic features of the stimuli: Udine variety

The stimuli presented to the native speakers of the Udine variety were selected among those produced by speakers in the second set (short sentences) of the production experiment in 5.3. Only one speaker was chosen (speaker F01), in order to guarantee homogeneity in the encoding of prominence relations. This one speaker was selected on the basis of the naturalness of the prosodic realisation and fluency and, again based on these criteria, only 10 utterances from this speaker were chosen. The characteristics of the utterances chosen as stimuli are presented here.

As shown in Figure 49, in most of the cases the target had a slightly rising contour in BF (H*), whereas in the remaining cases the pitch was falling (H+L*). The same accents were used (in different proportion) for the verb occurring in BF and the verb preceding the focus (in NF). Words in focus (verb in PF and target in NF) were always realised with a L+H* rising accent. The target in post-focal position was by contrast realised always without pitch accent. The final noun in BF was always realised with a falling movement, whereas when occurring in post-focal position (NF and PF conditions) the realisation was twofold: in PF always without an accent, in NF either with a falling pitch (H+L*) or without an accent. As already reported in Chapter 5, the target in BF always preceded a high intermediate phrase boundary tone (H-).

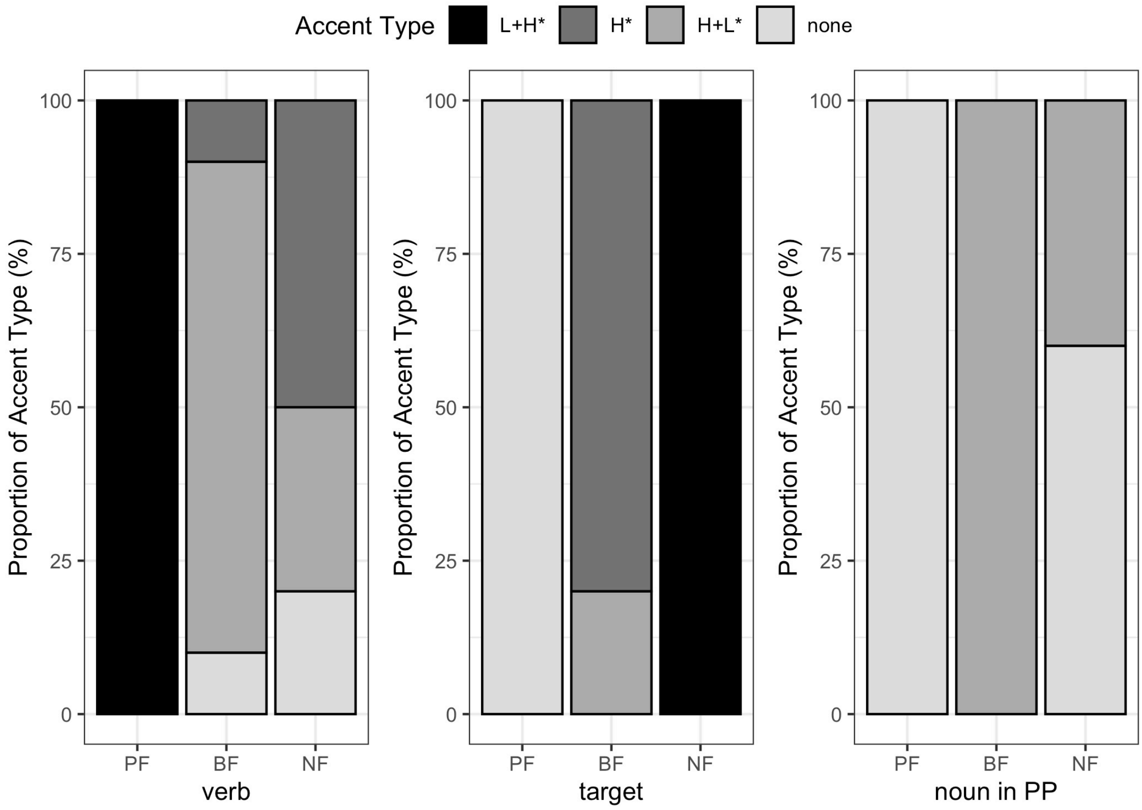


Figure 49. Accents type proportion of the 10 utterances (from speaker F01) selected as stimuli for the perception experiment.

Figure 50 (50B., 50D., 50F.) shows PEM for the stressed syllable of the three critical words. Table 26 and Table 27 display the results of the statistical analyses for the stressed syllables of all the critical words in all the conditions. The verb in BF presents a distribution of values similar to the target (difference not significant; see Table 26) and higher values compared to the noun in PP. The target in NF exhibits a pitch accent on the target and almost no movement on the verb. However, the difference between the PEM of the stressed syllable of the target and the verb does not register significant differences (see Table 26). By contrast, there is a difference in the noun in PP, with the verb and the target showing higher values, according to the occurrence in the PP after the focus, i.e. in a post-focal position (see Table 26). In PF, the verb is the only word that features a pitch accent and correspondingly exhibits high values of PEM. The following flat-contoured target word and the noun in PF exhibit low values of PEM.

The comparison between the stressed syllable of the targets occurring in different conditions reveals a lack of significance in the difference of PEM values between BF and PF (see Table 27). On the contrary, values of PEM for the stressed syllables of the target occurring in NF are significantly higher than the ones of the target occurring in BF and PF. Considering the stressed syllables of the verb, the highest values are exhibited in PF compared to BF and NF. For the noun in PP, the stressed syllable has higher values of PEM when occurring in PF and in NF compared to BF.

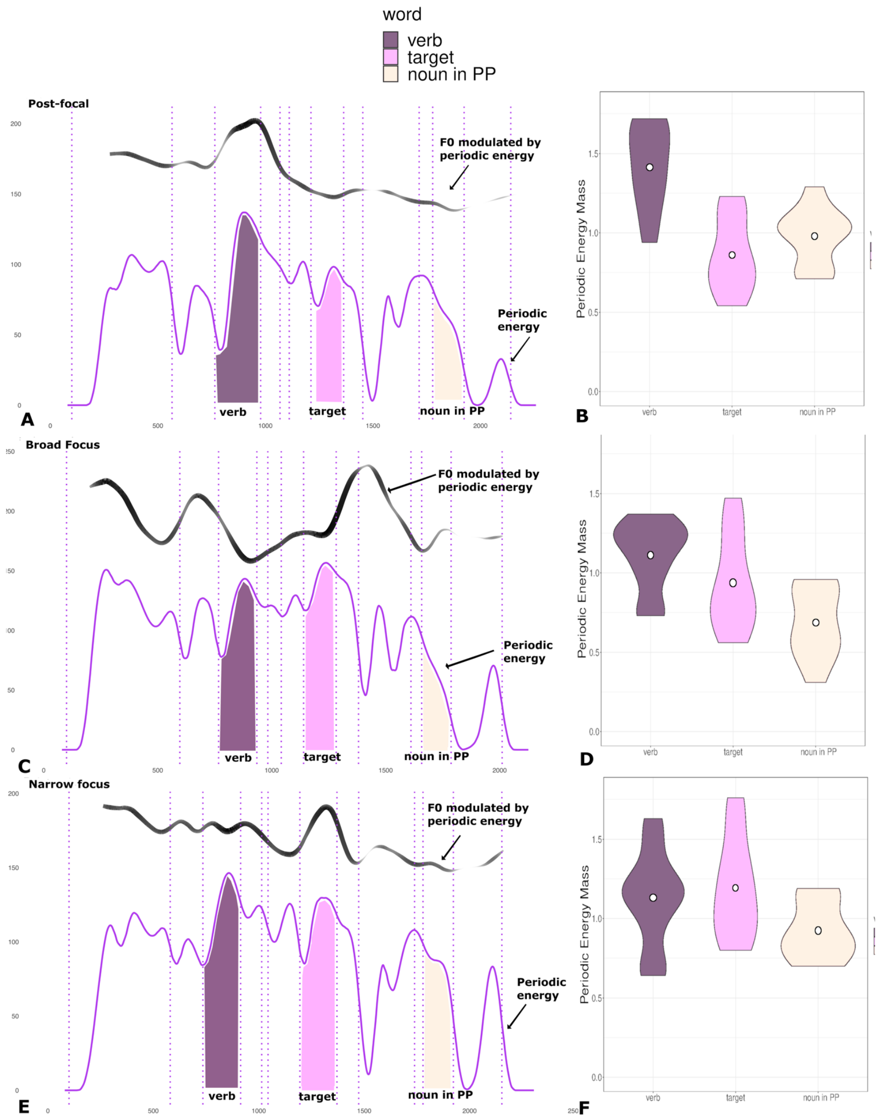


Figure 50. A., C. and E. (left panel) example of the sentence *Bisogna pesare la farina con la bilancia* (One should weigh the flour with the scale): upper curve F0 modulated by periodic energy, lower curve periodic energy. Y-axis show values in Hertz (Hz), x-axis show time in milliseconds (ms). Vertical lines show boundaries of the stressed syllable for every critical word. Parts of the periodic energy filled with colour correspond to the area under the periodic energy curve relative to each stressed syllable of the critical words. This is the value that corresponds to the Periodic Energy Mass. B., D. and F. (right panel) show Periodic Energy Mass for the critical words (verb, target noun and noun in PP) of every stimulus. Y-axis show relative values of Periodic Energy Mass, x-axis show the independent variable of the word. The figure shows the three prosodic conditions: post-focal (A. and B.), broad focus (C. and D.) and narrow focus (E. and F.). White dots indicate mean values.

	BF	**NF**	**PF**
verb	$\beta = 0.18 \pm 0.11$, $p = 0.13$	$\beta = -0.06 \pm 0.12$, $p = 0.62$	$\boldsymbol{\beta = 0.6 \pm 0.11}$, $\boldsymbol{p < 0.0001}$
noun in PP	$\boldsymbol{\beta = -0.25 \pm 0.11}$, $\boldsymbol{p = 0.04}$	$\boldsymbol{\beta = -0.27 \pm 0.12}$, $\boldsymbol{p = 0.04}$	$\beta = 0.12 \pm 0.11$, $p = 0.28$

Table 26. Results of the mixed model analyses of the differences between PEM of the stressed syllables in each word position within each condition, from left to right: broad focus, narrow focus and post-focal position. Significant results are displayed in bold.

	verb	target	noun in PP
PF	**β = 0.34 ± 0.06, p < 0.0001**	β = -0.07 ± 0.09, p = 0.38	**β = 0.29 ± 0.52, p < 0.0001**
NF	β = 0.02 ± 0.06, p = 0.77	**β = 0.26 ± 0.09, p = 0.01**	**β = 0.24 ± 0.52, p < 0.001**

Table 27. Results of the mixed model analyses of the differences between PEM of the stressed syllables in each condition within word position, from left to right: verb, target and noun in PP. Significant results are displayed in bold.

In the domain of the whole word (Figure 51), the PEM of the target in broad focus is not significantly different from the PEM of the target in narrow focus and in post-focal position. The same holds true for the verb and the noun in PP. In broad focus, the PEM of the target is not significantly higher than that of the verb and is significantly higher than that of the noun in PP. The PEM of the target in narrow focus is significantly higher than that of the verb and of the noun in PP. In post-focal position, the PEM of the target is significantly lower than that of the verb and significantly higher than that of the noun in PP. Results are reported in Table 28 and 29.

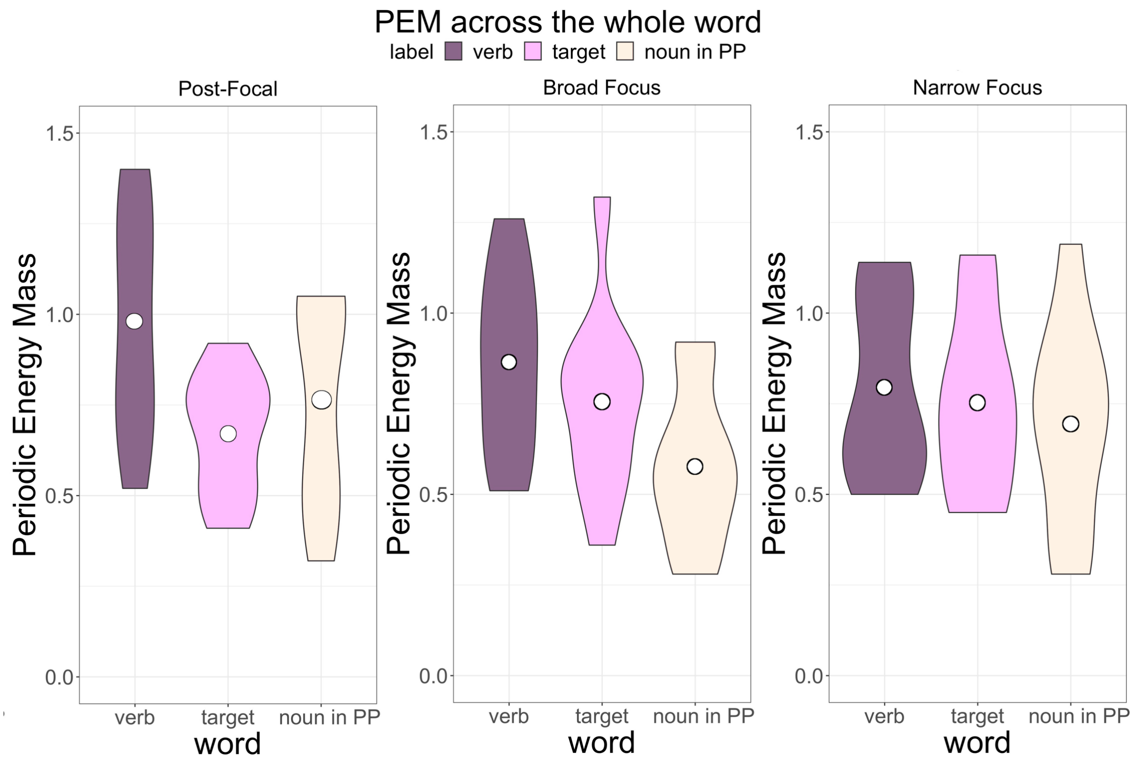


Figure 51. Relative PEM for the entire window of the words. Left panel shows values of PEM for verb, target and noun in PP in post-focal condition (verb in focus). Panel in the middle, shows values of PEM for verb, target and noun in PP for broad focus condition. Right panel shows values of PEM for verb, target and noun in PP for narrow focus condition (target in focus). Black points on the violin plots indicate mean values.

	verb	target	noun in PP
PF	β = 0.12 ± 0.06, p = 0.09	β = -0.09 ± 0.05, p = 0.11	**β = 0.18 ± 0.06, p = 0.005**
NF	β = - 0.07 ± 0.06, p = 0.29	β = -0.03 ± 0.05, p = 0.79	β = 0.12 ± 0.06, p = 0.06

Table 28. Results of the mixed model analyses of the differences between the entire word' values in each condition in each word position. Values are relative to the intercept (BF), corresponding to the relative critical word (verb, target and noun in PP) in broad focus. Significant results are displayed in bold.

	BF	NF	PF
verb	β = 0.11 ± 0.1, p = 0.33	β = 0.04 ± 0.11, p = 0.73	**β = 0.31 ±0.11, p = 0.01**
noun in PP	β = - 0.17 ± 0.1, p = 0.11	β = -0.06 ± 0.12, p = 0.62	β = 0.09 ± 0.11, p = 0.4

Table 29. Results of the mixed model analyses of the differences between each word position within each condition, from left to right: broad focus, narrow focus and post-focal position. Results refer to the domain of the whole word. Significant results are displayed in bold.

Figure 52 shows the values of scaling and synchrony for each syllable of the stimuli that present the lexical stress in penultimate position (indicated in the figure by the label *stress*). The interested reader can retrieve the figures in the remaining combinations at https://osf.io/5m8hw/. The values of scaling and synchrony confirm the intonation analysis that is reported in Figure 49. Negative values of both measures indicate that the verb in BF is mostly realised with a falling pitch accent on the stressed syllable. Mostly positive, but rather low values of synchrony for the stressed syllable of the target indicate that the majority of realisations comprise a slightly rising pitch (also indicated by the values of synchrony around zero), while some observations are realised with a slight fall (indicated by some negative values). A steep rise is realised at the end of the target word, indicating the presence of a phrase boundary. The noun in PP in this position is realised with a falling accent on the stressed syllable and a steep rise on the final syllable.

Negative values of synchrony, not very distant from zero, and values of scaling near zero indicate that the verb in the NF condition is mostly realised either with a slight fall or a slight rise, but does not show accents expanding in a high range. In some cases, it appears to show nearly absent movement. Very high and positive values of synchrony and scaling for the target in NF indicate that this word is realised with a rising pitch characterized by a great excursion. Following this great excursion, the very low values of scaling and synchrony that are around zero, show the (near) absence of movement for the first two syllables, and a steep rise on the final one.

The same pattern of values shown in the target in NF is observed also for the verb in PF. The critical words following the verb are also here realised with values of scaling and synchrony around zero, indicating a flat and low contour until the last syllable, which is also here realised with a final rather steep rise.

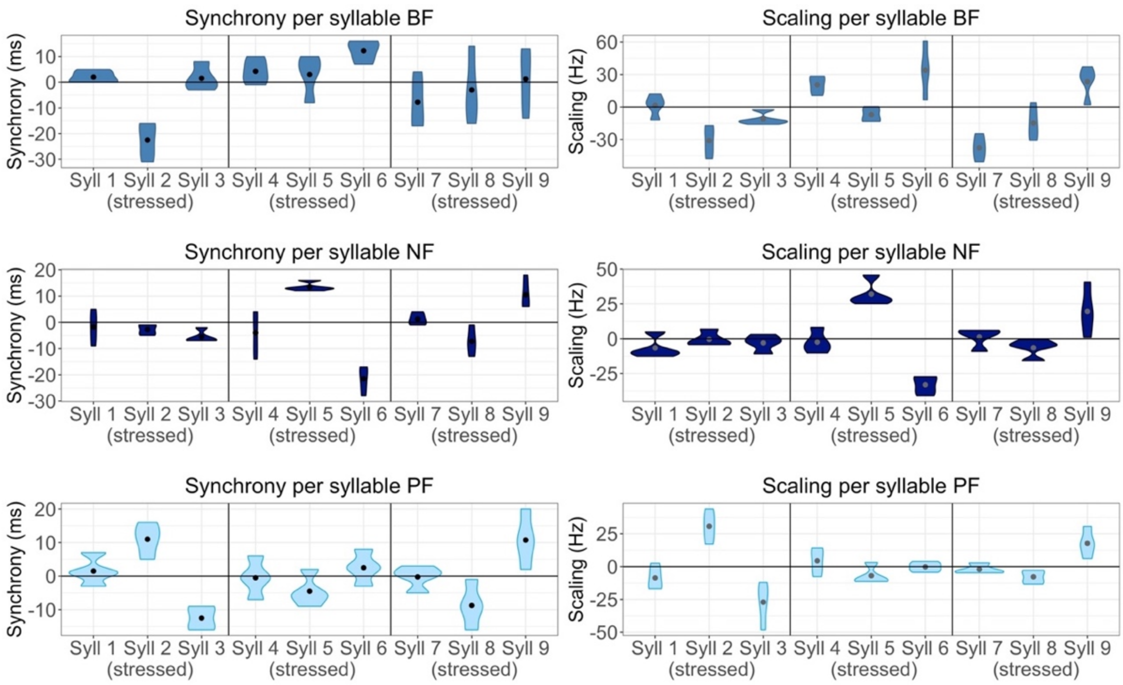


Figure 52. Values of scaling (on the left) and synchrony (on the right) for each syllable of the utterances in BF (first row), NF (second row) and PF (third row). The graph shows results of stimuli where both verbs and nouns in PP have penultimate stress (target nouns always have lexical stress in penultimate position). Dotted vertical lines mark each syllable, solid vertical lines mark critical words. Grey points indicate mean values.

In Table 30 means and standard deviations of scaling and synchrony for the stressed syllables of each prosodic condition and word position are reported. These measures are relative to all the stimuli. In Table 31 and 32, results of the statistical analysis are showed. As for scaling and synchrony, for the stressed syllable of the verb the models confirmed the difference between post-focal and broad focus, showing a high increase of values for the former condition. The stressed syllable in the verb showed a significant increase for NF over BF, attesting the presence of values more near zero for the former condition. As for the target, synchrony and scaling were proved to be increased for PF over BF, confirming values closer to zero for PF (see overall means in Table 30). The increase of values in NF in comparison to BF was also confirmed, indicating more rising contours for the former condition.

In the broad focus condition, the stressed syllables of the verb and the noun in PP have higher values of synchrony than the target. In the narrow focus condition, values of scaling and synchrony for the target are significantly higher than in the other two word positions. In the post-focal condition, the stressed syllable of the verb showed significantly higher values of scaling and synchrony than that of the stressed syllable of the target. Finally, the difference between synchrony of target and synchrony of noun in PP was significant, indicating more movement within the stressed syllable of the last constituent.

condition	verb	target	noun in PP
BF	Scaling: -12.31 (17.88) Synchrony: -10.56 (18.15)	Scaling: - 16.53 (20.33) Synchrony: - 5.06 (15.88)	Scaling: -8.46 (18.86) Synchrony: -14.81 (15.88)
NF	Scaling: -8.46 (18.86) Synchrony: -4.8 (15.88)	Scaling: 17.16 (22.60) Synchrony: 12.6 (16.15)	Scaling: -6.25 (13.58) Synchrony: -0.67 (17.08)
PF	Scaling: 17.85 (24.53) Synchrony: 9.96 (20.53)	Scaling: -7.13 (11.35) Synchrony: -7.26 (7.6)	Scaling: -5 (11.64) Synchrony: 1.53 (27.47)

Table 30. Mean and standard deviation of scaling and synchrony, in brackets, for stress syllables of the critical words for all the stimuli.

	verb	**target**	**noun in PP**
PF	**Scal: $\beta = 30.17 \pm 2.47$, $p < 0.0001$** **Synch: $\beta = 20.5 \pm 2.18$, $p < 0.0001$**	**Scal: $\beta = 9.59 \pm 2.22$, $p < 0.0001$** Synch: $\beta = -2.11 \pm 1.65$, $p = 0.2$	**Scal: $\beta = 25.74 \pm 2.34$, $p < 0.0001$** Synch: $\beta = 16.12 \pm 2.28$, $p < 0.001$
NF	Scal: $\beta = 3.85 \pm 2.46$, $p = 0.12$ **Synch: $\beta = 5.75 \pm 2.18$, $p < 0.01$**	**Scal: $\beta = 33.66 \pm 2.19$, $p < 0.0001$** **Synch: $\beta = 17.6 \pm 1.63$, $p < 0.0001$**	**Scal: $\beta = 24.47 \pm 2.32$, $p < 0.0001$** **Synch: $\beta = 13.93 \pm 2.6$, $p < 0.0001$**

Table 31. Results of the mixed model analyses on the values of scaling (scal) and synchrony (synch) of the differences between the stressed syllables' values in each condition in each word position. Values are relative to the intercept, corresponding to the relative critical word (verb, target and noun in PP) in broad focus. Significant results are displayed in bold.

	BF	NF	PF
verb	Scaling: β = 2.96, ± 2, p = 0.14 **Synch: β = -6.38 ± 1.67, p = 0.0001**	**Scaling: β = - 25.63, ± 2.24, p < 0.0001** **Synch: β = -17.43 ± 1.99, p < 0.0001**	**Scaling: β = 24.91, ± 2.05, p < 0.0001** **Synch: β = 17.16 ± 1.96, p < 0.0001**
noun in PP	Scaling: β = -16.56 ± 2, p < 0.0001 **Synch: β = 13.34 ± 1.67, p < 0.0001**	**Scaling: β = - 23.44, ± 2.33, p < 0.0001** **Synch: β = -13.28 ± 1.99, p < 0.0001**	Scaling: β = 2.06 ± 2.06, p = 0.32 **Synch: β = 8.73 ± 1.96, p < 0.0001**

Table 32. Results of the mixed model analyses of the differences between each word position in (from the left) broad focus (BF), narrow focus (NF) post-focal position (PF). Values are relative to the intercept, corresponding to the target in the relative condition. Significant results are displayed in bold.

6.2.5 Acoustic features of the stimuli: Bari variety

To see whether the distribution of prominence is different in another variety, a smaller corpus was collected from speakers of the variety spoken in Bari. The collection and the analysis of this corpus aimed at selecting stimuli for the perception experiment. In section 6.2.6.3 the corpus collected will be described.

6.2.5.1 Participants

For the present set of recordings, 4 female native speakers of the variety of Italian spoken in Bari were recorded. Participants' age ranged from 21 to 22 years old (mean age 21.5, *SD* 0.58). All speakers had been continuously exposed to the Bari variety of Italian, used it for everyday conversation and were all university students. None of them self-reported any speech or auditory impairment. In addition, none of the participants had a background in phonetics or prosody. All participants gave written informed consent.

6.2.5.2 Method

The method used to collect this data was the same used for the second set of recordings in the production study of Chapter 5 (5.3). Materials and procedure are described in the *Materials* and *Procedure* sections of the second set of recordings (5.3.1 and 5.3.3, respectively; the complete list of stimuli is presented in Appendix A2), the analyses conducted are described in the *Analysis* section of the first set of recordings (5.2.4).

6.2.5.3 Results of the whole set

As for the datasets in the production study in Chapter 5, utterances of the Bari dataset were intonationally analysed by two trained transcribers using Praat (Boersma & Weenink, 2020). Results of the pitch accent types, pooled over speakers are displayed in Figure 53. The figure shows that the degree of variability is not very high and that clear tendencies in the realisation of the focal conditions can be detected from the distribution of the accent types. BF is almost always realised with falling contours on the verb and the target, while in the noun in PP speakers either produce it with a falling contour or realise it without pitch movement. In PF and NF the focus exponents of the utterances are almost always marked with a rising-falling accent (H*+L, 96.25% of the cases for the verb in PF, 100% of the cases for target in NF), whereas the post-focal position is almost always realised without movement (100% of the cases for target in PF condition, 93.75% of the cases for noun in PP in NF condition, 98.75% of the cases for noun in PP in PF condition). Pre-focal elements (i.e., verb in NF) were realised without movement (44.3% of the cases), with H+L* (16.46% of the cases) accent or with H* accents (39.24% of the cases).

As expected from the low variability of the data pooled together, the production of individual speakers depicted in Figure 54 shows that all the speakers follow the general trend.

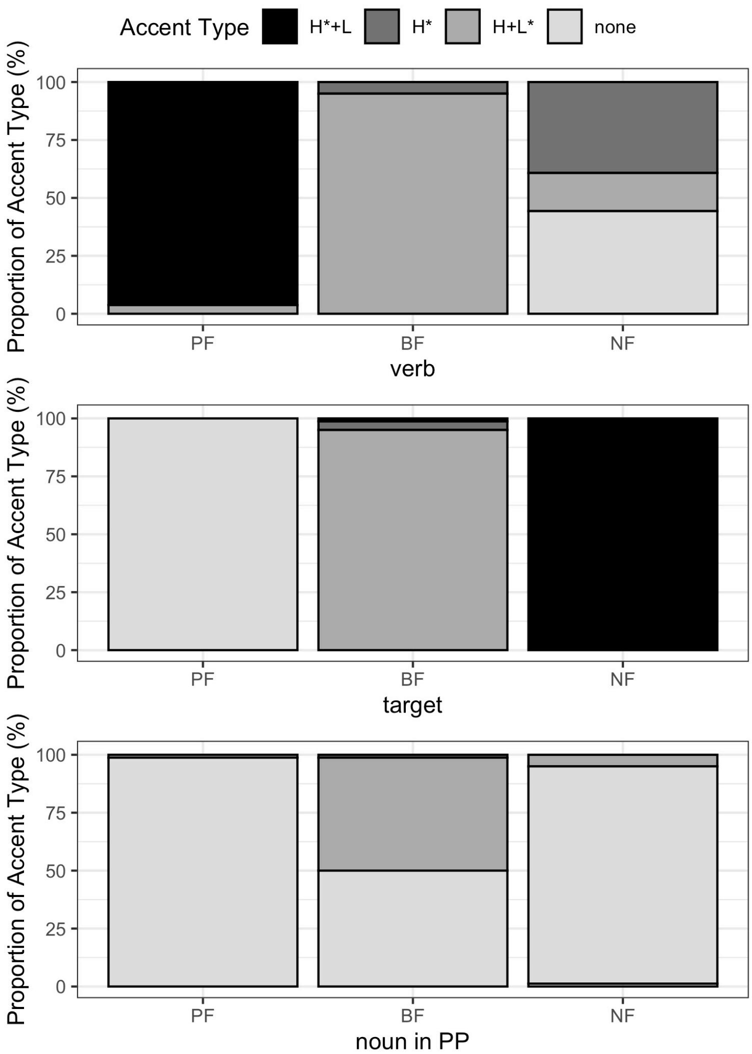


Figure 53. Accent types pooled across speakers. From the top results for the verb, the target and the noun in PP. From left to right, results for broad focus (BF), narrow focus (NF) and post-focal position (PF).

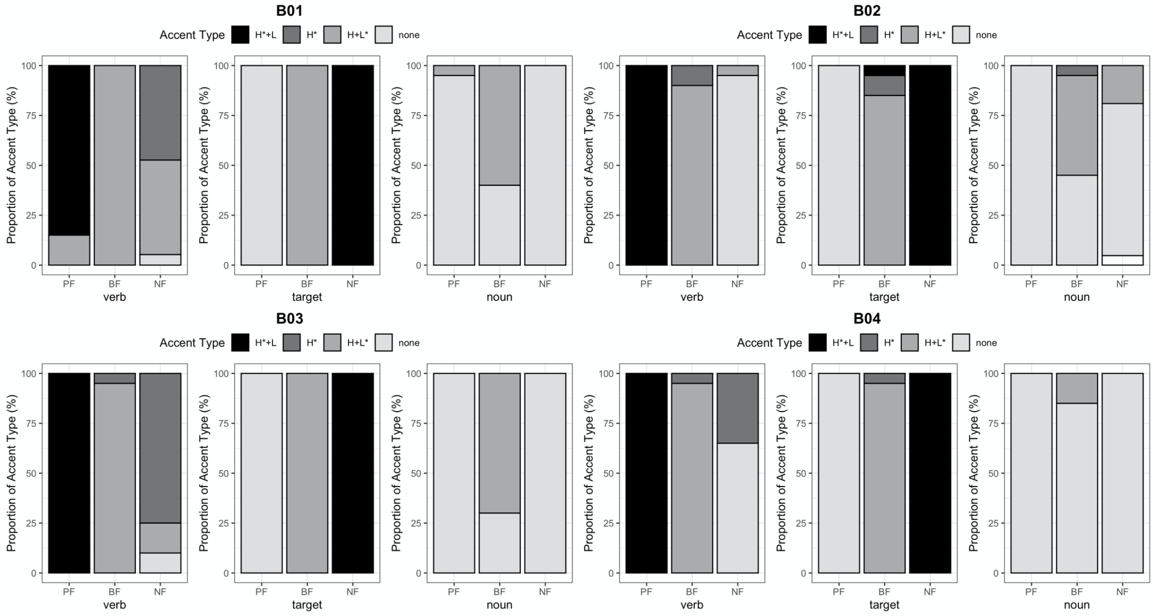


Figure 54. Distribution of pitch accent types across conditions and word position for each speaker separately.

The continuous parameters confirm what has been registered by the distribution of pitch accents. Figure 55 illustrates the measures of synchrony and scaling for the whole contour. It demonstrates that PF is realised with a rising-falling contour on the verb, NF is realised with a rising-falling contour on the target and BF with a falling contour on verb, target and noun in PP. The same tendency can be observed for every speaker (Figure 56 and Figure 57).

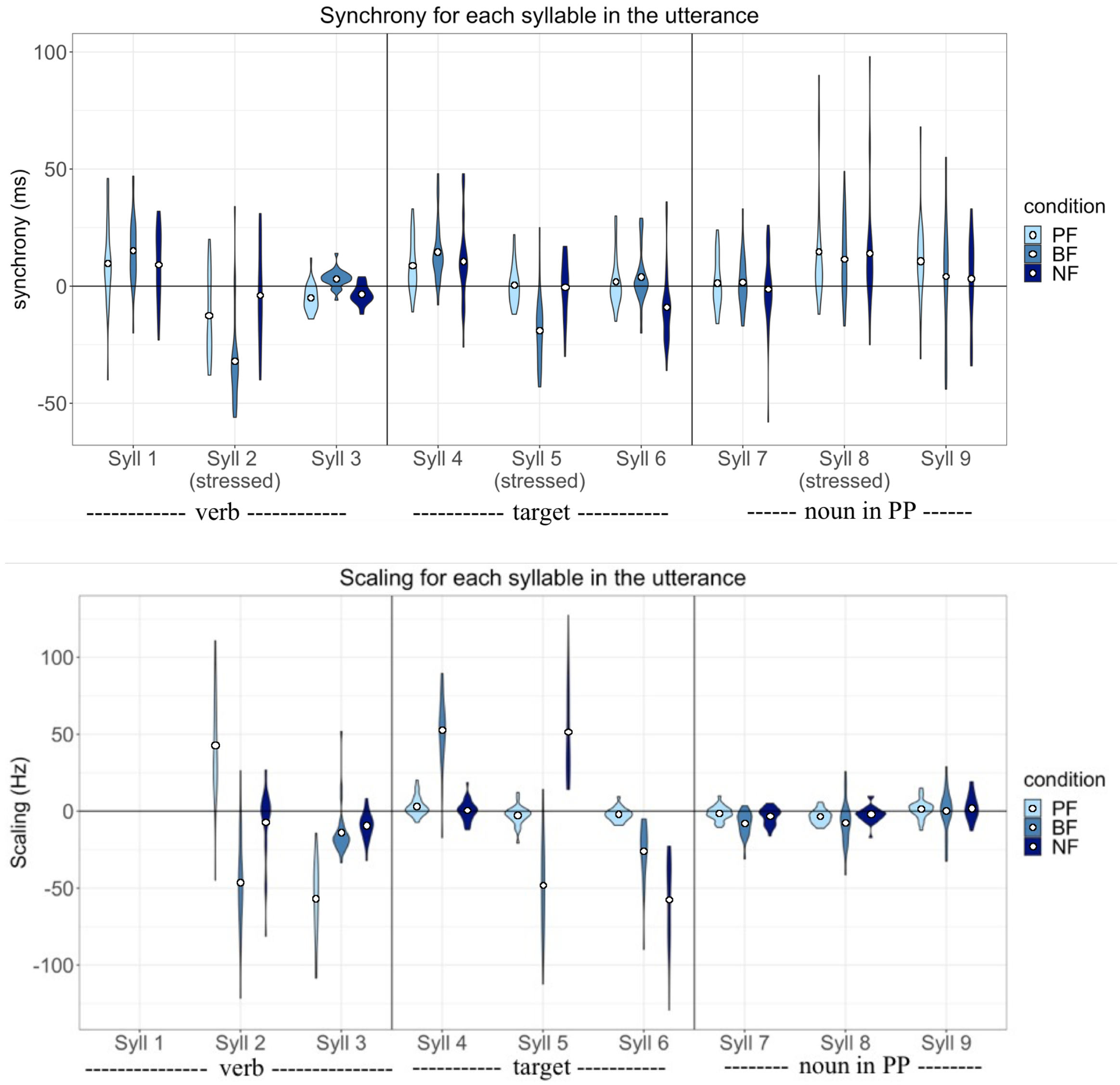


Figure 55. Synchrony and scaling of each syllable of the critical words. Stress on the penultimate syllable of each critical word. White dots on the violin plots indicate the mean value.

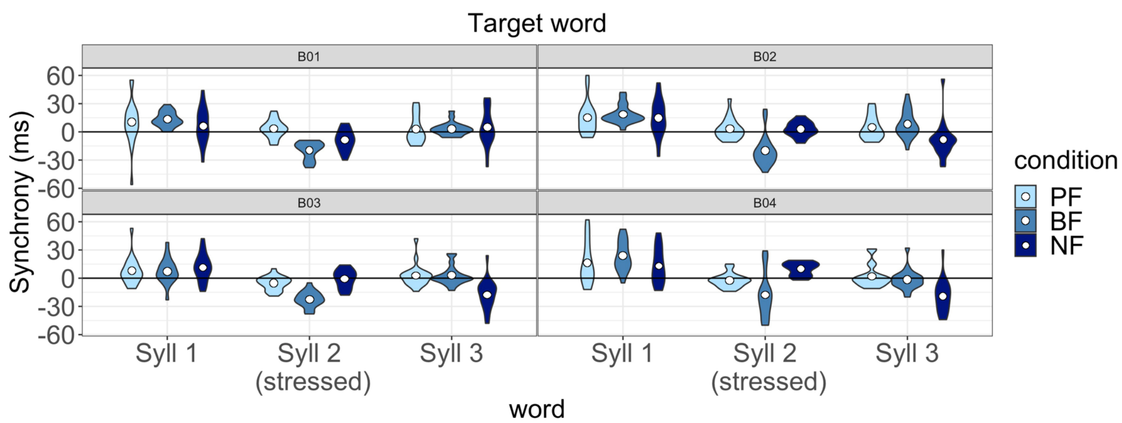


Figure 56. Synchrony of each syllable of the target word for each condition for each speaker (one speaker per panel). Stressed syllable is indicated by the label stressed in parentheses. White dots indicate mean values.

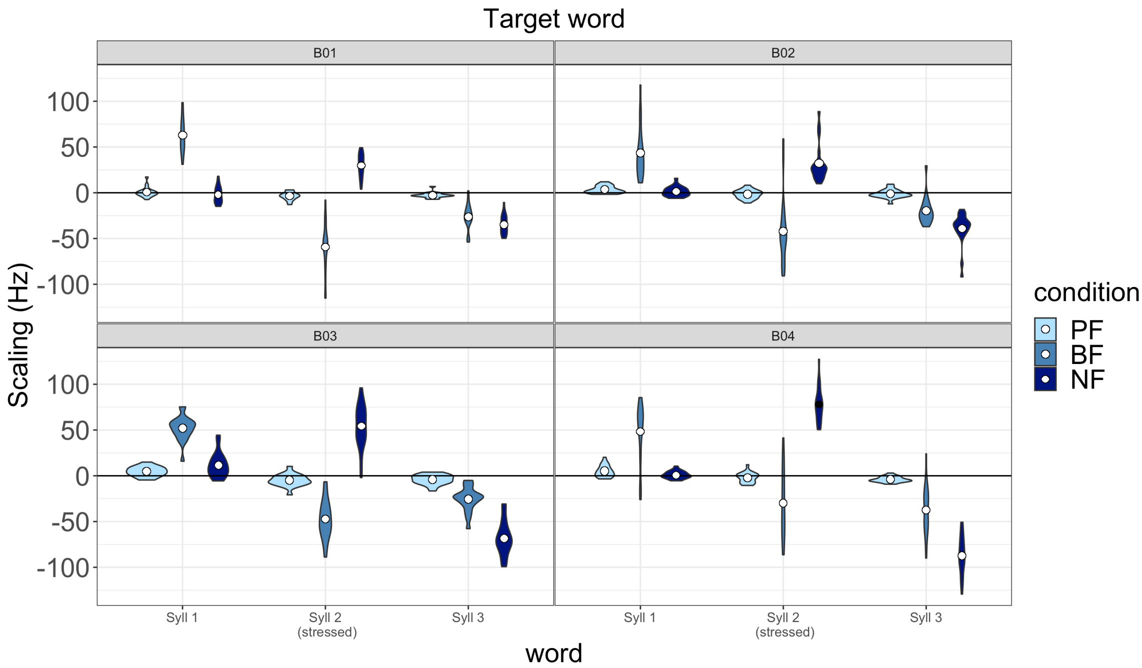


Figure 57. Scaling of each syllable of the target word for each condition for each speaker (one speaker per panel). Stressed syllable is indicated by the label stressed in parentheses. White dots indicate mean values.

Figure 58 shows values of PEM for the stressed syllables of each critical word. Results of the statistical model are reported in Table 33. The verb in PF has the highest values of PEM, whereas the noun in PP shows the same values for all three conditions. For the target, stressed syllables in NF have the highest values of PEM. In addition, results registered a difference between the PEM values of the target in BF and PF. The same pattern is shown for every speaker (Figure 59).

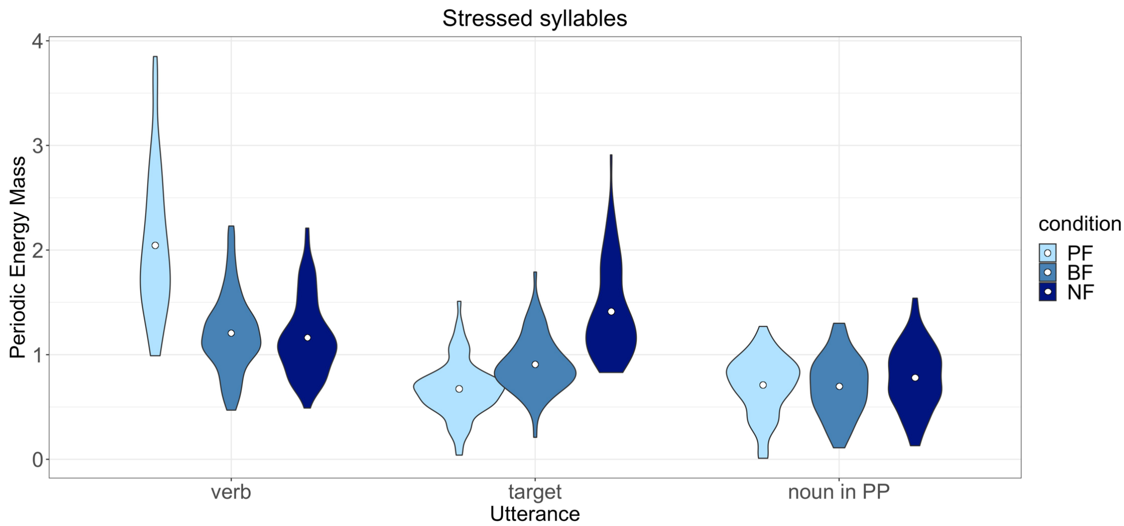


Figure 58. PEM values of the stressed syllables for each critical word in each condition. White dots indicate mean values.

	verb	target	noun in PP
PF	**β = 0.84 ± 0.23, p = 0.04**	**β = -0.23 ± 0.04, p < 0.01**	β = 0.01 ± 0.13, p = 0.93
NF	β = -0.04 ± 0.09, p = 0.68	β = -0.51 ± 0.13, p = 0.03	β = 0.08 ± 0.1, p = 0.45

Table 33. Results of the mixed model analyses of the differences between the stressed syllables' values in each condition in each word position. Values are relative to the intercept, corresponding to the relative critical word (verb, target and noun in PP) in broad focus. Significant results are displayed in bold.

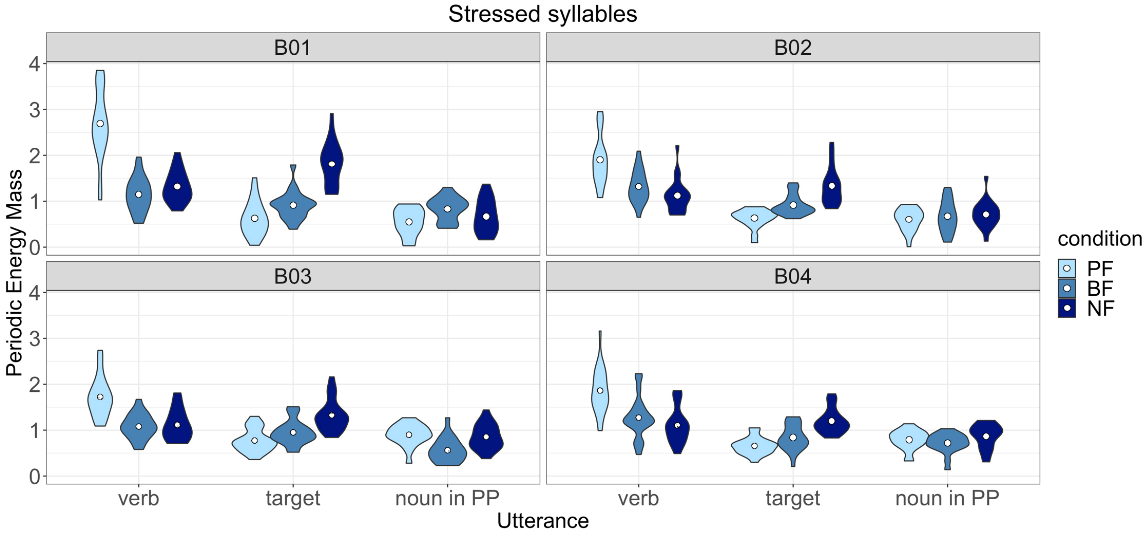


Figure 59. PEM of each syllable of the target word for each condition for each speaker (one speaker per panel). Stressed syllable is indicated by the label stressed in parentheses. White dots indicate mean values.

Figure 60 shows values of PEM for the target word. PEM of the first syllables of the target (Syll 1) is lower in PF compared to the other two conditions (β = -0.17 ± 0.05, p = 0.01). The same occurs in stressed syllables (Syll 2: β = -0.24 ± 0.05, p = 0.001), where in turn NF shows also higher values than BF (Syll 2: β = 0.51 ± 0.06, p < 0.0001). In the third syllable, again the only difference is between BF and PF, with the latter presenting lower values (Syll 3: β = -0.14 ± 0.05, p = 0.02).

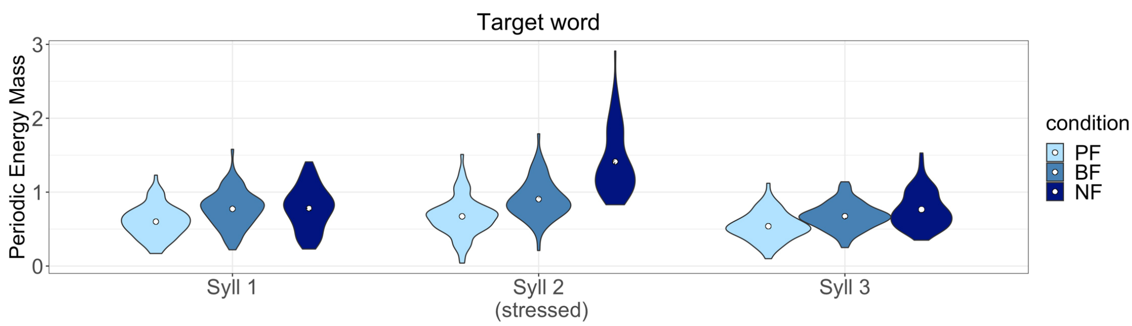


Figure 60. PEM for each syllable in the target word. The stressed syllable is indicated by the label stressed in parentheses. White dots indicate mean values.

Among the speakers of this set, one speaker was selected (B03), on the basis of fluency and naturalness of the prosodic realisation. For the 20 triplets produced by the speaker, 10 were chosen as stimuli for the perception experiment. In the next section (6.2.5.4), the acoustic features of the stimuli used for the perception experiment will be described.

6.2.5.4 Results of the selected stimuli

Figure 61 shows the results of the pitch accent analysis of the 10 utterances used as stimuli for the Bari variety. The verb in BF was realised with a falling pitch (H+L*; 100% of the cases), in NF either with the same falling pitch (20% of the cases) or with a slightly rising pitch (H*, in 80 % of the cases). When occurring in PF, the verb always bears a (rising-)falling pitch (H*+L, in 100% of the cases). All the occurrences of the target in the different conditions were realised with a different F0 contour: H+L* in BF, H*+L in NF and with no movement in PF. No movement is also found in the noun in PP when occurring in NF and PF (both post-focal positions for the noun in PP). By contrast the absence of movement was found for the noun in PP in BF only in a small percentage of cases (20%), while the remaining cases the pitch was falling (H+L*).

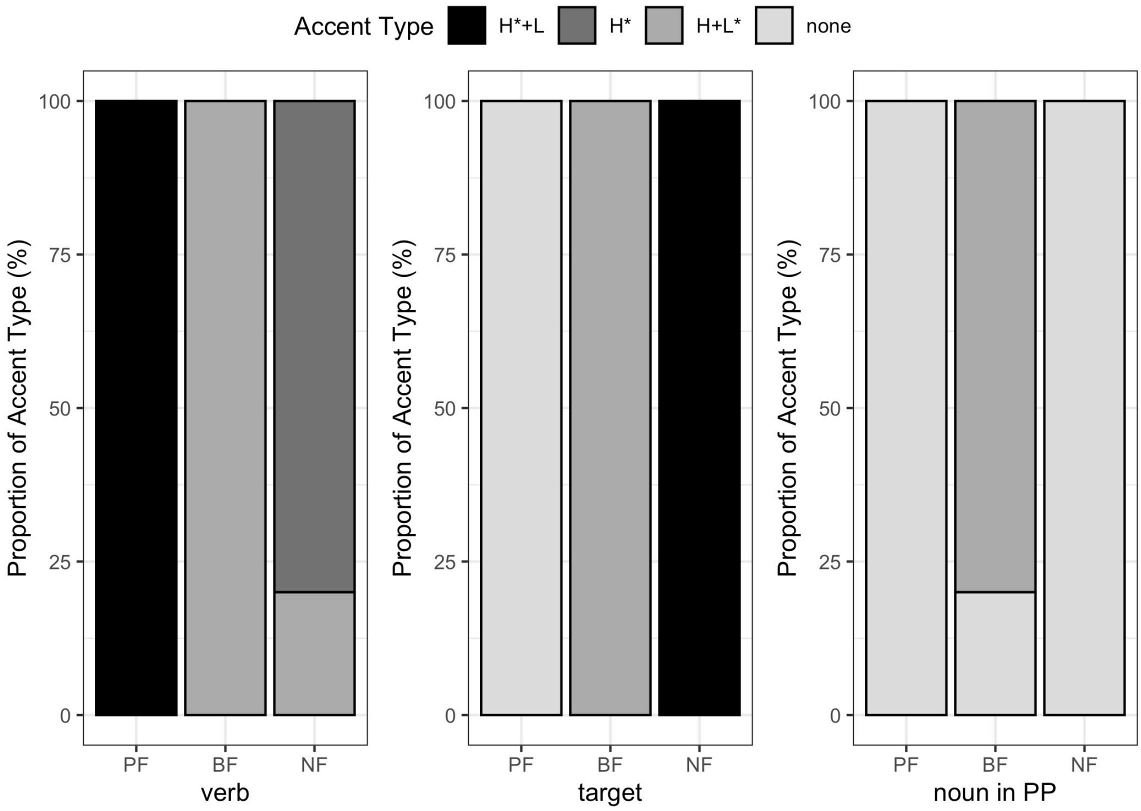


Figure 61. Accents type proportion in the 10 utterances selected as stimuli for the perception experiment (speaker B03).

Figure 62 (right panel, B., D. and F.) and Figure 63 show PEM for the domain of the stressed syllable and for the domain of the whole word respectively. Results of the statistical analyses are reported in Table 34, 35, 36 and 37. For the target word the difference in PEM between broad focus and the other two conditions was confirmed both for the domain of the stressed syllable and for the whole word.

In more detail, for the target in post-focal position the PEM values is lower in comparison to the PEM of the broad focus, while the value of narrow focus is increased. The verb shows higher values of PEM for the post-focal condition (where the verb is realised in narrow focus) compared to broad focus. The comparison between verb occurring in broad focus and in the narrow focus condition (where the verb is realised in pre-focal position) is not significant. Surprisingly, the noun in PP shows higher PEM values in the narrow focus condition compared to broad focus, while values of broad focus do not differ from the ones in post-focal condition. Results for these comparisons are reported in Table 34.

In post-focal position (Figure 62B.), the PEM of the stressed syllable of the target is significantly lower than that of the stressed syllable of the verb, by contrast, it is not significantly different from that of the stressed syllable of the word in PP. In broad focus (Figure 62D.) the PEM for the stressed syllable of the verb is significantly higher than that of the target, while the PEM for the noun in PP is significantly lower than for the stressed syllable of the target. In narrow focus (Figure 62F.), the difference between word position is also significant, showing a decrease in PEM of the two other word positions compared to the target. Results for these comparisons are shown in Table 35.

In the domain of the whole word (Figure 63), the PEM for the target in broad focus is significantly higher than that of the target in post-focal position, while it is significantly lower than the one in narrow focus. For the verb in the post-focal condition (where the verb is narrowly focussed), PEM is significantly higher than in broad focus and in narrow focus condition (where the verb is in pre-focal position). The noun in PP shows higher values of PEM when it occurs in post-focal condition compared to when it occurs in the broad focus condition. However, difference between the noun in PP in the broad focus condition and the same noun in the noarrow focus condition does not reach significance.

The PEM of the target in broad focus in not significantly higher than that of the verb though is significantly higher than that of the noun in PP. The PEM of the target in narrow focus is significantly higher than that of the verb and of the noun in PP. In post-focal position, the PEM of the target is significantly lower than that of the verb and significantly higher than that of the noun in PP. Results of the statistical analyses are reported in Table 36 and Table 37.

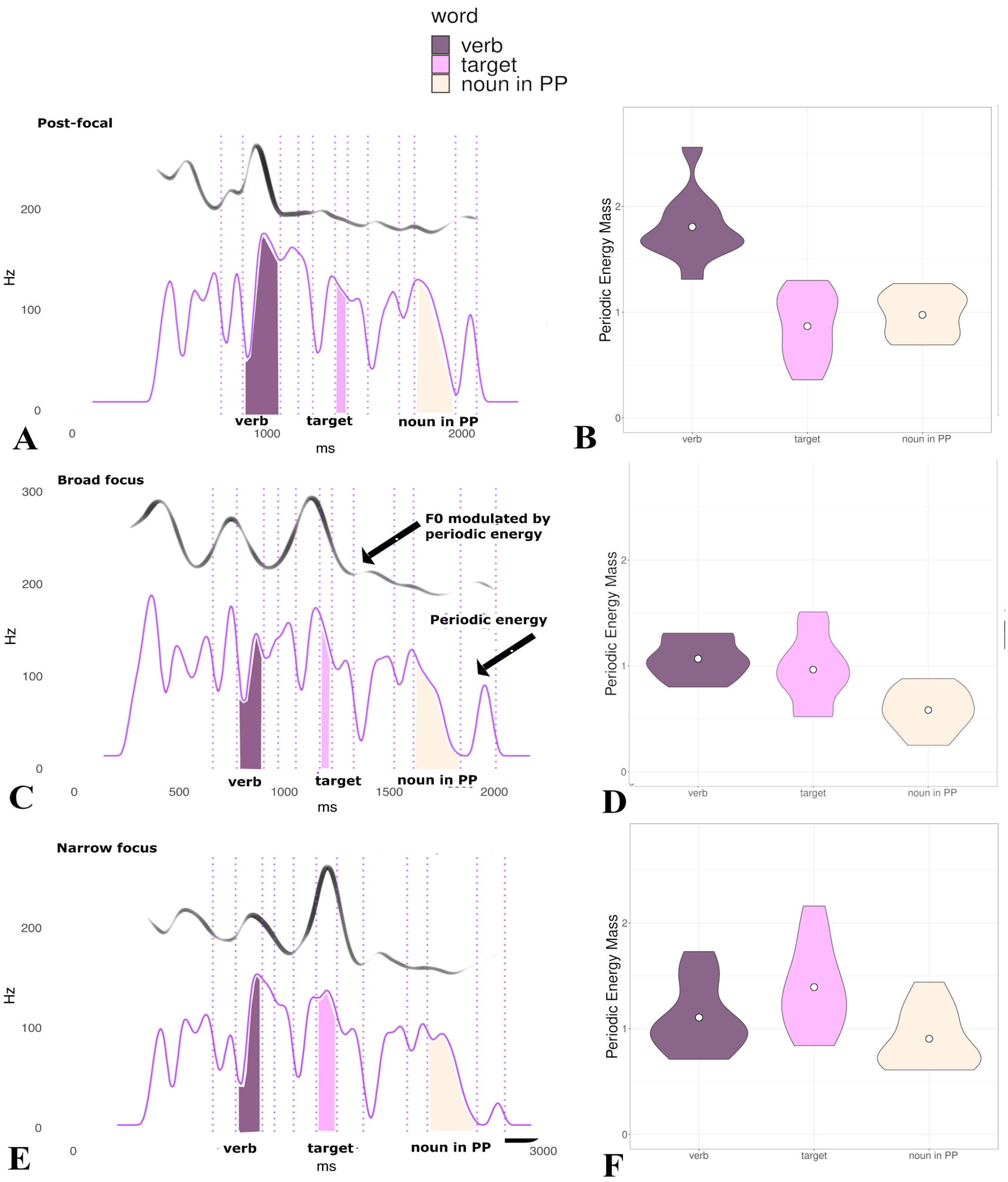


Figure 62. A., C. and E. (left panel) example of the sentence *Bisogna pesare la farina con la bilancia* (One should weigh the flour with the scale): upper curve F0 modulated by periodic energy, lower curve periodic energy. Y-axis show values in Hertz (Hz), x-axis show time in milliseconds (ms). Vertical lines show boundaries of the stressed syllable for every critical word. Parts of the periodic energy filled with colour correspond to the area under the periodic energy curve relative to each stressed syllable of the critical words. This is the value that corresponds to the Periodic Energy Mass. B., D. and F. (right panel) show Periodic Energy Mass for the critical words (verb, target noun and noun in PP) of every stimulus. The Y-axis shows relative values of Periodic Energy Mass, the x-axis shows the independent variable of the word. The figure shows the three prosodic conditions: post-focal (A. and B.), broad focus (C. and D.) and narrow focus (E. and F.). White dots on the violin plots indicate mean values.

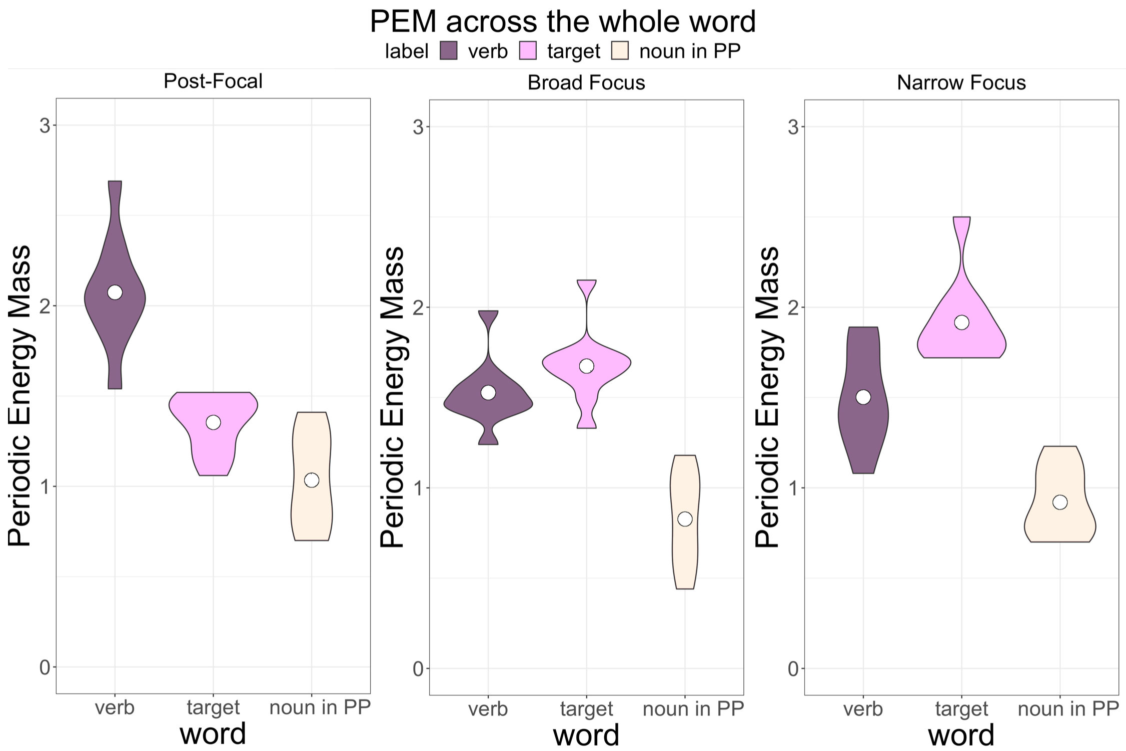


Figure 63. Relative PEM for the entire window of the words. Left panel shows values of PEM for verb, target and noun in PP in post-focal condition (verb in focus). The panel in the middle, shows values of PEM for verb, target and noun in PP for broad condition. The right panel shows values of PEM for verb, target and noun in PP for post-focal condition (target in focus). White dots on the violin plots indicate mean values.

	verb	**target**	**noun in PP**
PF	**β = -0.95 ± 0.08, p < 0.0001**	**β = -0.21 ± 0.06, p = 0.001**	β = 0.03 ± 0.05, p = 0.51
NF	β = -0.07 ± 0.08, p = 0.42	**β = 0.53 ± 0.06, p < 0.0001**	**β = 0.14 ± 0.05, p = 0.01**

Table 34. Results of the mixed model analyses of the differences between stressed syllables' values in each condition in each word position. Values are relative to the intercept, corresponding to the relative critical word (verb, target and noun in PP) in broad focus. Significant results are displayed in bold.

	BF	**NF**	**PF**
verb	**β = 0.23 ± 0.06, p < 0.001**	**β = -0.38 ± 0.09, p < 0.0001**	**β = 1.4 ± 0.09, p < 0.0001**
noun in PP	**β = - 0.22 ± 0.06, p = 0.001**	**β = -0.64 ± 0.09, p < 0.0001**	β = -0.02 ± 0.09, p = 0.79

Table 35. Results of mixed model analyses of the differences between each word position within each condition. From left to right: broad focus, narrow focus and post-focal position. Results refer to the stressed syllables. Significant results are displayed in bold.

	verb	target	noun in PP
PF	**β = 0.55 ± 0.1, p < 0.0001**	**β = - 0.32 ± 0.062, p < 0.0001**	**β = 0.21 ± 0.66, p = 0.01**
NF	β = -0.32 ± 0.14, p = 0.76	**β = 0.24 ± 0.064, p = 0.002**	β = 0.12 ± 0.68, p = 0.11

Table 36. Results of mixed model analyses of the differences between the entire word values in each condition in each word position. Values are relative to the intercept, corresponding to the relative critical word (verb, target and noun in PP) in broad focus. Significant results are displayed in bold.

	BF	NF	PF
verb	β = -0.15 ± 0.98, p = 0.15	**β = -0.41 ± 0.11, p = 0.001**	**β = 0.72 ± 0.11, p < 0.0001**
noun in PP	**β = - 0.85 ± 0.98, p < 0.0001**	**β = -0.1 ± 0.11, p < 0.0001**	**β = -0.32 ± 0.11, p = 0.01**

Table 37. Results of mixed model analyses of the differences between each word position within each condition, from left to right: broad focus, narrow focus and post-focal position. Results refer to the domain of the whole word.

Figure 64 shows the values of scaling and synchrony for each syllable of the stimuli that presents the lexical stress in penultimate position (indicated in the figure by the label *stress*). The figures presenting the remaining combinations can be retrieved at https://osf.io/5m8hw/. The values of scaling and synchrony confirm the intonation analysis reported above. Negative values of both measures indicate that the critical words in BF bear a falling pitch on the stressed syllable. Positive values of scaling and values of synchrony around zero, indicate that the verb in NF presents a slightly rising pitch. By contrast, on the target in NF highly positive values of scaling and synchrony values around zero indicate rising-falling pitch. This pattern of values is observed also for the verb in PF, whereas in the same condition the critical words following the verb are produced with a flat contour. Values around zero are registered also for the noun in PP occurring in the NF condition. In Table 38 means and standard deviations of scaling and synchrony for the stressed syllables of each prosodic condition and word position are reported. These measures are relative to all the stimuli. Results of the statistical analysis for scaling and synchrony are summarised in Table 39 and in Table 40. As for scaling and synchrony, for the stressed syllable of the target word the models confirmed the difference between post-focal position and broad focus, showing higher values for the former condition. Values of scaling and synchrony differed between broad focus and narrow focus, showing a high increase for the latter condition.

In broad focus condition, the stressed syllables of the verb and the noun in PP have higher values of scaling than the target, while values of synchrony differ only between target and noun in PP. In the narrow focus condition, values of scaling for the target are significantly higher than in the other two word positions. Values of synchrony are not significantly different in different word positions. In the post-focal condition, the stressed syllable of the verb showed significantly lower values of scaling than that of the stressed syllable of the target. By contrast, the difference between scaling of target and scaling of noun in PP was not significant. Values of synchrony did not show significant differences among position of the words.

condition	verb	target	noun in PP
BF	Scaling: -49.19 (17.43) Synchrony: -26.1 (5.7)	Scaling: -65.77 (22.69) Synchrony: -23.7 (10.11)	Scaling: -17.43 (7.44) Synchrony: -3.9 (6.49)
NF	Scaling: 13.83 (19.88) Synchrony: -2.5 (12.26)	Scaling: 55.45 (26.18) Synchrony: -1.7 (7.79)	Scaling: -8.71 (5.35) Synchrony: -6.0 (5.48)
PF	Scaling: 50.67 (23.90) Synchrony: -9.8 (10.12)	Scaling: -3.14 (8.05) Synchrony: -5.2 (3.33)	Scaling: -4.15 (3.99) Synchrony: -5.6 (4.30)

Table 38. Mean and standard deviation (in parenthesis) of scaling and synchrony for stress syllables of the critical words for all the stimuli.

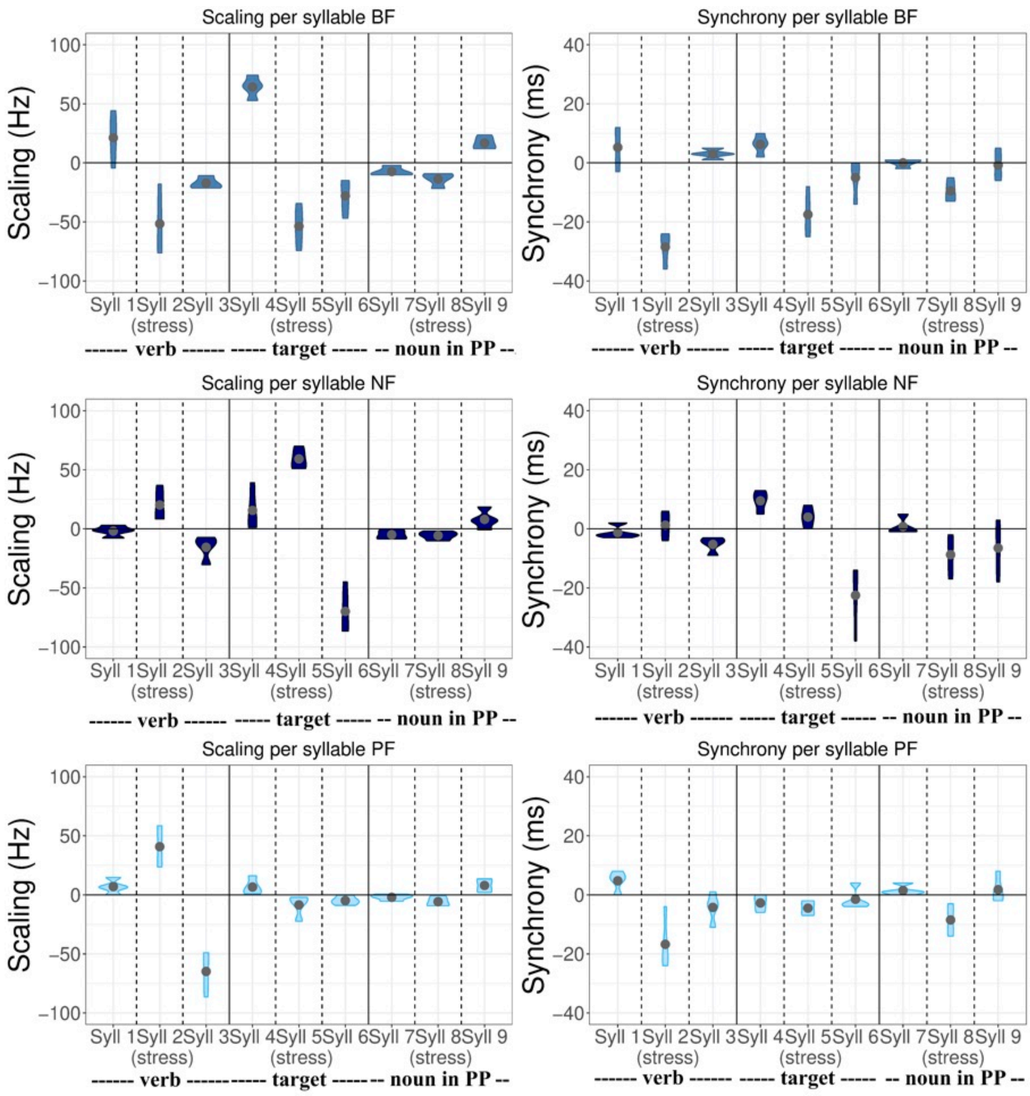


Figure 64. Values of scaling (on the left) and synchrony (on the right) for each syllable of the utterances in BF (first row), NF (second row) and PF (third row). The graph shows the results of stimuli where both verbs and nouns in PP have penultimate stress (target nouns always have lexical stress in penultimate position). Dotted vertical lines mark each syllable, solid vertical lines mark critical words. Grey points indicate mean values.

	verb	target	noun in PP
PF	**Scal: β = 99.86 ± 9.2, p < f0.0001** **Synch: β = 16.3 ± 3.97, p = 0.001**	**Scal: β = 62.63 ± 7.18, p < 0.0001** **Synch: β = 18.5 ± 2.74, p < 0.0001**	**Scal: β = 12.29 ± 1.79, p < 0.0001** Synch: β = -1.7 ± 1.6, p = 0.3
NF	**Scal: β = 63.02 ± 9.2, p < 0.0001** **Synch: β = 23.6 ± 3.97, p < 0.0001**	**Scal: β = 121.21 ± 1.18, p < 0.0001** **Synch: β = 22 ± 2.74, p < 0.0001**	**Scal: β = 8.72 ± 1.79, p = 0.0001** Synch: β = -2.1 ± 1.6, p = 0.21

Table 39. Results of the mixed model analyses on the values of scaling (scal) and synchrony (synch) of the differences between the stressed syllables' values in each condition and in each word position. Values are relative to the intercept, corresponding to the relative critical word (verb, target and noun in PP) in broad focus. Significant results are displayed in bold.

	BF	NF	PF
verb	**Scaling: β = 16.58, ± 6.3, p = 0.02** Synch: β = -2.4 ± 3.43, p = 0.49	**Scaling: β = - 41.61, ± 7.31, p < 0.0001** Synch: β = -0.8 ± 3.7, p = 0.83	**Scaling: β = - 53.08, ± 5.91, p < 0.0001** Synch: β = -4.6 ± 2.81, p = 0.12
noun in PP	Scaling: β = 48.33 ± 6.3, p = 0.87 **Synch: β = 19.8 ± 3.43, p < 0.0001**	**Scaling: β = - 65.15, ± 7.31, p < 0.0001** Synch: β = -4.3 ± 3.71, p = 0.26	Scaling: β = -1.01 ± 5.91, p = 0.87 Synch: β = -0.4 ± 2.81, p = 0.89

Table 40. Results of the mixed model analyses of the differences between each word position in (from the left) broad focus (BF), narrow focus (NF) post-focal position (PF). Values are relative to the intercept, corresponding to the target in the relative condition. Significant results are displayed in bold.

6.3 Summary of the acoustic parameters in UI and BI

In this section, an interim summary of the acoustic characteristics of the stimuli selected in the two varieties and described in 6.2.5 and 6.2.6.4 is provided.

In UI, both the categorical analysis of pitch contour and the continuous parameters of synchrony and scaling show that the critical words in BF are realised either with falling accents (H+L*) or with rising accents (H*, note that in the noun in PP all the realisations feature a falling accent), expanding in a limited range both within (synchrony) and between syllables (scaling). Characteristic of the target and the noun in PP is the presence of a rise associated with the last syllable of these two constituents. These rises are identifiable as high phrase accents (H-), at the end of the target word, and as a sequence of high phrase accents and boundary tones (H- H%) at the end of the utterance.

Again, both categorical analysis and continuous parameters show that in NF the target word is realised with a rising accent, with an increase in range compared to the rising accents found in broad focus, classified as L+H* and H* respectively. Values of synchrony and scaling reached in the stressed syllable of the target word in NF are rather extreme (relative to the range of the speaker) and make the word particularly stand out if compared with the verb and the noun in PP, characterised by very reduced movement. An exception to this reduced movement is the final rise associated with the end of the utterance (last syllable of the noun in PP), similar to the one found in the BF condition.

The verb in PF condition is characterised by a steep rise (L+H*) associated with the stressed syllable, followed by a rather extreme fall which reaches its target on the subsequent syllable. The target is characterised by a low and flat contour with (nearly) no movement. Slightly more movement is realised in the noun in PP, which again ends with a final rise.

Considering the F0 modulations of the UI stimuli, the target in NF and the verb in PF represent the most prominent words in their respective utterances, while all the critical words in BF seem to be more or less equally prominent. In particular, on the basis of the movement in pitch, a hierarchy could be established for the target, with NF at the top and PF at the bottom: NF > BF > PF. However, pitch movement is not the only cue to prominence. Particularly interesting as an additional measure for prominence are values of PEM. In the stimuli, the highest values of PEM are registered by verb in PF and target in NF, both in the stressed syllable and in the entire word domain. They demonstrate increased values not only compared to the other words in the same utterance, but also compared to critical words in other conditions and positions. By contrast, in BF values of verb, target and noun in PP are similarly distributed in the domain of the whole word, and show a difference between verb and target on the one hand and noun in PP on the other only for the domain of the syllable.

Interestingly, for UI stimuli, values of PEM of the target in BF and in PF show a similar distribution, both when considering the domain of the stressed syllable and in the domain of the entire word. These values could hypothetically cue a similar level of prominence for the target in PF compared to the target in BF, despite the lack of pitch movement for the former condition, but see 6.4.

The stimuli in BI generally show similar patterns to the ones of the stimuli in PF, even though some differences are registered. Also for BI, the categorical analysis and the continuous measures shows that there is a gradual increase in prominence from the target in PF to the one in BF and the one in NF, this time also more clearly shown by PEM values, which differ between the target of PF and BF (both in the stressed and in the entire word domain). More in detail, the critical words of BF in stimuli of Bari are mostly realised with falling accents (H+L*), presenting a great excursion for verb and target, while a highly reduced one for the noun in PP. The target in NF is marked by a (rising-)falling accent (H*+L), showing high excursion. The target in this condition is therefore characterised by more prominent pitch movement than the BF condition. In addition, the target in NF is also characterised by greater pitch excursion compared to verb and noun in PP, similarly to what is attested for UI.

The target in PF is realised with a flat and low pitch, which is preceded by a (rising-)falling accent characterised by rather high excursion. Similar to UI stimuli, for BI stimuli this characteristic of PF suggests that this word should occupy the lowest position in the hierarchy of acoustic prominence of the target. This is in addition confirmed by PEM values, which for both the domain of the word and the one of the syllable, show the lowest values for PF compared to NF and BF. The greatest values of PEM are registered for the verb in PF condition, which, similarly to the target in NF is characterised by the highest acoustic prominence compared to the other words in the sentence. Interestingly, in BF verb and target, despite showing a similar contour, are characterised by different PEM values where the latter shows higher values compared to the former (though only in the domain of the whole word). Finally, the noun in PP shows the lowest values within the conditions, while across conditions, it shows higher values when occurring in PF condition than when occurring in BF condition.

The next section comprises an overview of the predictions on the ratings of prominence in light of the acoustic characteristics of the stimuli.

6.4 Prominence Rating Task

In this experiment, the role of language-specific expectations in the perception of prominence is investigated by recording prominence ratings of the same signal by different groups of listeners. Listeners were presented with utterances out of context in order to minimise the effect of expectations derived by an immediately preceding context, which would elicit a specific information structure for the trial utterances. In this way, the only expectations created in the perception of the acoustic signal were language-specific, i.e., derived from the knowledge of the grammar of the language.

The ratings recorded corresponded to the ones of: (i) native speakers of the Udine variety of Italian rating their own variety (UI), (ii) native speakers of the Bari variety of Italian rating their own variety (BI) and (iii) native speakers of German (henceforth, G), learning Italian, rating both UI and BI. Section 6.5.1 will present the comparison between the groups of native speakers of the two varieties of Italian. Section 6.5.3 will present the comparison between learners and Bari participants, both rating BI (6.5.3.1), and the comparison between learners and Udine participants, both rating UI (6.5.3.3). In the experiment, ratings of all critical words in all conditions will be discussed. However, specific hypotheses are not present for all these positions and conditions.

A general hypothesis would be that narrowly focussed words (target in NF and verb in PF) would be perceived as more prominent than words in broad focus and in post-focal position, given the enhanced cues to prominence characterising them (highly prominent (rising-)falling accents with great excursion and higher values of PEM) compared to the other two conditions. For the three groups of participants that are considered in this experiment (delineated above), the perception of the difference between the target in narrow focus and the target in the other two conditions should not vary. Indeed, following the signal, the contrast between narrow focus and the other two conditions should be clearly perceived by all listeners. In addition, expectations regarding the differences in the degree of prominence between these conditions should not differ between groups.

Considering only the acoustics, another hypothesis is drawn, this time for the comparison between the target in broad focus and in post-focal position in the ratings relative to the variety of Bari: ratings for the target in broad focus should be higher than ratings for the target in post-focal position for both groups of participants rating this variety (learners and Bari participants). In fact, the higher acoustic prominence of broad focus (higher PEM and presence of pitch movement) should yield higher results for the broad focus condition. A different scenario can be thought of for the Udine stimuli, in which the similar distribution of PEM between the target in BF and the target in PF could lead to a more similar distribution of the prominence ratings. However, the presence of the pitch movement associated with the stressed syllable of the target in BF, should still yield a difference between the two conditions, since the presence of an accent featuring pitch movement should increase the likelihood of prominence perception (see Baumann & Winter, 2018), but this difference can probably be attenuated by the PEM distribution.

Making predictions only based on the acoustics, the two groups of listeners rating the Bari variety (BI and G) should not differ among them, but could be different from the two groups rating the Udine variety (UI and G): the latter two groups could rate target in BF with more similar values to target in PF than the ones found in the former two groups (BI and G). However, since prominence perception is not only influenced by signal-based factors, but also by language-specific expectation-based factors, the perception of prominence of target in BF and target in PF could be different for different groups. In the UI group, values of perceived prominence of the target in post-focal region are expected to be lower than the values of perceived prominence in broad focus. In fact, in addition to the differences conveyed by the F0 contour in the two conditions in the Udine stimuli, which should lead to a lower level of perceived prominence in post-focal position, Udine listeners should expect the degree of prominence of post-focal position to be low not only for the lack of movement but also because the production experiment in Chapter 5 showed that most of the speakers differentiated between PEM distribution in broad focus and in post-focal position. Therefore, the most frequent distribution of PEM values implies the differentiation between post-focal and broad focus. PEM values are not expected to play a role in this distinction and listeners are expected to rate prominence based on their expectations and on F0 movement. Moreover, the distribution of accents in the Udine variety is such that also in questions the post-focal region of the utterances does not present movement associated with the stressed syllable, differently from the Bari variety (see 2.3.3.1). Thus, participants of the Udine group, should not have any expectations of a lack of attenuation in the post-focal position.

For the BI group, the ratings of the target word are expected to have values that are similar between BF and PF. This prediction originates from the fact that BI listeners should expect a higher degree of prominence in the post-focal position, given the relatively high probability of finding cues to prominence in post-focal position. The comparison between BI group and UI group, should yield different results in the prominence level assigned to the post-focal position. Indeed, given the difference in the distribution of prominence in the two varieties, expectations concerning the prominence in the post-focal position should differ.

For the German learners, the hypothesis is that there will be a three-way differentiation in the prominence ratings of the target for both the varieties rated: NF would be rated as more prominent than BF, which in turn would be rated as more prominent than PF. These predictions are derived not only from the acoustic signal, which would clearly yield these results for learners rating the Bari variety, but also from the expectations of finding attenuation in the post-focal position. These are matched by the signal in the Bari variety, and are to some extent not matched by the signal for the Udine variety (similar levels of PEM between BF and PF). In this latter case, effects of PEM could yield a difference between the ratings of the two varieties, resulting in a lower distance between BF and PF for the Udine variety compared to the Bari variety.

Ratings of learners are expected to differ, to some extent, to ratings form native speakers. This differentiation could be a more general differentiation coming from the higher difficulty in processing prominence relations in the L2 (see for example Akker & Cutler, 2003 and 3.9) and from language specific trends that might be more generally present (see for example Chen, Rietveld & Gussenhoven, 2001). This could yield differences in the overall level of perceived prominence, for example, resulting in a general upward or downward shift of all the values. Learners are expected to be influenced by their proficiency, showing a pattern more similar to natives with the increase in proficiency (see 3.9). In addition, more precise differentiation regarding the perception of prominence relations would come from language-specific expectations concerning the distribution of prominence in an utterance, which, being different from the one found in the Bari variety, could lead to differences in the two groups.

In the prominence ratings of all the groups a sizeable interindividual variability is expected. Indeed, listeners have previously been reported to react in different ways to prominence-lending cues and to their interaction (Baumann & Winter, 2018; Cangemi et al., 2015; Cole, Mo & Baek, 2010; Cole, Mo & Hasegawa-Johnson, 2010).

6.4.1 Procedure

To determine perceived prominence of the critical words the task designed was inspired by the *Rapid Prosody Transcription* (RPT) method developed by Cole, Mo and Hasegawa-Johnson (2010). The method used incorporates certain modifications: while in the RPT, participants listen to long excerpts of recorded speech and are asked to underline the words perceived as prominent while listening to them, the present experiment required the participants to listen to utterances in isolation and to rate three words of the utterances (the critical words, see 6.2.3). Moreover, participants could listen to the utterances as many times as they wanted (but note that in order to assure a better homogeneity the trials that presented more than five playbacks were excluded, together with their corresponding items in the other two conditions). Participants generally listened to the stimuli less than 5 times (median of the Bari group = 1; median of the Udine group = 1; median of the German group rating BI = 2; median of the German group rating UI = 1). In the experiment with BI stimuli one participant of the German group listened to 5 trials more than 5 times (range from 6 to 13 times). Therefore, this participant was excluded from the analysis, since the number of trials to be excluded (5 trials × 3 conditions) was too high. Of the German group (henceforth, G), one other participant listened to 2 trials more than 5 times. In the BI group 3 participants listened to either 1 or 2 trials more than 5 times. From the UI group none of the participants listened to any of the trials more than 5 times. By contrast, learners listening to UI stimuli had more trials to which they listened more than 5 times: one participant listened to 4 trials from 6 to 11 times and another participant listened to one trial 6 times. The former participant was excluded from the analysis. Of the latter participant, one trial with its relative items was excluded from the analysis.

In the present task, judgments were not binary, as the prominence rating task was implemented using a visual analogue scale (see for example Arnold, Wagner & Möbius, 2011; Gussenhoven & Rietveld, 1998; Terken, 1997). The thickness and darkness of the scale indicated iconically the degrees of prominence, with thicker lines representing the degrees of prominence (as in Baumann & Röhr, 2015). An example of the scale is provided in Figure 65. The visual analogue scale and the fact that three different words were rated per item allowed to consider the relational nature of prominence and to investigate fine-grained details of prominence perception. Participants were, therefore, not restricted to predefined values to interpret the levels of prominence, but were free to interpret those levels along the visual analogue scale. In addition, stimuli were presented in isolation, in order to prompt the listeners to address the acoustic information and to prevent influence of the context on the perceptual judgements (see Chapter 3).

The task was performed online through the ‘SoSci Survey’ software (Leiner, 2014). In the instructions for participants, the term prominence was replaced with “salienza” (*salience*), for Italian and with “hervorgehoben” (*highlighted*) for German. This decision was made because the term prominence is not straightforward to understand. Moreover, particular attention was paid to the wording of the instructions, since previous studies have shown that it may also entail meaning-based interpretations and not only acoustically-based ones (Cole et al., 2019). The formulation of the instruction in the present experiment was the following: “how salient/highlighted do the words sound *at your ear*”, explicitly created to direct listeners towards an acoustically-based interpretation. The left pole of the visual analogue scale was labelled “per niente saliente” (*“not at all salient”*), the right pole as “al massimo saliente” *(“at most salient”*). For German listeners, the labelling was “gar nicht hervorgehoben” (*“not at all highlighted”*) and “sehr stark hervorgehoben” (*“at most highlighted”*). Each position of the slider corresponded to a number range from a minimum of 1 to a maximum of 101, see Figure 65 for the example of the setup.

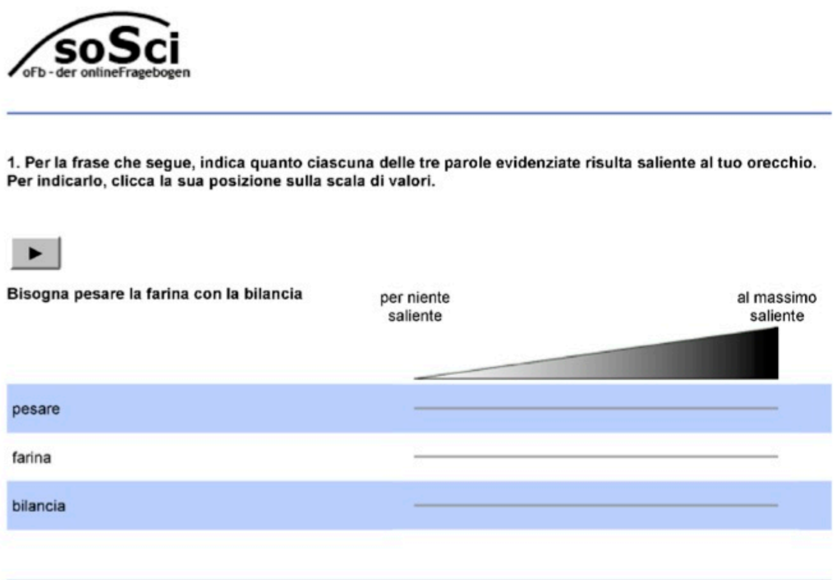


Figure 65. Example of the experimental setup in the web-based perceptual evaluation task.

Stimuli were presented both acoustically and orthographically on the computer screen. Materials were presented in a pseudo-randomised order. Each utterance composed of the same word was considered as an item. No item was repeated before an interval of 6 items and no more than two items in the same prosodic condition were presented consecutively. The perceptual evaluation task was preceded by a short training session.

6.4.2 Statistical analysis of the prominence ratings

To statistically analyse the prominence ratings, the numbers corresponding to the positions of the slider in the visual analogue scale were considered as dependent variable. The values presented a high degree of variability, both endemic to the investigation of prominence ratings and to the fact that participants could use the scale freely. To account for variation in participants' use of the scale, prominence ratings were transformed to z-scores on a by-participant basis (see Schütze & Sprouse, 2014). The three focal conditions, the native language (G, UI or BI) of the participants and the position of the word in the sentence (verb, target and final noun in the PP) were considered as independent variables. For all the analyses we used R (R Core Team, 2019) and *lme4* (Bates et al., 2012) to perform linear mixed effects models of the relationship between condition, critical word and perceived prominence. To obtain p-values either type II Wald test of the package *car* (Fox & Weisberg, 2019), or the *afex* package (Singmann et al., 2018) were used.

To test for group effects (UI vs. BI, UI vs. G and BI vs. G), linear mixed effects analyses of the relationship between prominence ratings, word position, condition and group were performed. As fixed effects, POSITION, CONDITION and GROUP were entered into the model. As random effects, intercepts for SUBJECT and ITEM as well as by-subject and by-item random slopes were considered.

Crucially, the prediction was that the three different groups would present differences in rating the target, but not elsewhere. As a consequence, given that for the other contrasts we did not predict a difference, the effect on the comparison between conditions in the target was expected to be small in the full model. Therefore, additional analyses on the perceived prominence ratings within each group were conducted. The corresponding model had as fixed effects CONDITION and random effects intercepts for SUBJECT and ITEM as well as by-subject random slopes.

To test the specific hypotheses for the effect of word position (verb, target and noun in PP) in each condition, a follow up on interactions with the factor position was conducted by making a subset of the data per condition. On these subsets, linear mixed effect analyses with POSITION and GROUP as fixed effects and the intercepts and random slopes for SUBJECT and ITEM as random effects were performed.

Moreover, to test the hypothesis that German learners and the UI group attend more to the acoustic cues, while the BI group uses top-down information to a higher extent, correlation analyses between PEM and prominence ratings were computed. In addition, correlation analyses between prominence ratings with scaling and with synchrony were also run. The differences between these correlations in the groups were then tested by mean of Fisher's r-to-Z transformation.

Finally, to test possible differences in the comparison between the ratings of Germans in BI and UI, a model with CONDITION, WORD and VARIETY as fixed effects and intercepts for SUBJECT and ITEM as well as by-subject random slopes as random effects was fitted.

All stimuli, the data tables and the script for the statistical analyses can be retrieved at https://osf.io/t35he/.

6.5 Results of the Prominence Rating Task

6.5.1 Results Bari and Udine

Results of all the critical words (target, verb and noun in PP) in all conditions and for both groups (Bari and Udine) are illustrated in Figure 66 and Figure 67. Figure 66 shows the mean of ratings in all the words per condition, whereas Figure 67 shows their distribution.

A visual inspection of the graphs (Figure 66 and Figure 67) shows the tendency for the Udine listeners to perceive the prominence of the target word in broad focus with a higher level of prominence than the target word in post-focal position. This tendency is confirmed by the Wald type II test on the full model, which registered an interaction of CONDITION, POSITION and GROUP [$\chi2$ (4) = 12.56, p= 0.01]. A further inspection of the model indicates that the difference between the ratings of the target word in broad focus and post-focal position differs between the two groups, showing a more pronounced difference for the group of Udine listeners (β = 0.36 ± 0.17, $p = 0.04$). Figure 68 shows the pairwise comparison of the effects extracted from the model.

As predicted, the position of the word in the sentence had a general significant effect on listeners' ratings [$\chi 2(2) = 19.34$, $p < 0.0001$]. For both groups the target in narrow focus (bearing a rising accent) was perceived as significantly more prominent than the verb and the final noun (verb: β = -1.47 ± 0.09, $p < 0.0001$, noun: β = -1.1 ± 0.09, $p < 0.0001$). Again, for both groups, in post-focal condition the verb was perceived higher in prominence than the following target word (verb: β = 1.29 ± 0.09, $p < 0.0001$) and the final noun lower than the target (noun in PP: β = -0.34 ± 0.09, $p < 0.0001$). As far as the broad focus condition is concerned, the target word was perceived as more prominent than the final noun and not differently from the verb (noun: β = -0.25 ± 0.09, $p < 0.01$, verb: β = -0.12 ± 0.09, $p = 0.2$).

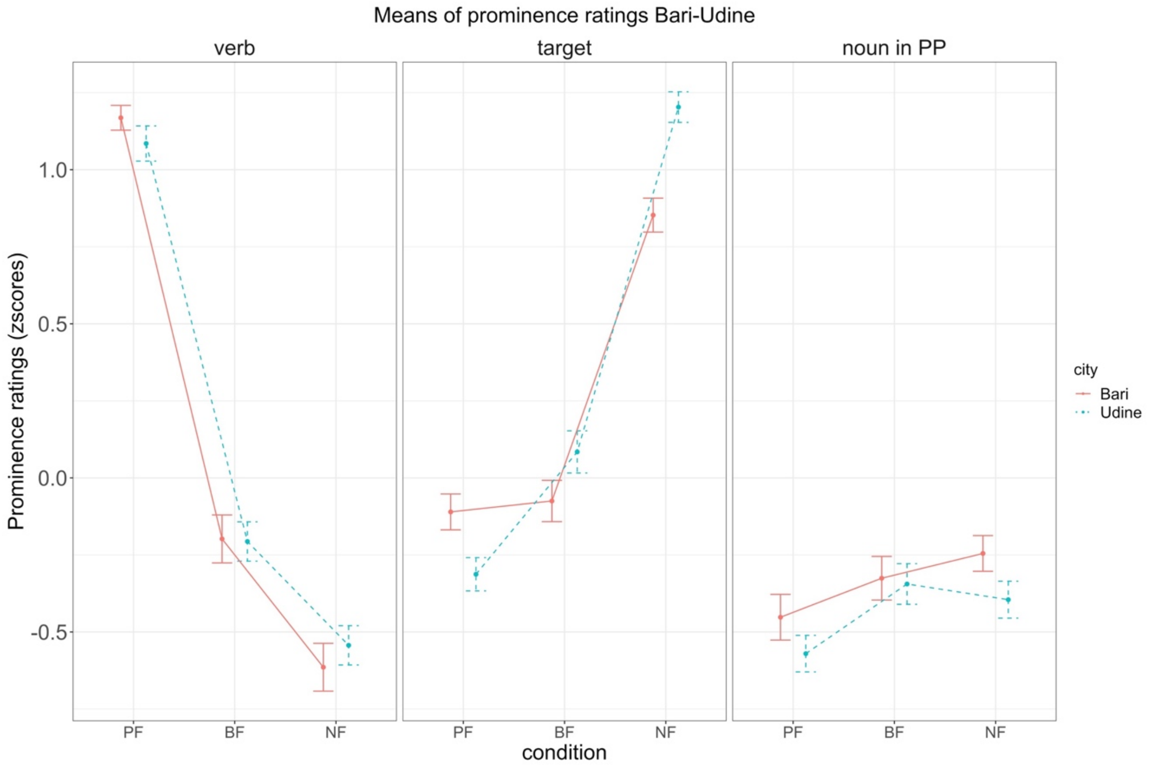


Figure 66. Means of prominence ratings for BI and UI for each word in each condition. Solid lines indicate the results for Bari, dotted lines the results for Udine. The different panels show from left to right verb target and noun in PP. On each panel, the x-axis shows from left to right PF, BF and NF.

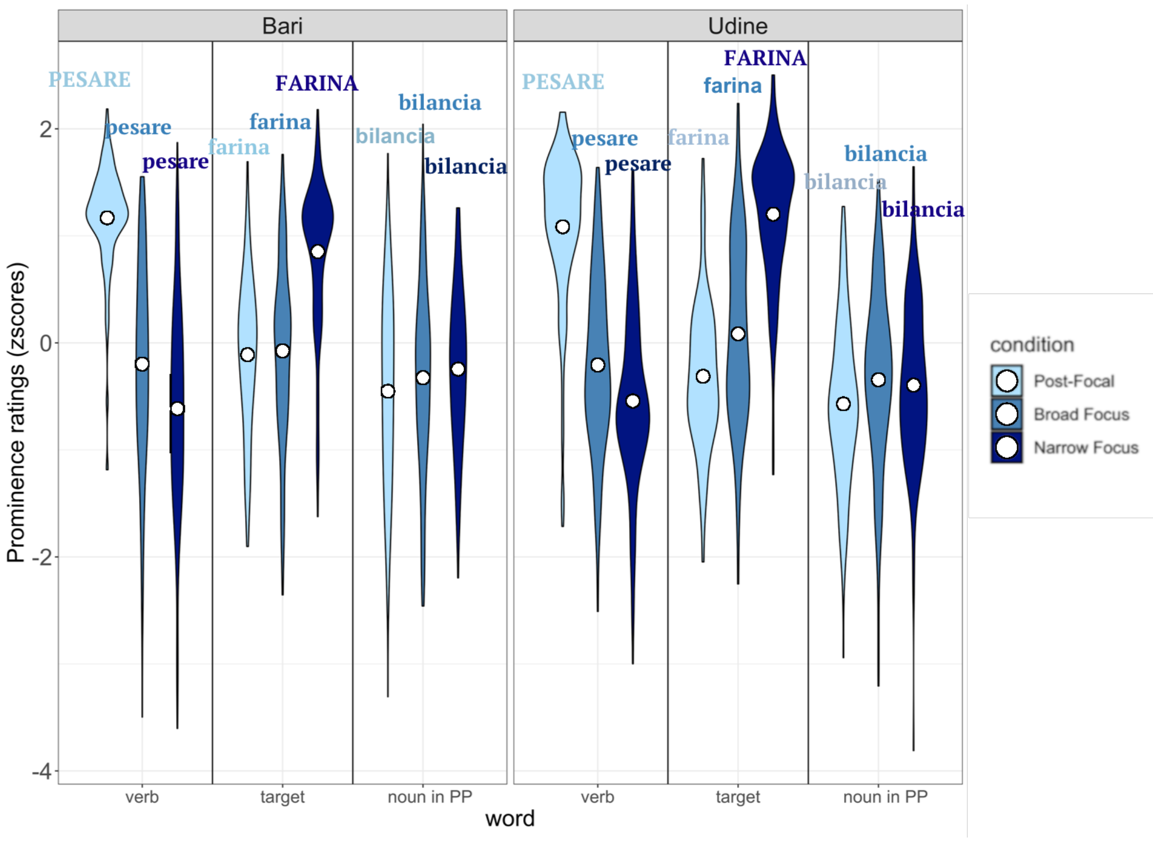


Figure 67. Prominence ratings for the target word by the two groups of participants: Bari group on the left, Udine group on the right. Each panel within the panel group shows values for the words in one position in the utterance, from the left, verb, target and noun in PP. The y-axis shows prominence scores (z-scored transformed). The x-axis shows the independent variable of the condition. In both panels: from left to right ratings for the post-focal condition (PF), broad focus (BF) and narrow focus (NF). Words in contrastive narrow focus are indicated by capital letters.

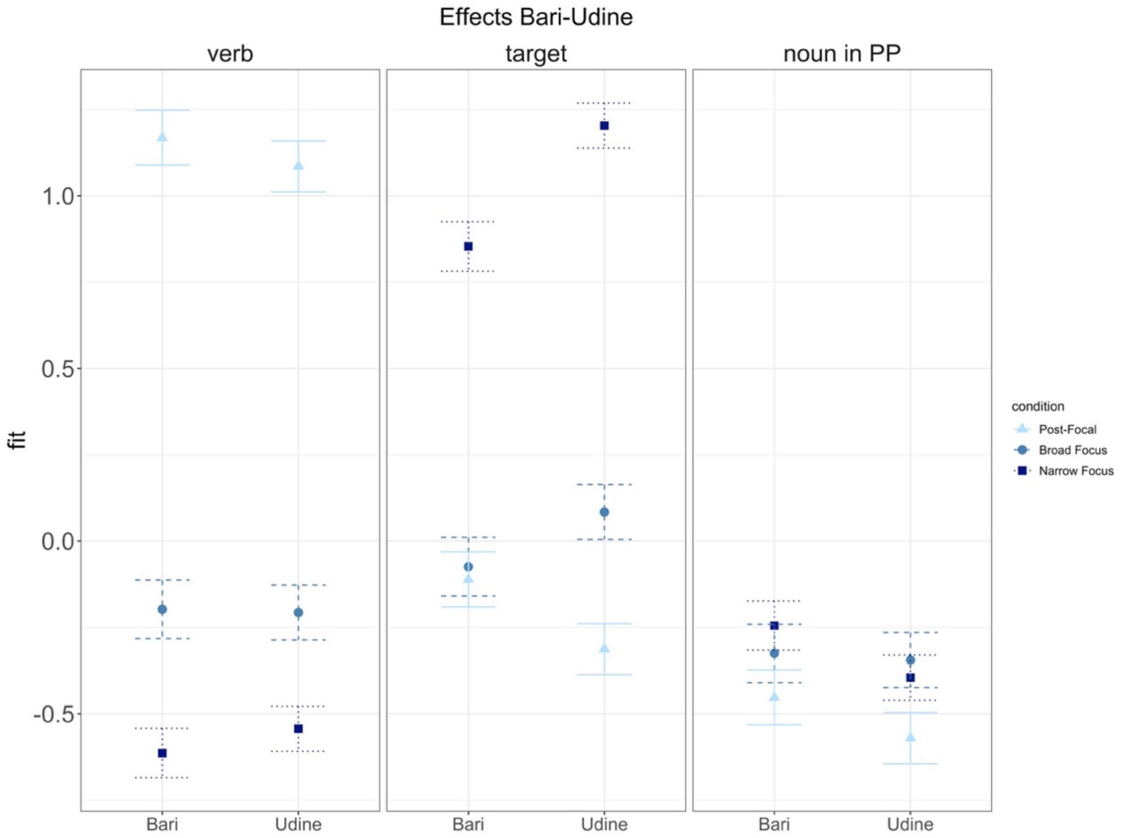


Figure 68. Effects registered by the full model. Error bars represent 83 % confidence intervals. Solid lines, dashed lines and dotted lines indicate error bars for PF, BF and NF respectively. Triangles, circles and squares indicate the means of PF, BF and NF respectively.

Figure 69 displays for each participant the means of the differences between the ratings for the target in narrow focus and post-focal position (NF-PF), between the ratings in narrow focus and broad focus (NF-BF) and between the ratings in broad focus and in post-focal position (BF-PF). As expected, there are many individual differences, mostly between broad focus and post-focal ratings. A visual inspection of the graph showed that the narrow focus condition was always perceived with a higher level of prominence than the post-focal condition, with the exception of one listener of the Bari group (25_BI). By contrast, in the comparison between broad focus and post-focal, there are differences between the two groups. For the majority of the Udine listeners the broad focus was rated higher than the post-focal position, with just three participants (3_I, 16_I, 14_I) rating post-focal higher than broad focus and with three participants making little differences (10_I, 17_I, 2_I). By contrast, the Bari group shows a higher tendency to rate post-focal target words more prominently than broad focus, with 7 out of 16 participants rating PF higher than BF.

To test whether in the target, the comparison between narrow focus and the other two conditions is higher than the comparison between broad focus and post-focal, a model with these differences as dependent variable was run. Results show that ratings for narrow focus differed from broad focus to a higher extent compared to the extent to which ratings for broad focus differed from post-focal position ($\beta = 0.89 \pm 0.24$, $p = 0.001$). The same tendency was also confirmed for the difference between narrow focus and post-focal, which was higher than the one registered between broad focus and post-focal ($\beta = 0.93 \pm 0.15$, $p < 0.0001$). The variable group showed had no effect in these comparisons.

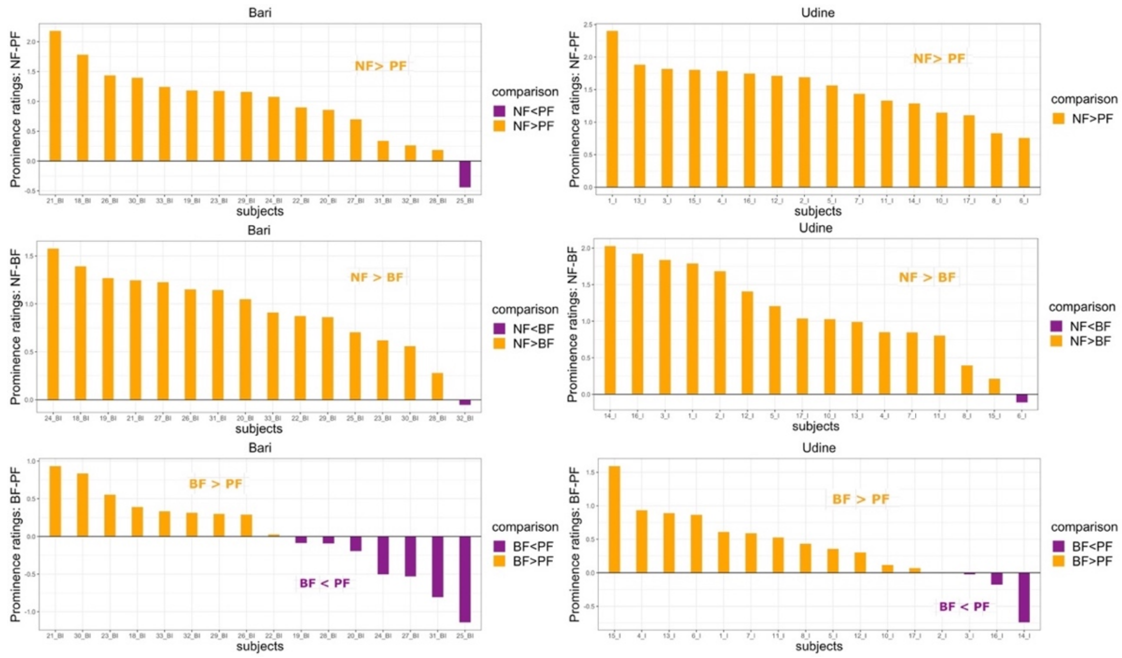


Figure 69. Differences of the prominence ratings of the target word between conditions for each subject of the two groups of participants. On the left listeners from Bari, on the right listeners from Udine. On the y-axis mean of the difference of the prominence scores (z-scored transformed). On the x-axis subjects. On the top, mean of the difference between the scores for narrow focus (NF) and the scores for post-focal (PF), in the middle mean of the difference between the scores for narrow focus (NF) and broad focus (BF), on the bottom, mean of the difference between the scores for broad focus (BF) and post-focal (PF). In the graphs on the top positive values indicate that the rating for NF are higher than the ones for PF. In the graphs on the middle positive values indicate that the ratings for NF are higher than the ones for BF. In the graphs on the bottom positive values indicate that the ratings for BF are higher than the ones for PF.

Figure 70 shows the results for the two groups (Bari and Udine) of the correlations between prominence ratings and PEM across word position. Both groups presented a significant correlation (Bari: $r(1363) = 0.4$, $p < 0.0001$; Udine: $r(1438) = 0.14$, $p < 0.0001$). Results of the Fisher's r-to-Z transformation for the difference in the correlation between the two groups registered a significant difference, with the group of Bari showing a higher correlation in comparison to the group of Udine (z = -7.54, p < 0.0001). This higher correlation was expected, given that the difference in the PEM of the entire word in the Udine dataset where present only for the noun in PP, while ratings presented a three-way distinction.

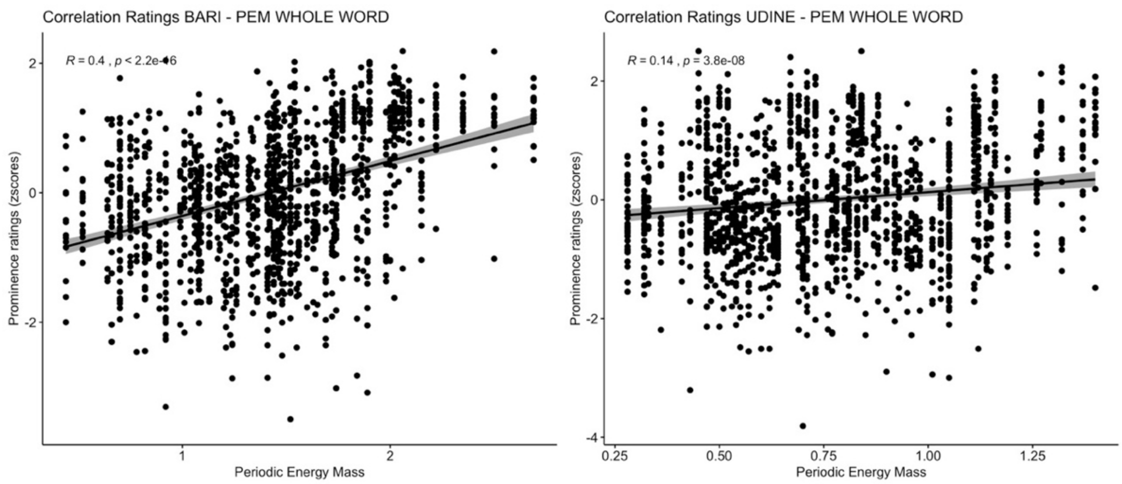


Figure 70. Correlations between PEM and prominence ratings. On the left Bari, on the right Udine. Results are pooled across conditions and word position.

For the target, the expectations where that the correlations would not be significant for either of the groups, since on the one hand, Bari data showed no distinction in ratings between BF and PF where the distinction in PEM was present, and on the other, the ratings of BF and PF, showed a differentiation in the Udine group, which was not present in PEM. Figure 71 shows the results of the correlation between prominence ratings and PEM for the target word. The correlation reached significance for both groups (Bari: r (453) = 0.34, p < 0.0001; Udine: r (478) = 0.1, p = 0.01). Results of the Fisher's r-to-Z transformation revealed a significant difference between the two groups, with the correlation for the Bari group being higher (z = -3.9, p = 0.0001).

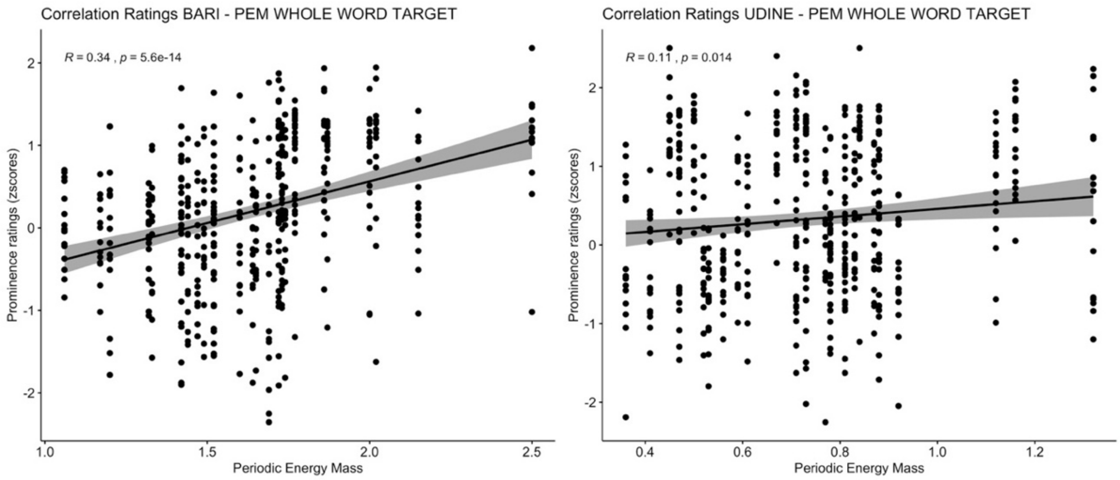


Figure 71. Correlations between PEM and prominence ratings for the target. On the left Bari, on the right Udine. Results are pooled across conditions.

This was contrary to the expectations, but might have been the result of the narrow focus. Therefore, Figure 72 shows the correlation of values considering only BF and PF. Results show that while the correlation is significant for the Udine group, it is not significant for the Bari group (Udine: r (318) = 0.18, p = 0.001, Bari: r (312) = 0.04, p = 0.44). This shows that despite the more similar distribution between PEM values in BF and PF, for the Udine group to the higher values in the distribution correspond indeed higher ratings, while this is not the case for the Bari group.

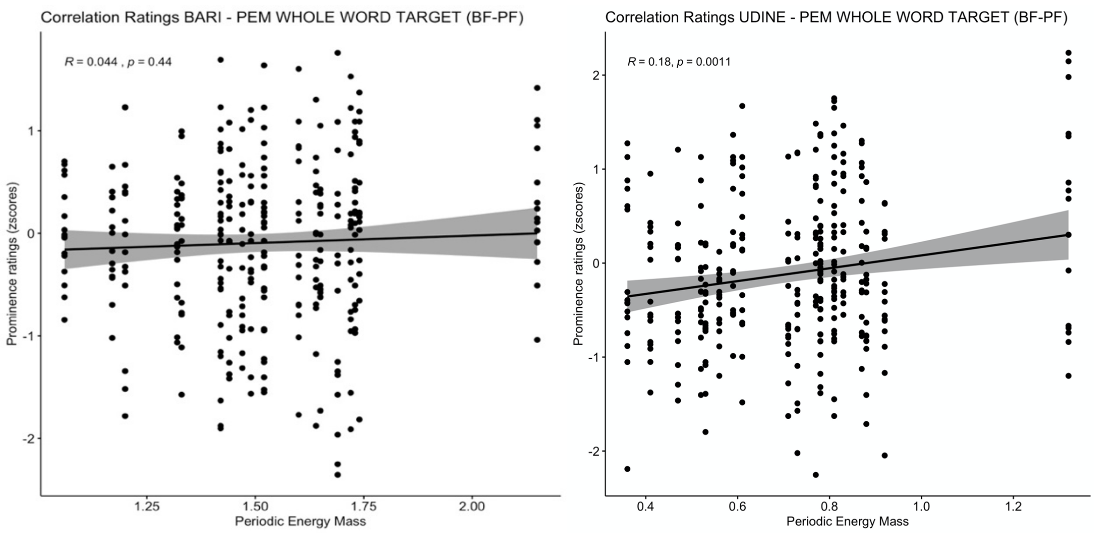


Figure 72. Correlations between PEM and prominence ratings for the target. On the left Bari, on the right Udine. Results are relative to broad focus and post-focal conditions.

The presence of a correlation between PEM and prominence ratings in the noun in PP was tested for both groups. Figure 73 shows the results of the correlation. For both groups, the correlation is not significant (Bari: r (453) = -0.01, p = 0.83; Udine: r (478) = -0.09, p = 0.06). Results of the Fisher's r-to-Z transformation for the correlation within the noun in PP in the two groups show that there is no difference between the two groups (z = -1.23, p = 0.22).

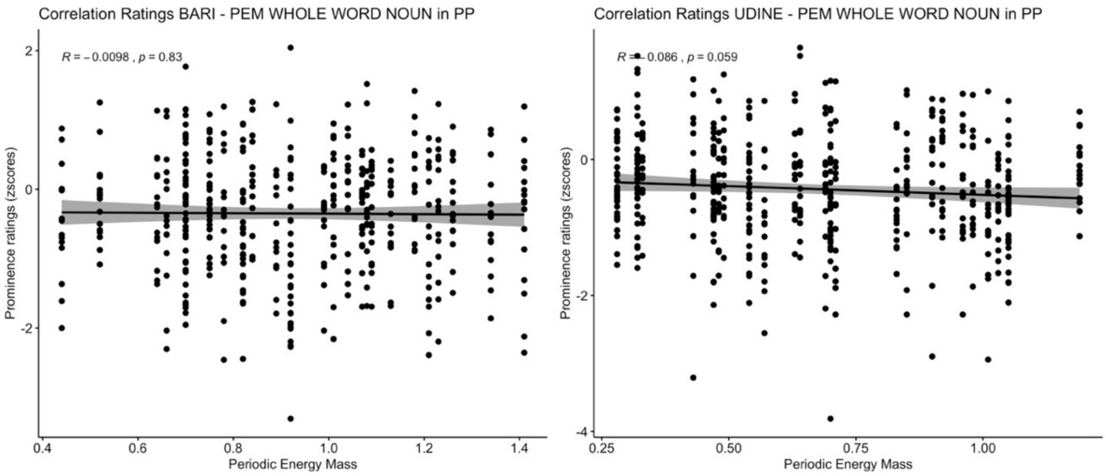


Figure 73. Correlations between PEM and prominence ratings for the noun in PP. On the left Bari, on the right Udine. Results are pooled across conditions.

Figure 74 shows the results of the correlations between scaling and prominence ratings across the word and between synchrony and prominence ratings across the word for the group of Bari. The correlation of prominence ratings with scaling was significant (scaling: r (1411) = 0.33, p < 0.0001), while the correlation with synchrony was not (synchrony: r (1411) = -0.01, p = 0.65). This was expected given the smaller variation of synchrony values between conditions and word position. For the Udine group (Figure 75) a significant correlation was found both for scaling and synchrony (scaling: r (1438) = 0.45, p < 0.0001; synchrony: r (1438) = 0.17, p < 0.0001). Results of the Fisher's r-to-Z transformation for the difference in the correlation between the two groups revealed significant differences, showing for the Udine group a higher correlation between F0 modulation and prominence ratings (scaling: z = -3.79, p < 0.001; synchrony: z = -4.85, p < 0.0001).

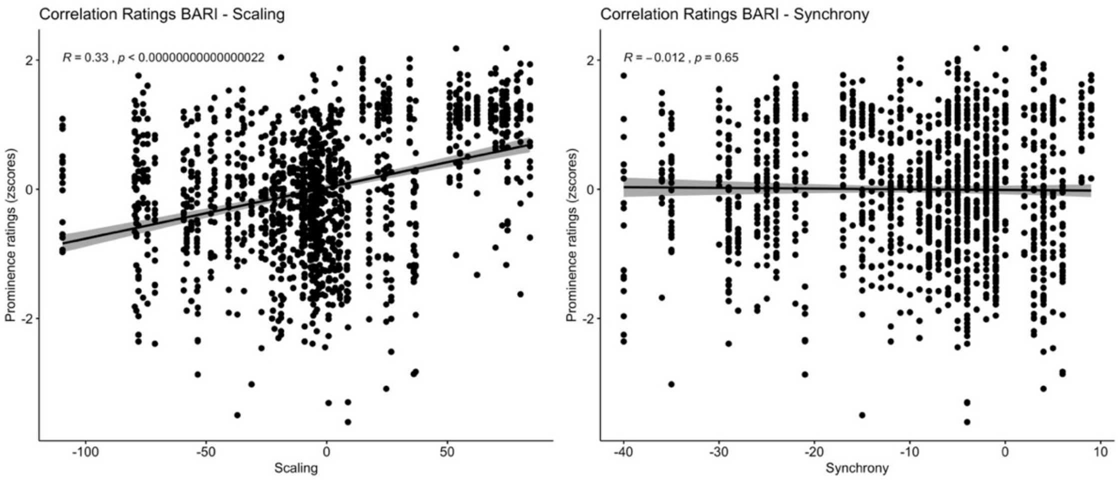


Figure 74. Left panel: Correlations between scaling and prominence ratings. Right panel: correlation between synchrony and prominence ratings. Results are relative to the stimuli and ratings of Bari. Results are pooled across conditions and word position.

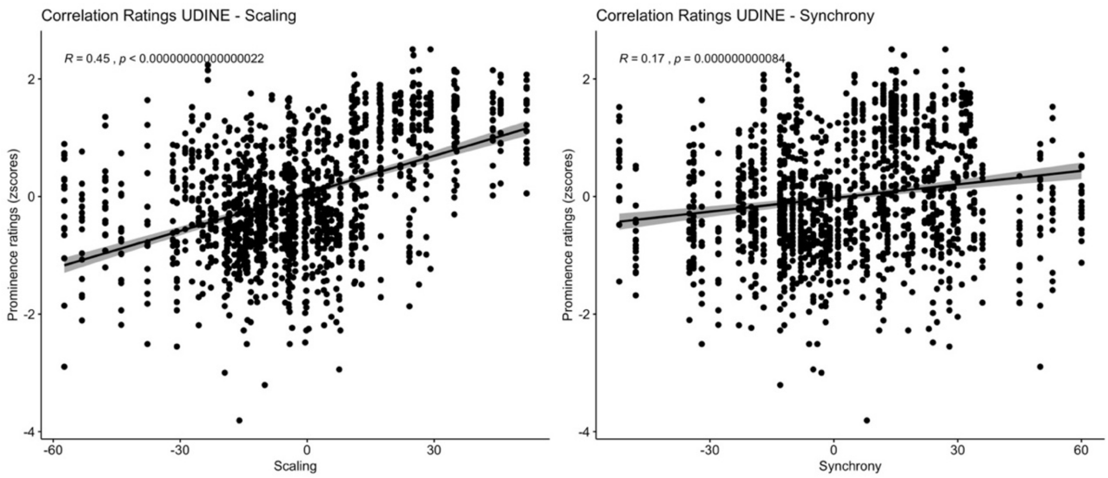


Figure 75. Left panel: Correlations between scaling and prominence ratings. Right panel: correlation between synchrony and prominence ratings. Results are relative to the stimuli and ratings of Udine. Results are pooled across conditions and word position.

6.5.2 Interim summary of the results: Bari and Udine

Firstly, the comparison between the ratings from Bari and Udine participants revealed characteristics of the perception of prominence that can be generalised across the two groups. Results show that in both varieties, the words presenting the wider expansion in pitch and the highest values of PEM (namely, the verb in PF and the target in NF) were perceived as most prominent compared to the other two critical words occurring in the same utterance (target and noun in PP in PF condition, both occurring in post-focal position; verb and noun in PP in NF condition, occurring in pre-focal and in post-focal position respectively). Secondly, these acoustically very prominent words (verb in PF condition and target in NF condition) were also perceived as more prominent than the words in the same position but occurring in different conditions and presenting lower acoustic prominence (verb in PF higher than verb in BF and verb in NF; target in NF higher than target in BF and target in PF). These results were generally expected and provided additional evidence to studies conducted on other languages, showing that listeners are sensitive to the variation that speakers make to convey prominence and that gradient prominence distinctions are perceived (i.e., rising pitch is perceived as more prominent than falling pitch, and pitch showing more excursion is perceived as more prominent than the one showing less excursion, see for example Baumann & Röhr, 2015; Cole et al., 2019; Roettger et al., 2019; among others). Thirdly, for the target, the difference between the ratings for narrow focus and the ratings for broad focus was higher than the difference between the ratings for narrow focus and the ratings for post-focal position, which suggests a generally closer relation in the prominence perceived between the words in broad focus and in post-focal position, compared to the relation between two words in in the domain of focus (target in NF and target in BF; see the discussion in 6.7).

Interestingly, a difference between the two groups is registered in the ratings of the target word across conditions. While the group of Udine makes a three-way differentiation between ratings of the target (NF rated higher than BF, in turn rated higher than PF), the group of Bari makes only a two-way differentiation of the conditions, rating NF higher than BF, but showing the same distribution of ratings for BF and PF. The differentiation between target in BF and target in PF was therefore higher for the Udine group. This is surprising because the distribution of PEM in the target of the Udine stimuli was similar between BF and PF, while in the Bari stimuli it was higher for BF than PF. In addition, the modulation of F0 in BF presented more excursion for the target in these latter stimuli than in the former. This higher value of prominence conferred to the target in PF, is hypothesised to be driven by the specific expectations of prominence that the Bari group has for the post-focal position and the differences in these expectations with the Udine group. The discussion around this result will be deepened in 6.7.

Additional evidence to differences in the perception of prominence in the two groups is presented by the results of the correlations. Considering all words in all conditions, the correlation between PEM and ratings is significant for both groups, showing to be higher for the Bari group. This result is explained by the similar distribution of PEM values across conditions in the Udine stimuli. By contrast, considering only the target in BF and in PF, the correlation between PEM values and rating values was significant only in the Udine group. This points to the fact that, for Udine listeners, higher values of PEM corresponded to higher values of perceived prominence. For Bari listeners, this is not the case and PF is generally rated as having a similar prominence compared to BF.

Additional results from the correlations between ratings and synchrony show a higher correlation between F0 modulation and ratings for the Udine group compared to the Bari group. This result confirms the higher tendency for the Udine group to rate words with falling pitch (i.e., negative values of synchrony and scaling) as less prominent than words featuring a rising contour (i.e., positive values of synchrony and scaling) and to rate words presenting no pitch movement (i.e., values of synchrony ad scaling around zero) as lower in prominence.

6.5.3 Results: Germans and Italians

6.5.3.1 Germans and BI

Results of all the critical words (target, verb and noun in PP) in all conditions and for both groups (Bari and Germans) are illustrated in Figure 76 and Figure 77. Figure 76 shows the mean of ratings in all the words per condition, whereas Figure 77 shows their distribution. A visual inspection of the graph in Figure 76 shows a higher differentiation between the mean in PF and BF in the group of Germans compared to the group of Bari listeners. To test for differences between the two groups, mixed analyses were conducted as described in 6.5.1). Firstly, the full model was run. A visual inspection of the residual plot of the full model did not reveal any deviations from homoscedasticity or normality. Results of the type II Wald test on the full model registered an interaction of CONDITION, POSITION and GROUP [$\chi2(4) = 16.18$, $p= 0.003$]. Figure 78 shows the pairwise comparison of the effects extracted from the model. The tendency that can be visually observed in the graph is that the German listeners perceived a different level of prominence for the target word in post-focal position compared to the one in the broad focus condition, whereas in the group of Bari listeners the prominence level of the target in the two conditions overlaps. Further inspection of the model showed that the group of learners rated the target in broad focus with significantly higher values than in post-focal position (β = 0.28 ± 0.12, p = 0.03) but, while the effect of this contrast is lowered for the group of Bari, the difference between the two groups is not significant (β = -0.26 ± 0.18, p = 0.14). In contrast, for both groups the target word in narrow focus was rated higher than the target word in broad focus (β = 1.24 ± 0.10, p < 0.0001).

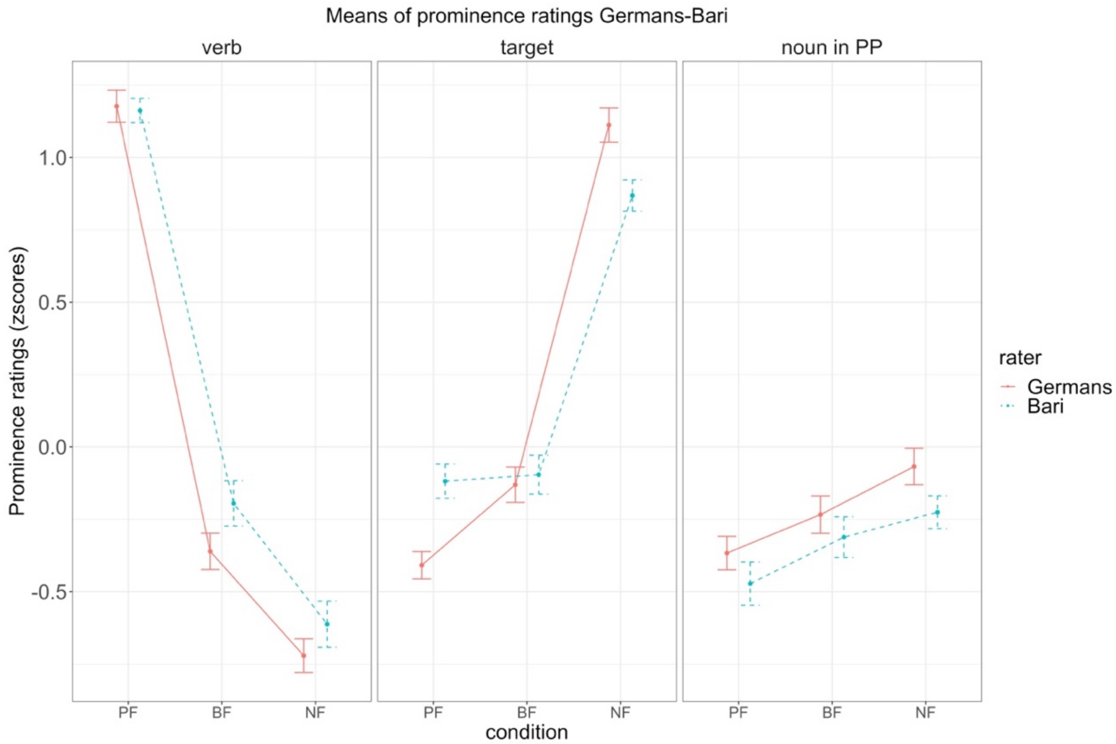


Figure 76. Means and standard errors of the ratings for each position of the word (from the left verb, target and noun in PP) in each condition (from the left post-focal, broad focus and narrow focus, coded as PF, BF and NF respectively). Dashed lines indicate ratings for Italians, solid lines for Germans.

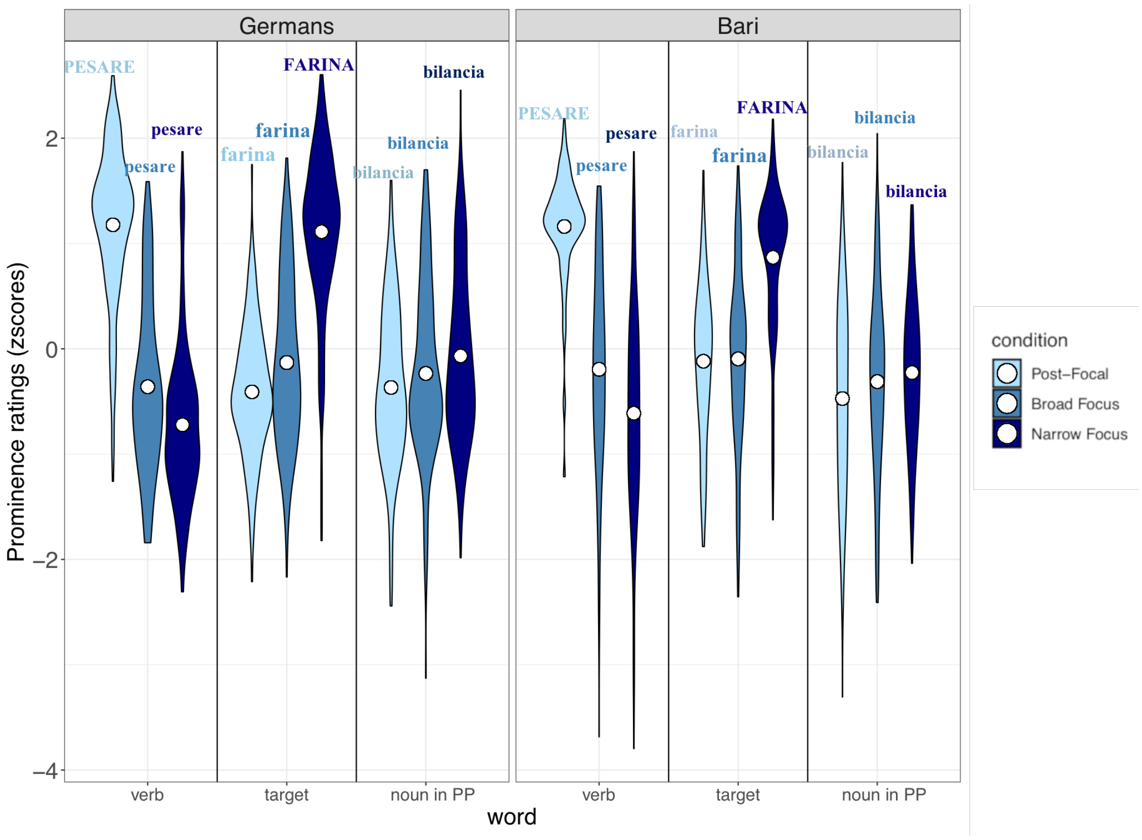


Figure 77. Prominence ratings for the three critical words by the two groups of participants (German and Bari). The y-axis depicts prominence scores (z-scored transformed). The x-axis shows the independent variable of the position of the word in the sentence. In both panels, verb, target word and final noun are plotted from left to right as well as values for each word group: post-focal condition (PF), broad focus (BF) and narrow focus (NF). Words in contrastive narrow focus are indicated by capital letters. The left panel illustrates the ratings by German learners, the right panel, the ratings by the Italians.

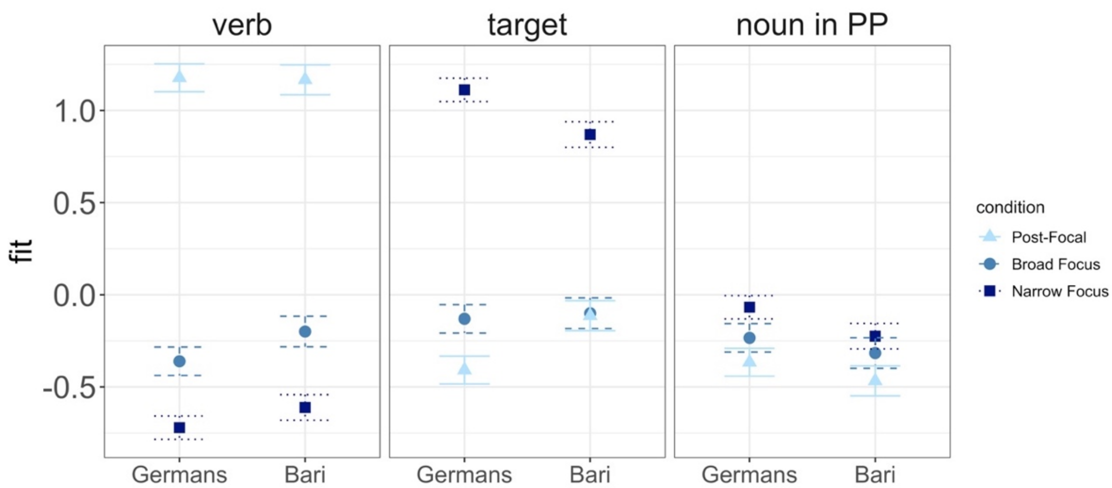


Figure 78. Means and standard errors for condition effects for each position of the word (from the left verb, target and noun in PP) for the two groups of participants (Germans on the left of each panel, Italians on the right of each panel). Triangles and solid lines indicate post-focal condition, circles and dashed lines indicate broad focus condition, squares and dotted lines indicate narrow focus condition.

Since a specific hypothesis about the ratings of the target within each group was formulated, two separate models, one for each group, were fitted to test this hypothesis. The model for the group of learners revealed a significant effect of CONDITION on the target, with the level post-focal significantly lowering the ratings' values compared to the level broad focus ($\beta = -0.28 \pm 0.10$, $p = 0.01$), as already shown in the full model. The model for the group of Bari listeners revealed that the slightly decreasing values of the ratings for the target in post-focal position compared to broad focus is not significant ($\beta = -0.01 \pm 0.15$, $p = 0.93$), whereas for both groups the target word in narrow focus was rated higher than in broad focus ($\beta = 1.24 \pm 0.09$, $p < 0.0001$ for Germans; $\beta = 0.97 \pm 0.11$, $p < 0.0001$ for Italians).

The position of the word in the sentence had a generally significant effect on participants' evaluations [$\chi2(2) = 201.17$, $p < 0.0001$]. In both groups the target in narrow focus was perceived as significantly more prominent than the other two words (verb: $\beta = -1.83 \pm 0.09$, $p < 0.0001$, noun: $\beta = -1.18 \pm 0.18$, $p < 0.0001$). In post-focal position, where the verb bears the contrastive accent, it was perceived as higher in prominence than the target word (verb: $\beta = 1.59 \pm 0.1$, $p < 0.0001$). The ratings for the noun in PP did not significantly differ from the ones of the target in the German group (noun in PP: $\beta = -0.01 \pm 0.08$, $p = 0.21$), while they differed in the Bari group (noun in PP: $\beta = -0.22 \pm 0.09$, $p = 0.02$). In broad focus, the ratings for the verb were lower than the target (verb: $\beta = -0.23 \pm 0.10$, $p = 0.04$) and no difference was registered between target and final noun (noun in PP: $\beta = -0.10 \pm 0.14$, $p = 0.48$). The variable group had no effect in these comparisons.

Figure 79 displays the means of the differences between conditions for each participant. As expected, there are many individual differences, mostly between broad focus and post-focal ratings. For both groups of participants, the narrow focus condition was almost always perceived with a higher level of prominence than the broad focus condition. On the contrary, in the comparison between broad focus and post-focal, there are differences between the two groups. For the German group, the majority of participants rated broad focus higher than post-focal. Just two participants rated post-focal higher than broad focus and six participants made very little differences. By contrast, the Bari group shows a higher tendency to rate post-focal target words more prominent than broad focus. Bari listeners generally demonstrate more variability than Germans. In particular, they not only show high values of ratings for post-focal compared to broad focus, but also compared to narrow focus. In fact, while for Germans the difference between narrow focus and post-focal is always positive and shows values always higher than 1, except for one subject, six native Italian speakers exhibit a difference lower than 1 point, with one subject even rating post-focal higher than narrow focus.

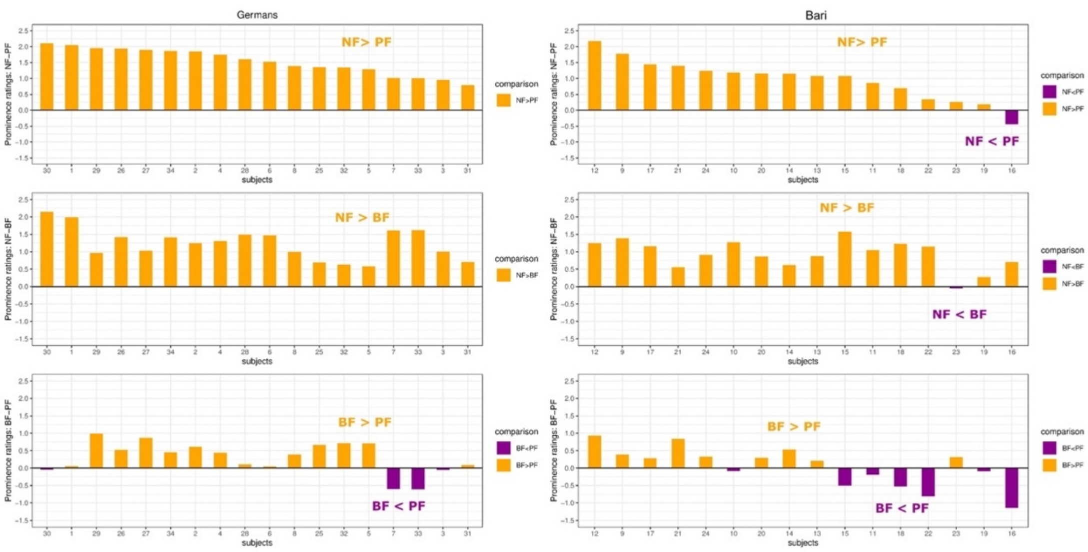


Figure 79. Differences of the prominence ratings of the target word between conditions. Values for every subject of the two groups of participants. On the left Germans, on the right native speakers. On the y-axis mean of the difference of the prominence scores (z-scored transformed). On the x-axis subjects. In both panels: on the top, mean of the difference between the scores for narrow focus (NF) and the scores for post-focal (PF), in the middle, mean of the difference between the scores for narrow focus (NF) and broad focus (BF) and on the bottom, mean of the difference between the scores for broad focus (BF) and post-focal (PF). In the graphs on the top, positive values indicate that the rating for NF are higher than the ones for PF. In the graphs on the middle positive values indicate that the rating for NF are higher than the ones for BF. In the graphs on the bottom, positive values indicate that the ratings for BF are higher than the ones for PF.

Figure 80 shows for the two groups the results of the correlations between PEM and prominence ratings across word position. For the correlation with PEM values across the word, both groups presented a significant correlation (Germans: $r (974) = 0.53$, $p < 0.0001$; Bari: $r (1363) = 0.4$, $p < 0.0001$). Results of the Fisher's r-to-Z transformation for the difference in the correlation between the two groups revealed a significant difference, with the Germans showing a higher correlation in comparison to Bari listeners ($z = 3.17$, $p = 0.002$).

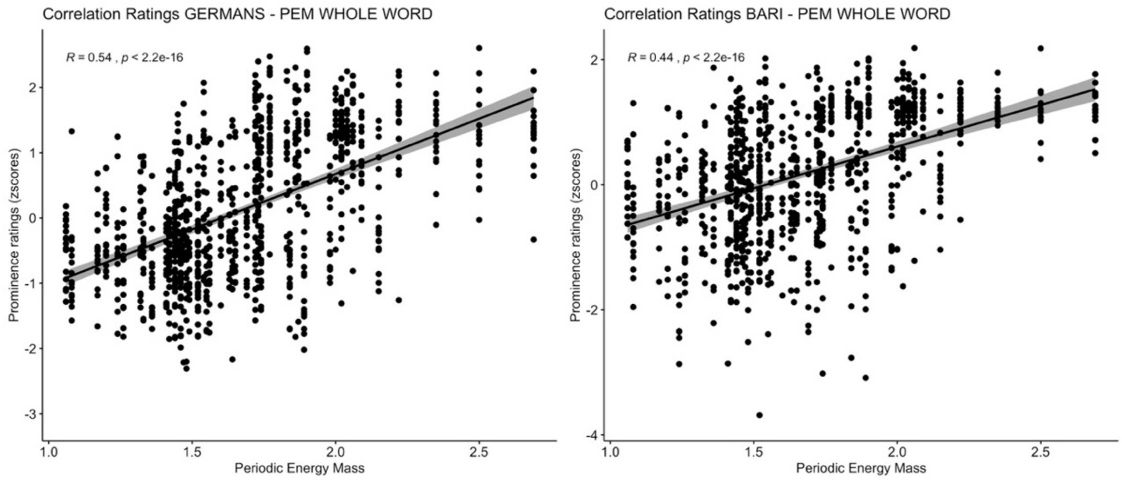


Figure 80. Correlations between PEM and prominence ratings. On the left German, on the right Bari. Results are pooled across conditions and word position.

Given that the two models for prominence ratings computed within groups yielded significant results for the group of Germans and not for the Bari group, for each group the correlations between PEM and prominence ratings of the target were tested. Figure 81 shows the results of this correlation for both groups, which is significant (Germans: $r(486) = 0.48$, $p < 0.0001$; Bari: $r(453) = 0.34$, $p < 0.0001$). Results of the Fisher's r-to-Z transformation revealed a significant difference between the two groups, registering a higher correlation for Germans in comparison to the Bari group ($z = 2.27$, $p = 0.02$).

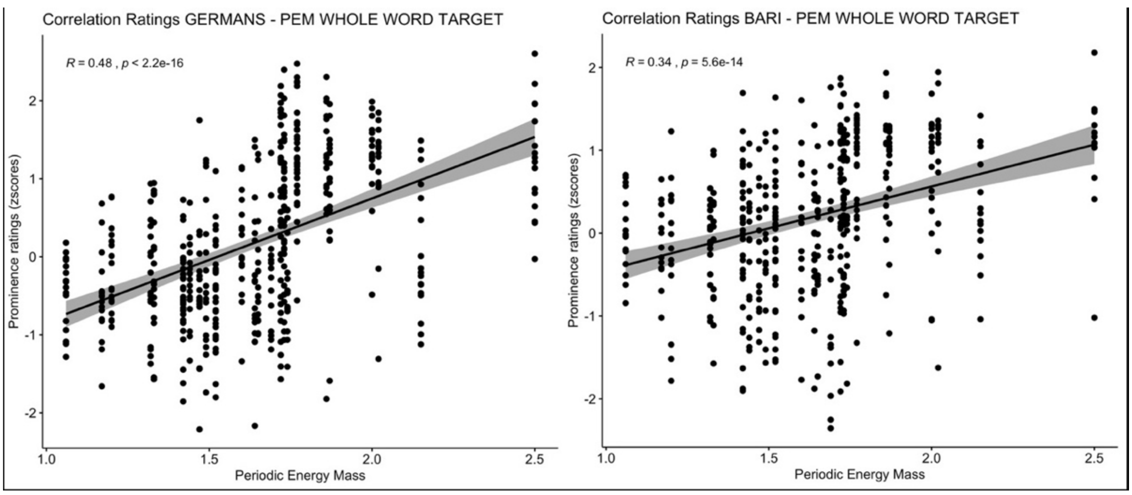


Figure 81. Correlations between PEM and prominence ratings for the target. On the left Germans, on the right Bari. Results are pooled across conditions.

In addition, while considering only BF and PF condition (Figure 82) results show that while the correlation is significant for the German group, it is not significant for the Bari group (Germans: $r(334) = 0.15$, $p = 0.005$, Bari: $r(312) = 0.04$, $p = 0.44$).

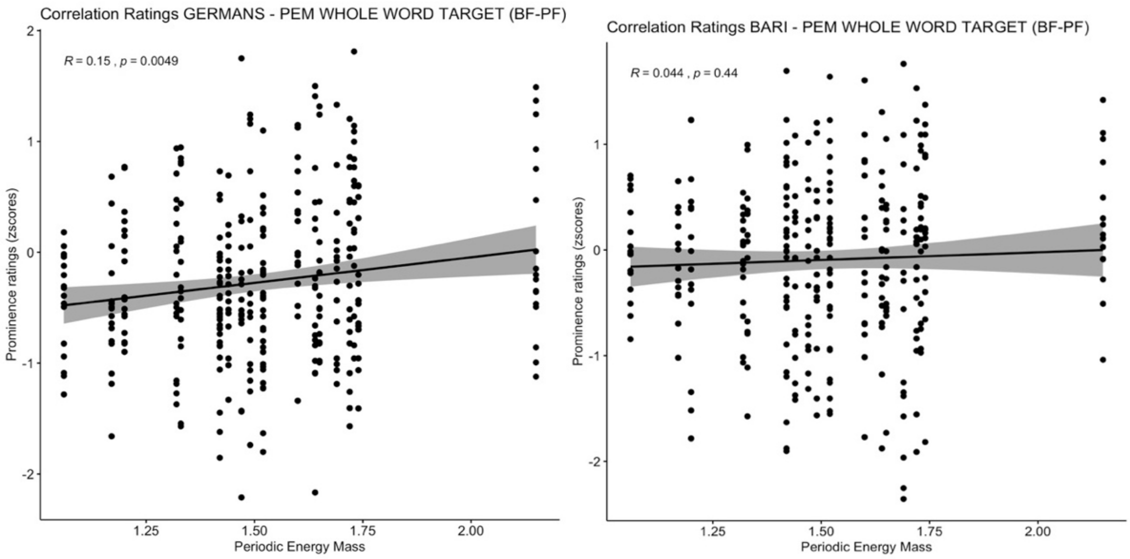


Figure 82. Correlations between PEM and prominence ratings for the target. On the left Germans, on the right Bari. Results are relative to broad focus and post-focal conditions.

Results of the previous model testing the variable word position showed that the noun in PP in broad focus is not differently rated in comparison to the target for the group of Germans. By contrast, the difference was present for the group of Bari. Therefore, the presence of a correlation between PEM and prominence ratings in the noun in PP was tested. Figure 83 shows the results of this correlation for both groups. For both groups, the correlation is not significant (Germans: r (486) = -0.02, p = 0.72; Bari: r (453) = -0.01, p = 0.83). Results of the Fisher's r-to-Z transformation showed the lack of difference between the two groups (z = -0.26, p = 0.79).

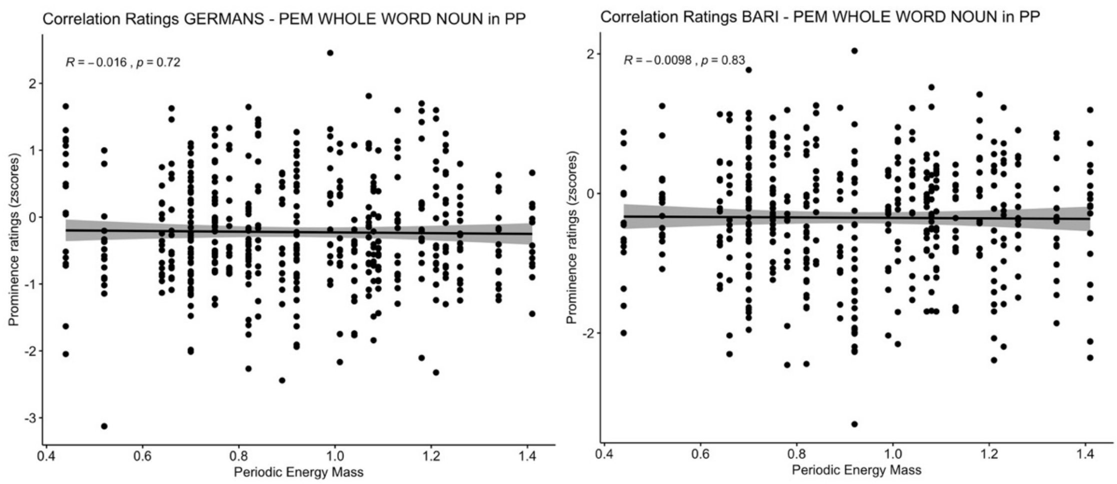


Figure 83. Correlations between PEM and prominence ratings for the noun in PP. On the left Germans, on the right Bari. Results are pooled across conditions.

Figure 84 shows for the group of Germans the results of the correlations between scaling and prominence ratings across the word and between synchrony and prominence ratings across the word. The correlation of the prominence ratings with scaling was significant (r (1510) = 0.39, p < 0.0001), while the correlation with synchrony was not (r (1510) = 0.03, p = 0.19). This was expected given the smaller variation of these value between conditions and word position. As already shown in Figure 74 in 6.2.2, the correlations for the Bari group (Figure 84) showed the same pattern registered for Germans (scaling: r (1411) = 0.33, p < 0.0001; synchrony: r (1411) = -0.01, p = 0.65). Results of the Fisher's r-to-Z transformation for the difference in the correlation between the two groups revealed no significant difference (for scaling: z = -1.86, p = 0.06; for synchrony: z = -1.07, p = 0.28).

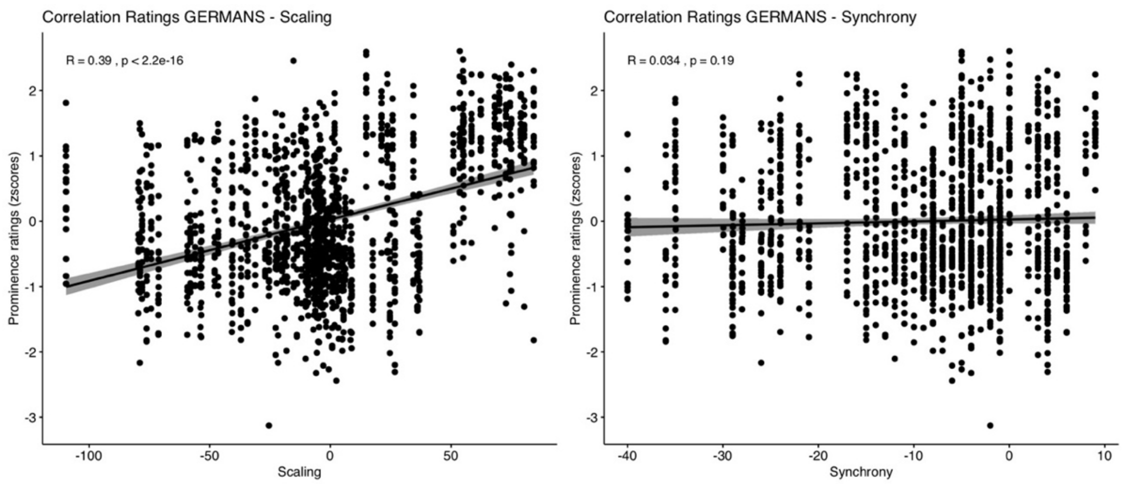


Figure 84. Left panel: Correlations between scaling and prominence ratings. Right panel: correlation between synchrony and prominence ratings. Results are relative to the stimuli and ratings of German. Results are pooled across conditions and word position.

6.5.3.2 Germans and UI

Figure 85 and Figure 86 show the results of all the critical words (target, verb and noun in PP) in all conditions, for both groups (G and UI). While Figure 85 shows the mean ratings (in all the words per condition), Figure 86 allows us to examine their distribution.

Visually inspecting the graphs (Figure 85 and Figure 86) a tendency to perceive the prominence of the target word in broad focus with a higher level of prominence than the target word in post-focal position can be recognised for the two groups. Indeed, the type II Wald test on the full model registered no interaction of CONDITION, POSITION and GROUP [$\chi 2(4) = 7.03$, p= 0.13]. A further inspection of the model showed that the difference between the ratings of the target word in broad focus are higher than the ratings of the target word in post-focal position ($\beta = 0.41 \pm 0.09$, $p < 0.0001$) and that the level GROUP does not have an effect on this difference ($\beta = -0.07 \pm 0.13$, $p = 0.6$). Figure 87 shows the pairwise comparison of the effects extracted from the model.

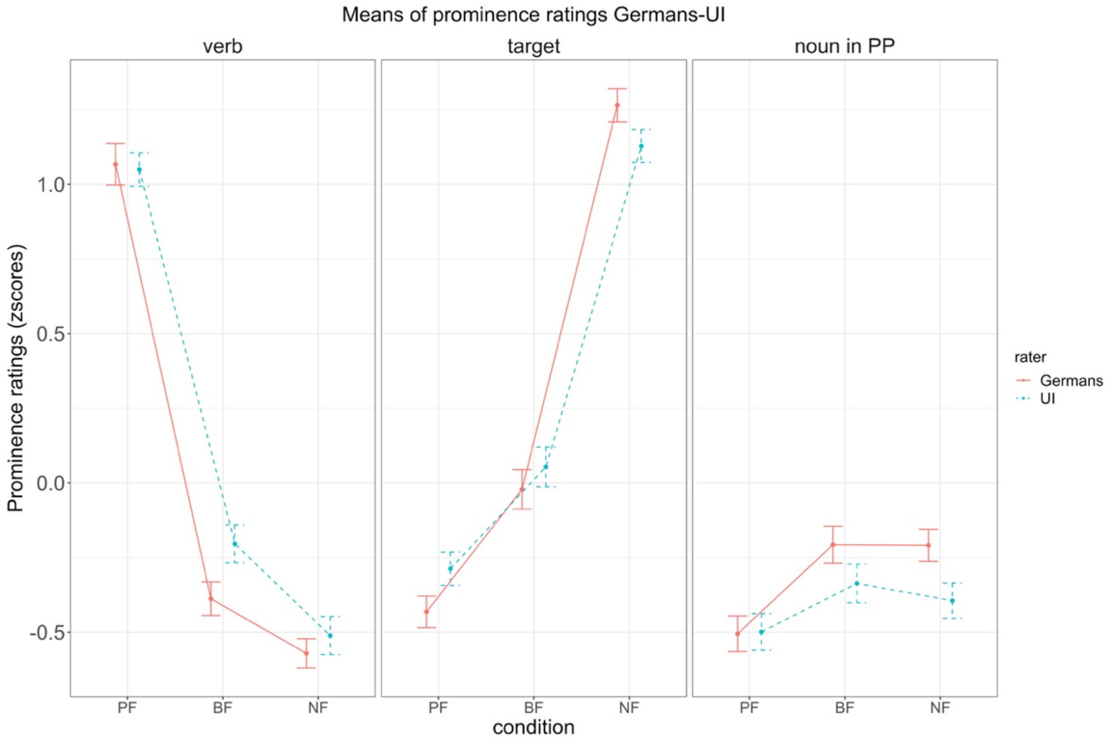


Figure 85. Means of prominence ratings for Germans and UI for each word in each condition. Solid lines indicate the results for Germans, dotted lines the results for Udine. The different panels show from left to right verb target and noun in PP. On each panel, the x-axis shows from left to right PF, BF and NF.

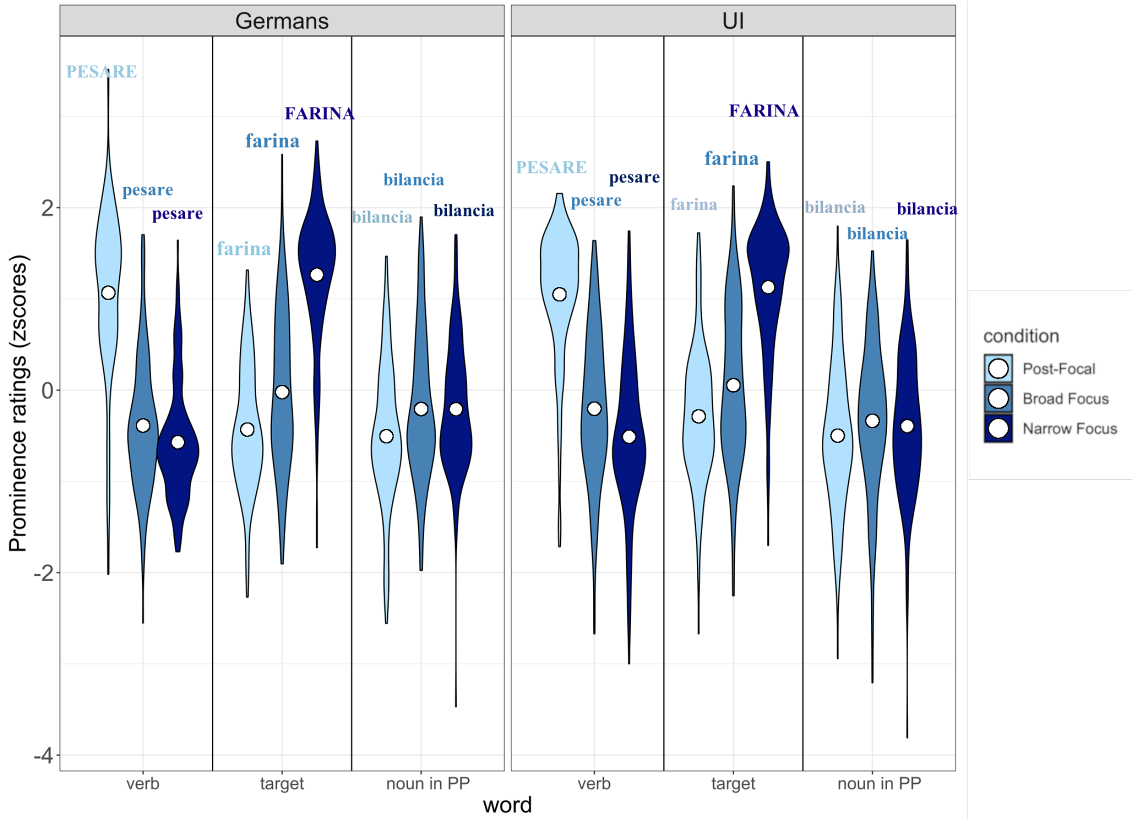


Figure 86. Prominence ratings for the target word by the two groups of participants: Germans on the left, UI on the right. Each panel within the panel group shows values for the words in one position in the utterance, from the left, verb, target and noun in PP. The y-axis shows prominence scores (z-scored transformed). The x-axis shows the independent variable of the condition. In both panels: from left to right ratings for the post-focal condition (PF), broad focus (BF) and narrow focus (NF). Words in contrastive narrow focus are indicated by capital letters.

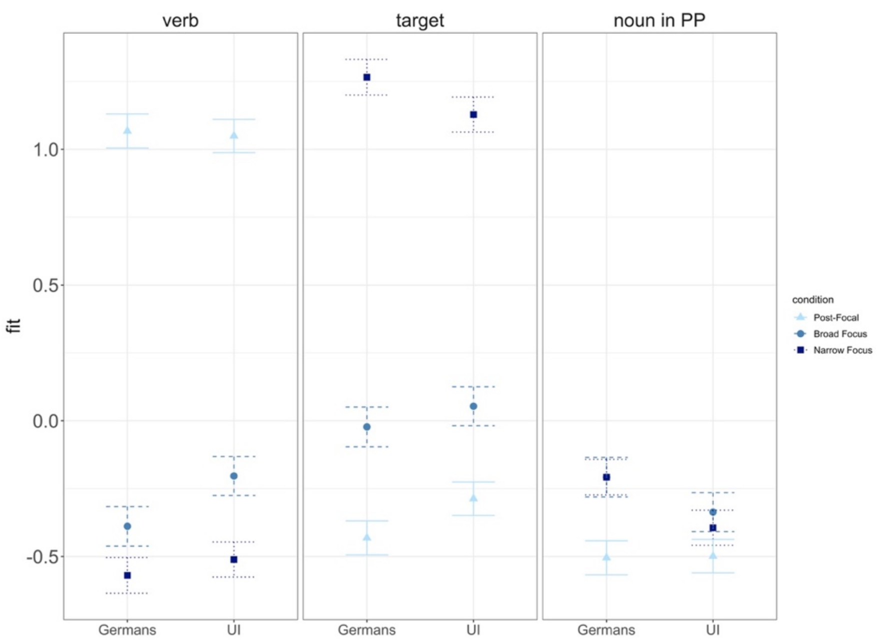


Figure 87. Effects registered by the full model. Error bars represent 83 % confidence intervals. Solid lines, dashed lines and dotted lines indicate error bars for PF, BF and NF respectively. Triangles, circles and squares indicate the means of PF, BF and NF respectively.

Figure 88 displays the means of the differences between conditions for each participant for the target word. For both groups of participants, the narrow focus condition was always perceived with a higher level of prominence than the broad focus condition. In addition, in the comparison between broad focus and post-focal, there are almost no differences between the two. For both groups, the majority of participants rated broad focus higher than post-focal. In both groups only one participant rated the two conditions with very few differences and only four participants in the German group and three participants in the Udine group rated BF with a higher prominence than PF.

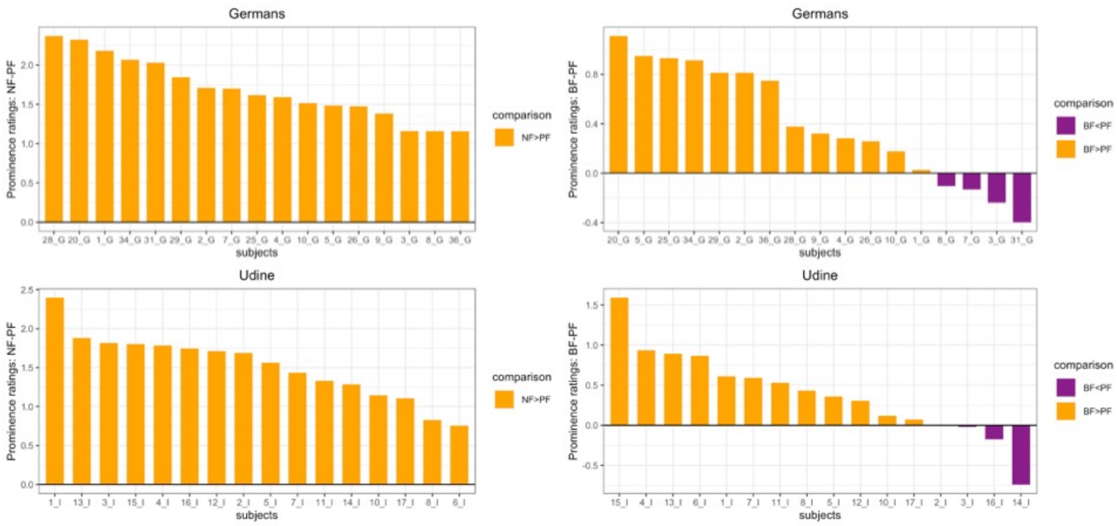


Figure 88. Differences of the prominence ratings of the target word between conditions for each subject of the two groups of participants. On the top Germans, on the bottom listeners from Udine. On the y-axis mean of the difference of the prominence scores (z-scored transformed). On the x-axis subjects. On the left, mean of the difference between the scores for narrow focus (NF) and the scores for post-focal (PF), on the right mean of the difference between the scores for broad focus (BF) and post-focal (PF). In the graphs on the left, positive values indicate that the rating for NF are higher than the ones for PF, in the graphs on the right, positive values indicate that the ratings for BF are higher than the ones for PF.

Figure 89 shows the results for the two groups (Germans and Udine) of the correlations between prominence ratings and PEM across word position. Both groups presented a significant correlation (Germans: $r (1510) = 0.14$, $p < 0.0001$; Udine: $r (1438) = 0.14$, $p < 0.0001$). The Fisher's r-to-Z transformation did not register a significant difference between the two groups ($z = -0.28$, $p = 0.78$).

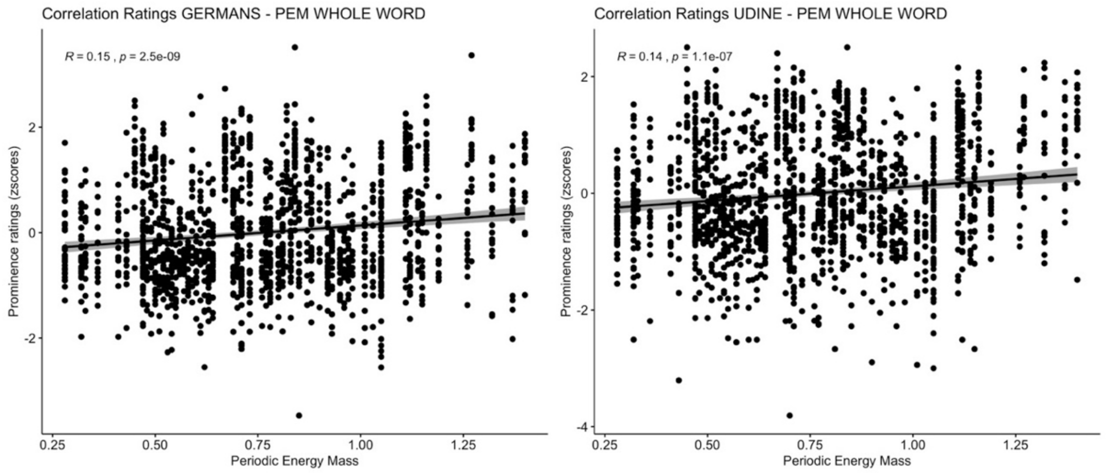


Figure 89. Correlations between PEM and prominence ratings. On the left German, on the right Udine. Results are pooled across conditions and word position.

As expected, no difference between groups was registered in the correlation of PEM and prominence ratings in the target word (Figure 90; z = 0.64, p = 0.5). For the target, the low correlation in both groups can be motivated by the fact that while in the Udine stimuli no difference is registered between BF and PF, the ratings of the group of Germans and the group of Udine show a three-way distinction between the conditions. A lack of correlation between PEM and prominence ratings for the noun in PP (Figure 91) is registered for both groups (Germans: r(502) = -0.03, p = 0.55; Udine: r(508) = -0.07, p = 0.12). This lack of correlation can be attributed to the rather high values assigned to the prominence in this position, which are associated to low values of PEM.

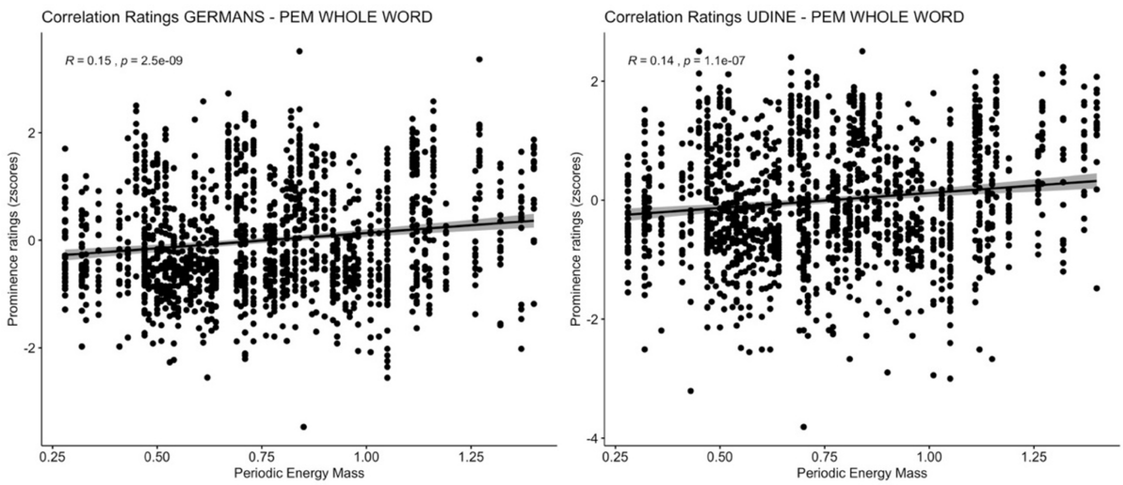


Figure 90. Correlations between PEM and prominence ratings for the target. On the left Germans, on the right Udine. Results are pooled across conditions.

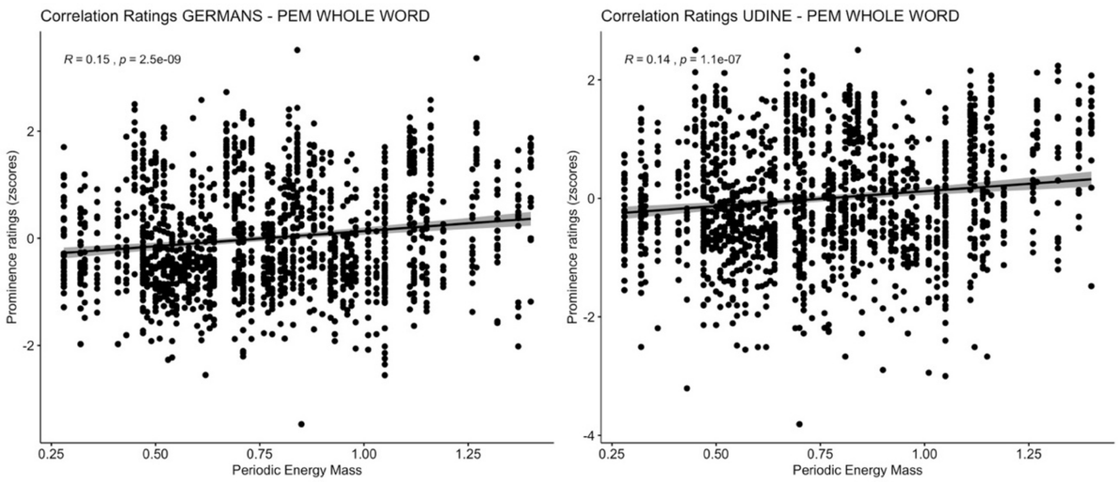


Figure 91. Correlations between PEM and prominence ratings for the noun in PP. On the left Germans, on the right Udine. Results are pooled across conditions.

Figure 92 shows for the group of Germans the results of the correlations between scaling and prominence ratings across the word and between synchrony and prominence ratings across the word. The correlation of the prominence ratings showed a significant result both for scaling and for synchrony (scaling: r (1510) = 0.46, $p < 0.0001$; synchrony: r (1510) = 0.14, $p < 0.0001$). Results of the Fisher's r-to-Z transformation for the difference in the correlation between the two groups (Germans and Udine; for Udine see Figure X in x.x.x.) revealed no significant difference (for scaling: z = 1.03, p = 0.3; for synchrony: z = -0.28, p = 0.78).

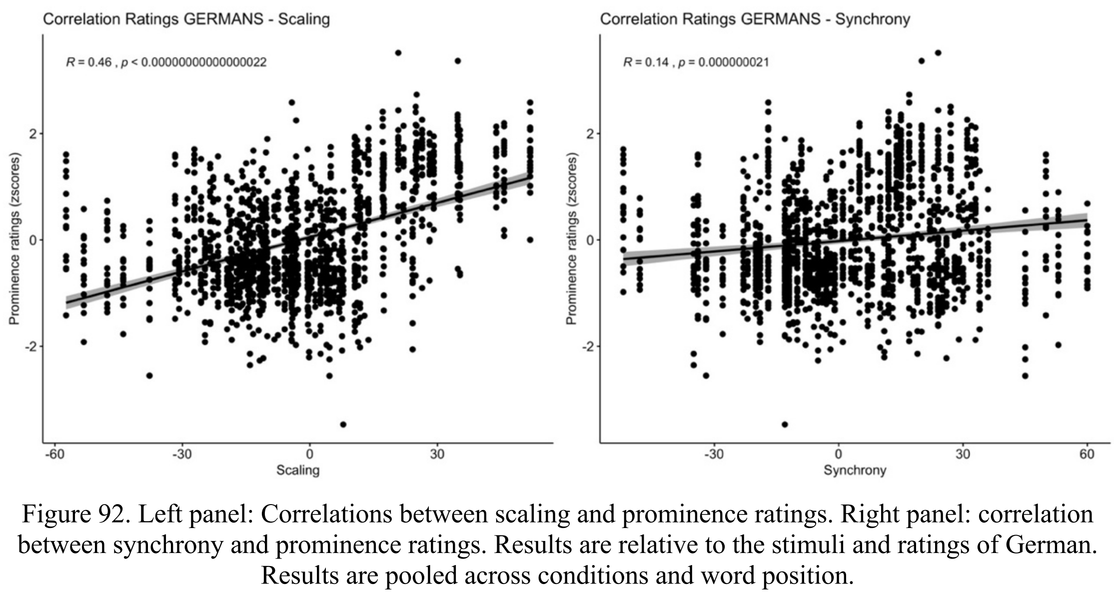


Figure 92. Left panel: Correlations between scaling and prominence ratings. Right panel: correlation between synchrony and prominence ratings. Results are relative to the stimuli and ratings of German. Results are pooled across conditions and word position.

6.5.3.3 Germans: BI and UI

The model run for the comparison of the ratings of Germans in the two varieties showed a significant interaction between the variables CONDITION, WORD and VARIETY [$\chi2(4) = 10.59$, $p = 0.03$]. A further exploration of the model revealed that this significant effect is engendered by a difference relation between the ratings of the verb in BF and NF for the two varieties, where the difference between verb in PF and verb in NF is perceived as higher when rating BI compared to UI ($\beta = 0.39 \pm 0.18$, $p = 0.04$). This difference can be re-conducted to the lower difference in PEM between the verb in PF and in NF in the UI stimuli. In addition, these stimuli also show less excursion on the verb in NF compared to the excursion found in the same condition for BI stimuli. Figure 93 shows the pairwise comparison of the effect extracted from the model. In order not to create redundancies, in this section only these results are presented as figures. The graphs displaying means and distributions of the ratings are retrievable from 6.5.3.1 and 6.5.3.2.

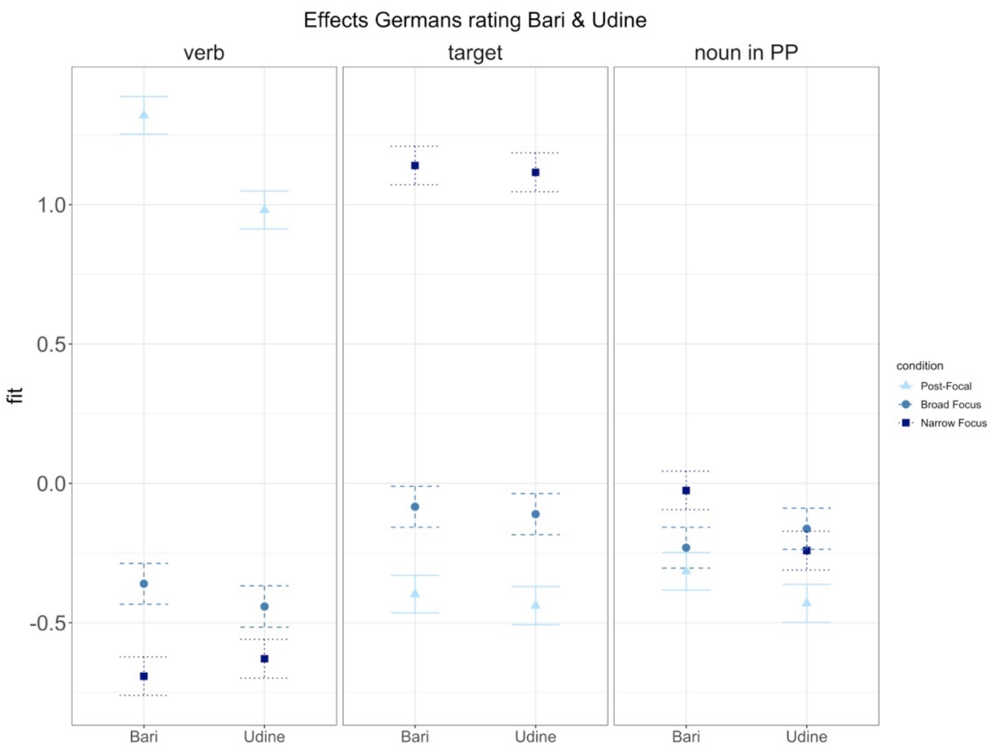


Figure 93. Effects registered by the full model. Error bars represent 83 % confidence intervals. Solid lines, dashed lines and dotted lines indicate error bars for PF, BF and NF respectively. Triangles, circles and squares indicate the means of PF, BF and NF respectively. Each panel corresponds to one critical word (from the left, verb, target and noun in PP). Within each panel results for the two varieties are displayed (on the left Bari, on the right, Udine).

6.5.4 Interim summary of the results: Germans and Italians

For German participants (G) a general tendency similar to Udine (UI) and Bari (BI) native speakers is observed. The verb in PF and target in NF are rated with higher prominence than the other critical words in the same utterance and are rated as the most prominent compared to the same critical words in the other conditions. This shows that, despite rating a non-native language, participants could identify the most striking general patterns and recognise which were the words that the speaker wanted to convey as most important for the message.

Looking at more subtle differences, an interesting pattern in the comparison between G and BI on the one hand, and G and UI on the other, is that these latter two groups (G and U) are more similar in the ratings of prominence compared to the former two groups (G and BI). This difference mostly consists in the ratings of the target, which show a three-way differentiation among the conditions for G and UI, while for BI only showing a distinction between NF and BF and not between BF and PF. This difference can be attributed to the more similar probabilistic distribution of prominences in the post-focal position between UI and G compared to BI. Indeed, BI seems to find more expectations for prominence after the focal constituent compared to the other two groups, resulting in higher values of perceived prominence for the target in PF.

Considering only the German group, no difference is registered between the ratings for one variety and the other for the target in post-focal position (target in PF). The comparison between target in PF and BF, reveals that this latter condition is rated as more prominent in both varieties. The lack of a difference between the two datasets in this position is interesting, when one considers the difference in the distribution of PEM for this position in the stimuli of the two varieties.

The next section will discuss in more detail the implications of the results found and will try to explain the results in light of the behaviour of all the different groups.

6.6 Discussion and conclusion

Results on the ratings of perceived prominence generally confirm that listeners follow both acoustic signal and structural expectations form their native language when rating the degree of prominence of words. In the former case, the strategy followed is bottom-up (driven by the stimulus), in the latter, the strategy used is top-down (driven by the native language and structural expectations derived from it).

Concerning the signal-driven contribution, this study has shown that a word bearing a pitch accent with high excursion and presenting high values of PEM (as the target word in NF and the verb in PF), is perceived as the most prominent word in the utterance both by native speakers of the Bari and the Udine varieties of Italian and by German learners of Italian. This result reveals that the intention of the speaker to convey a high level of prominence on a particular word (recall that the stimuli were recorded in the production study presented in Chapter 5), was successfully interpreted by both native and non-native listeners. On the perspective of non-native understanding of the prosodic marking of intended meaning, these results, even though not focussing on the retrieval of the context in which an utterance was produced, might indicate that when the focus is signalled by similar strategies, it is easily picked up by non-natives. A finding connected to this result was that for all groups the difference between the prominence ratings of a word that greatly and uniquely stands out of its utterance's context and the prominence ratings of a word in the wide focus domain is greater than the difference between this latter word and the word in post-focal position. This suggests that the change in the degree of prominence between a word that greatly and uniquely stands out of the utterance (i.e., narrow focussed words) and a word that occurs in broad focus domain is perceived as higher than the more categorical change between showing a pitch contour on the one hand (word in the broad focus domain) and being flat and low on the other (word in post-focal position).

Moreover, the present results provide additional evidence to previous studies attesting that the strong acoustic cues are more likely to be marked as prominent and produce a higher agreement in the assignment of prominence (in this case among different groups), suggesting that the perception of prominence given these cues is robust (Baumann & Winter, 2018; Cole et al., 2010a, b; Mahrt et al., 2012). Indeed, in the absence of such strong cues (in the target in broad focus and in the target in post-focal position) differences in top-down expectations are registered. These differences emerged between the two groups of Italian native speakers (Bari and Udine) and between the group of German learners and the group of Bari native speakers: Udine and German listeners assigned a higher degree of prominence to words (target) in broad focus compared to words (target) occurring in post-focal position, while ratings for Bari listeners were similarly distributed in the two conditions.

As far as the comparison between Bari and Udine is concerned, the direction in which the difference is realised, is particularly interesting considering the acoustic characteristics of the stimuli. Bari stimuli present considerable differences in duration and energy profile between broad focus and post-focal position, while in Udine stimuli these differences were not present. The differences presented by the Bari stimuli consisted of considerably lower distribution of values in the target in post-focal position compared to the target in broad focus. In addition, the target in post-focal position presented the least acoustic prominence within its utterance, suggesting that the perceived prominence should be markedly lower compared with the target in broad focus. Moreover, the target in the latter condition presented not only the highest values of PEM within its utterance, but also showed considerably wider excursion in pitch compared to the other two words (verb and noun in PP). By contrast, in the Udine stimuli, the target in BF showed a similar distribution of PEM values compared to verb and noun in PP. In addition, the F0 on verb and target showed similar modulations within-syllable and across-syllables, while the noun in PP was less compressed than the one realised in the Bari variety, where in turn the target showed very high pitch excursion compared to verb and noun in PP. These distinctions in the acoustic characteristics of the target in broad focus in the two varieties were expected to yield increased differentiation between the target's prominence ratings in the Bari group, not in the Udine group.

Ratings by Bari listeners were interpreted as being influenced by top-down expectations on the level of prominence in the post-focal position, derived by the distribution of prominences in questions in this position for this variety (see 2.3.3). Additional support for this interpretation is derived by the correlation between PEM and prominence ratings, which was absent for the Bari group in the ratings of prominence in BF and PF, while it was present in the Udine group. This shows that the latter group followed a more pitch-related bottom-up strategy. Participants of the Bari group, who did not expect a consistent lack of prominence in the post-focal position, did not rate the post-focal target as low in prominence. By contrast, both the Udine group and the group of German learners rated the post-focal position as low in prominence: they do not show the expectations on the degree of prominence shown by Bari listeners, since the distribution of prominences brought about by pitch movement in their variety (or language) is different.

However, an additional consideration to be made is the presence of the high phrase accent at the end of the target in broad focus in the Udine variety, which could have led to a perception of increased prominence. Yet, the presence of increased prominence in the target in broad focus for the Bari group should still lead to a high difference in the perception of prominence within the group of Bari and this difference could still be attributed to the role of expectations in this variety. This characteristic is confirmed by the comparison with the German group, whose participants, in line with Udine participants, made a higher differentiation between broad focus and post-focal position compared to Bari participants, rating the prominence of post-focal position as low in the stimuli of both varieties. In addition, they did not make a difference in the ratings of the two varieties, despite the differences found in the signal.

In rating the Bari variety, German learners either employed their native expectations of attenuation of post-focal position, or prioritised bottom-up inferences when judging the non-native language (or used both strategies). Either way, they were not biased by expectations for post-focal prominences, both because they do not have knowledge of the distributional properties of prominences in Bari Italian, and possibly also because this is how flat pitch and low energy would be rated in their native language (Röhr & Baumann, 2010). In the experiment with the Udine variety, the German learners either prioritized the F0 movement over the energy profile or employed their expectations of attenuation derived from their native language. Interestingly, the level of proficiency of learners had no effect on the difference in the perceived prominence of the word (target) occurring in broad focus and in post-focal position, suggesting that the ratings of prominence were not influenced by semantics, since different interpretation of prominence relations did not correspond to higher level of semantic knowledge (presumably attributed to learners of a higher level).

In addition, top-down expectations appear to contribute to some extent to the prominence assigned by Germans to the final word in the utterance (e.g., *bilancia*, scales). This word, structurally prominent, but acoustically weak, is assigned by German participants more prominence than expected from the bottom-up cues. Indeed, German participants assign to the noun in PP in broad focus condition values similar to the ones assigned to the target, which is acoustically more prominent. This stems from the fact that the noun in PP is both occurring in final position and bearing the nuclear accent, whose structural importance is recognised by listeners. Moreover, the final position seems to be considered particularly important by Germans, since, in rating the Bari variety, the word occurring in this position (noun in PP) in NF and PF condition (occurring post-focally) showed an enhanced level of perceived prominence. In the PF condition, the noun in PP is rated with similar values attributed to the preceding word (target), despite having lower values of PEM, while the noun in PP occurring in NF is rated higher than in BF. This latter result (of a word in post-focal position rated similarly to a word in broad focus) is possible because of the very reduced acoustic characteristics of the latter word. The lack of significance of the correlation between prominence ratings and PEM for this final word also confirmed that the acoustic signal is not followed in this position. In addition, when rating the Udine variety, Germans' rating of the noun in PP in BF and NF were similarly distributed, as expected from the similar distribution of PEM values in these two varieties. Interestingly, the noun in PP occurring in PF, although having a similar distribution of PEM values compared to the other two conditions, was rated with lower values. This could be due to reduced movement within and across the syllables for this condition compared to NF and BF. The absence of the correlation between prominence ratings and PEM was also present in both the Italian groups (BI and UI), with participants of both groups rating the noun in PP in the three conditions as similar in prominence, despite the noun in PP in BF condition being the nuclear accent. Taken together, these results, both of learners and native speakers of Italian, indicate that the position, rather than the acoustics, influences the ratings of this words and is in line with the findings of Cole et al. (2019).

Overall, the findings indicate that the perception of prominence may rely on the language-specific distribution of accents. While both German learners and Udine native speakers seem to interpret low flat pitch of post-focal targets with a low degree of prominence, the degree assigned by Bari listeners was higher. For Germans the explanation for the ratings on the Bari variety can be traced back both to the low acoustic prominence of the word (no movement in pitch and low values of PEM) and to the lack of recourse to distributional properties of prominences in the Bari variety, and, possibly, to the fact that flat pitch would be interpreted this way in their native language (Röhr & Baumann, 2010). This latter explanation is also supported by the fact that in the ratings of the Udine variety, learners interpret the flat contour of the target as indicating that the word is low in prominence, despite it showing a level of PEM similar to broad focus. In addition, Udine listeners interpret the flat pitch in their variety as low in prominence. This interpretation is also due to the fact that in the majority of the occurrences of post-focal material in this variety PEM values are distinguished from the ones characterising broad focus, as shown in the production study (Chapter 5). Thus, Udine listeners possibly relied to some extent also on their knowledge of the language and not only on acoustic features of the stimuli. The probabilistic distribution of energy, together with the lack of F0 dynamics, resulted in the low prominence ratings of the post-focal position. By contrast, Bari listeners rate low and flat F0 as similarly prominent to higher F0 dynamics, appearing to rely on their native language-driven expectations that in this case would point them to find cues for prominence in post-focal position. The explanation for this tendency could be found in the higher-level structures of the language that contribute to generate expectations regarding the production of upcoming words in the utterance.

Lastly, it appears that language-driven expectations would direct Bari Italian speakers to designate more attention to post-focal information than reported for other languages such as English or German. The next experiment presented in Chapter 7 will test the effects of this outcome on online processing and attention (re-)orienting.

Chapter 7

ERP study: online processing of prominence in post-focal position

7.1 Introduction

The present chapter is a longer and more elaborate version of the paper by Ventura, Grice, Savino, Kolev, Brilmayer and Schumacher (2020). The experiment reported is generally concerned with the influence of prosodic prominence on selective attention and the subsequent depth of semantic processing during on-line speech comprehension. More specifically, the first issue that the experiment aims to address deals with the role of fine-grained cues to prominence in orienting attention. In particular, this experiment investigates whether the cues to prominence found in the post-focal position of questions in the variety of Italian spoken in Bari can (re-)orient attention in this position. The second issue arises from the results that have been found in the rating study reported in the previous chapter (Chapter 6). These results show that ratings on the prominence of the post-focal words given by native speakers of the Bari variety are affected by their native language's probabilistic distribution of accents in post-focal position. They show that Bari listeners expect a rather high degree of prosodic prominence for post-focal words and might therefore pay particular attention to them. This experiment aims at further investigating this possibility and, at the same time, at assessing the role of expectation-based inferences derived from the probabilistic distribution of prosodic prominence on online processing. Thus, the aim of this experiment is to contribute to the understanding of how signal-based and expectation-based factors connected to prosodic prominence orient attention and how these two processes can be better disentangled.

To investigate the real-time correlates of prosody comprehension and its potential role in attentional processes an event-related brain potential (ERP) study was conducted. The rationale for the present experiment is based on a series of ERP studies which have shown that the prosodic marking of information status and structure influences the ease by which discourse information is processed online (Baumann & Schumacher, 2012; Heim & Alter, 2006; Toepel et al., 2007; Schumacher & Baumann, 2010, among others, see also 3.7). For example, Baumann and Schumacher (2012) suggest that the processing of the prosodic marking of entities not only has an influence on the detection of the mismatch between prosody and information status, but that its intrinsic function in discourse has an independent influence on processing. In their study they found an increase in the biphasic N400-Late Positivity modulation connected to deaccentuation. This result suggests an increase in the prediction error and an increase of the update of the mental model, both connected to the request of the search for the respective given entity in the discourse, a search that is inherently triggered by deaccentuation. Other studies (e.g. Magne et al., 2005; Röhr et al., 2020) have shown that an increase in prominence can trigger an Early Positivity (P300), the type of positivity found by Baumann and Schumacher (2012; see Sassenhagen et al., 2014 for the interpretation of the P600 as part of the P3 family), which can be connected with attentive speech comprehension. Results of these studies lead to consider prosody as an attentional orienting device, in that strong prosodic prominences can attract attentional resources (cf., Röhr et al., 2020:12; see below for further explanation).

The orienting of resources results in a deeper processing of the stimulus: attentive stimuli are processed more deeply compared to non-attentive stimuli (Luck & Kappenman, 2012). The processing depth is reflected in the ERP components' amplitude, which is enhanced for stimuli that have attentional resources devoted to them. Thus, attention orientation allows the relevant information to be processed deeper in comparison to non-relevant information. This mechanism is particularly useful in light of the so called "good enough processing" principle (Ferreira et al., 2002; Sanford, 2002). It assumes that the linguistic input is only partially analysed, given that the amount of attention involved in processing words or constituents is not homogeneously distributed across the entire linguistic input. While part of the input is deeply processed, another part is only processed in a shallow manner (i.e. overlooked, see 3.8.1). Prosody is considered to modulate the relation between shallow and deep processing across the utterance (e.g., Cole et al., 1978; Cutler & Foss, 1977; Dimitrova et al., 2012; Fraundorf et al., 2010; Li et al., 2018; Sanford et al., 2006; among others, see below). An interesting case to test the depth of the processing of the stimulus as guided by prosody is that of the semantic processing of incongruence. Indeed, the elicitation of N400 effects by semantic incongruences is well attested as a special case of prediction error (Brown & Hagoort, 1993; Kutas & Hillyard, 1980; Kutas & Federmeier, 2011; among others). This paradigm has been used by Wang et al. (2011) to test the idea that the prosodic marking of focussed information helps in the online processing of semantic incongruence and prevents the shallow processing of the stimulus. They investigated the processing of semantic (in)congruence in utterances that presented prosodic information that either matched with the previous context or created a mismatch with it. In particular, the previous context was composed by questions eliciting narrow focus on the target word or on a preceding word, causing the target to occur in background in this latter case (i.e. post-focal position). An example of the stimuli that they used in their experiment is given in (25)-(26), which present both semantically congruent and semantically incongruent words (the examples are a translation from Dutch, taken form Wang et al., 2011:815). The target word is in bold, while the accented words are reported in capital letters. Here we see that the same context preceded two prosodically different realised stimuli (a. and b. in the examples). The prosodic realisation of the target in (25a.) and (26a.) matched the previous context, in that it was either realised with accentuation of the target occurring in focal position (25a.) or with deaccentuation of the target occurring in background (26a.; and accentuation of the word *mum*). The utterances in (25b.) and (26b.) represented mismatching between prosodic realisation and previous context, in that they were either realised with the deaccentuation of the focus (25b.) or with the accentuation of the background (26b.).

(25) **Context:** What kind of vegetable did mum buy for dinner today?

a. Today mum bought **EGGPLANT** (congruent)/**BEEF** (incongruent) for dinner.

b. Today MUM bought **eggplant** (congruent)/**beef** (incongruent) for dinner.

(26) **Context:** Who bought the vegetable for dinner today?

a. Today MUM bought **eggplant** (congruent)/**beef** (incongruent) for dinner.

b. Today mum bought **EGGPLANT** (congruent)/**BEEF** (incongruent) for dinner.

Results reported by the authors show that when the semantic incongruence occurs in focal position and is accented (25a.), the effect on the N400 is larger than the one in all the other conditions. Therefore, while the incongruence is deeply processed in the first scenario (25a.), in the others it is shallowly processed. Particularly interesting for the experiment that will be reported below is that this effect was larger, not only compared to the mismatching conditions (prosody mismatching with information structure), but also compared to the effect elicited by semantically incongruent words that occur in the background and are deaccented. Accordingly, attentional resources are allocated to the accented focus of the utterance, which seems to draw attention away from the following part of the utterance (i.e., post-focal position), resulting in its shallower processing.

This study is of great interest, because it suggests that in Dutch (and other West-Germanic languages) attentional resources are not devoted to the utterance's background. However, it could not disentangle the role of expectations generated by the context from the role of the acoustic signal in the orienting of attentional resources. In fact, the study of Wang et al. (2011) is preceded by a large body of research pointing to the fact that focus alone increases the allocation of attentional resources, without the mediation of accentuation (Cutler & Fodor, 1979; Birch & Rayner, 1997; Wang et al., 2009; Ward & Sturt, 2007; among others). By contrast, Kristensen et al. (2012) and Li and Ren (2012) tried to disentangle the role of accentuation alone on the processing of incongruence, therefore analysing the contribution of signal-based processing only. Both studies investigated utterances in isolation, to prevent the generation of expectations derived from a previous context. Results of both experiments show that pitch accents orient attention, with a consequent ease in language comprehension. Kristensen et al. (2012) used the classical paradigm of the *Moses illusion* (where the target was anomalous on the basis of word knowledge, such as the word *Moses* in *How many animals of each kind did Moses take on the Ark?* See 3.8.1), comparing accented targets with non-accented targets. Their results show that the presence of pitch accents leads to a larger activation of the general attentional network in comparison to their absence, which interacts with the semantic/pragmatic processing of the linguistic input.

Results of Li & Ren's (2012) study on the processing of semantic incongruencies in prominent and attenuated targets, supports the findings by Kristensen et al. (2012). As mentioned before, contrary to Wang et al.'s experiment (2011), this study did not provide a preceding context for the critical utterances, allowing to investigate the effect of bottom-up cues on processing, as distinguished from top-down effects. In addition, they also considered the processing of two degrees of prominence, with the target presenting more excursion in pitch and increase in duration cuing enhanced prominence. Their stimuli consisted of Chinese utterances in which the target word was either semantically congruent or semantically incongruent with the sentential context and was realised in three different prosodic conditions. An example of the stimuli used in this experiment is reported in (27)-(29). Note that while the original examples are in Chinese, here only the

English translation is provided. In order to describe the design in the clearest way, words in the translation maintain the same position in the utterance of the Chinese originals. This way, the three targets are in the same position and present a similar contour on the preceding part of the utterance (the proper name *Wang Yan* was realised with a similar pitch contour in all conditions), but are marked with different degrees of acoustic prominence: (27) a relatively prominent condition; (28) a more prominent condition (target realised with a greater excursion and longer duration compared to (27)); and (29) a condition in which the target word is attenuated (target realised with reduced pitch range and duration compared to (27)). Note that the semantic congruence or incongruence is driven by the presence or the absence of the word *bèi* (underscored in the examples).

(27) Just now (bèi) WANG YAN found the **KEY** is of great help

(28) Just now (bèi) WANG YAN found the **KEY** is of great help

(29) Just now (bèi) WANG YAN found the **key** is of great help

Li and Ren's findings suggest shallow processing of semantic information in the case of attenuation (29), while an increase in processing concomitant with the increase of the pitch range was observed. Indeed, a pronounced N400 effect was found both in (27) and (28), with the latter presenting a broader and larger distribution of the effect. By contrast, no N400 effect was found in (29).

Taken together, results of Wang et al. (2011), Kristensen et al. (2012) and Li and Ren (2012) suggest that both information structure and acoustic information contribute to orienting attention to the part of the message that is most important. Top-down inferences deriving from the context direct the attention where the important part is expected, while signal-based factors connected to different degrees of prosodic prominence orient attention in the absence of expectations. Thus, these studies provide evidence for the shallow processing of the semantic information in the post-focal position. This shallow processing is caused on the one hand by the lack of expectations to find important information after the focal constituent of the utterance, which draws attention away from this position (Wang et al., 2011), and on the other hand by the lack of attention orienting due to the absence of prominence cues (Kristensen et al., 2012; Li & Ren, 2012). In addition, attentional resources are largely allocated to the constituent bearing the focal accent (or the enhanced prominence) and do not shift to the subsequent part of the utterance.

The present study aims at further testing the role of prominence-lending features in attention orienting (i.e. signal-driven modulation of attention), asking whether the acoustic prominence in post-focal position would cause a reorienting of attention from the precedent accented constituent. This hypothesis will be tested by investigating utterances in isolation (i.e. without preceding context) realised in the Bari variety of Italian presented to native speakers of this variety. In this experiment, monolingual native speakers of the Bari variety listened to utterances in isolation realised in two different sentence modalities (statements and questions), and with the critical word occurring in two focal conditions (critical word in narrow focus and in post-focal position). The two sentence modalities were adopted because of the differences in their prosodic realisation in post-focal position. Recall from 2.3.3 that in the Bari variety of Italian questions are realised with a post-focal compressed pitch accent. By contrast, post-focal position in statements is realised with a flat and low contour (see 2.3.3, and see also 6.2.6).

More specifically, the current experiment investigates the online processing of semantically congruent and incongruent words realised in utterances in two sentence modalities, questions and statements, and two prosodic conditions, focal (in narrow focus, NF) and post-focal (PF). An example of the experimental material is given in (30)-(31). Target words are reported in bold, narrowly focused words are reported in capital letters. Note that in questions and statements words in narrow focus have a contrastive accent (high prominence; 30a. and 31a.). Words in post-focal position in statements were realised with a flat, low F0 (31b.), whereas the same words in questions bore a compressed pitch accent (30b.), presenting enhanced cues to prominence in comparison to statements. Both the realisation of the post-focal position in the two modalities are the common realisations found in the Bari variety (see 2.3.3).

(30) QUESTIONS

Per aprire la porta,
To open the door,

a. **NF** bisogna girare la **MANIGLIA** (congruent) vs. **FRITTATA** (incongruent)?
3 SG .INF DET N N
should turn the handle omelet

b. **PF** bisogna GIRARE la **maniglia** (congruent) vs. **frittata** (incongruent)?
do you have to turn the handle vs. omelet?

(31) STATEMENTS

Per arieggiare la stanza,
To ventilate the room,

a. **NF** bisogna aprire la **FINESTRA** (congruent) vs. **LATTINA** (incongruent)
3 SG .INF DET N N
should open the window can

b. **PF** bisogna APRIRE la **finestra** (congruent) vs. **lattina** (incongruent)
you need to open the window vs. can

The fact that the post-focal position may hold two different degrees of prosodic prominence allows us to better disentangle the top-down expectations of prominence from the bottom-up inferences and their role in the orienting of attention. In fact, the post-focal position might be presumed to be processed in a shallow manner, since the early encountering of the accent might suggest that important information is no longer presented further downstream (top-down expectations). This can happen in statements (31b.), which do not present acoustic cues to prominence in the post-focal position. By contrast, in the case of questions (30b.), the presence of cues to prominence might preclude the possible shallow processing of post-focal material, redirecting attention from the precedent focal accent to the post-focal position.

In the present experiment semantic congruence is predicted to have a general effect on language processing, eliciting a more pronounced N400 for incongruent over congruent critical words (Kutas & Federmeier, 2011; see also 3.7.1). The relative difference of congruence is further predicted to be modulated by prosody (Li & Ren, 2012). In fact, the ERP amplitude has been shown to be more enhanced for attended stimuli that were prosodically highlighted (Li & Ren, 2012). The difference in the amplitude of the N400 between incongruent and congruent conditions is predicted to be higher when the processing of the incongruity is facilitated by prominence. In particular, the processing of the incongruence will be facilitated by the narrow contrastive accent (30a., 31a.), yielding a large N400 difference between semantically incongruent and congruent target words. This difference in amplitude will remain stable between statements and questions (30a., 31a.). The hypothesis for words in post-focal position is, by contrast, twofold: in the case of statements (31b.), the missing acoustic cues for prominence in this position should not encourage deep processing of the word (less pronounced N400 difference curve), in the case of questions (30b.), the cues for prominence should prevent shallow processing of the word (larger N400 difference in the questions compared to the statements). A further indicator of signal-driven attention which could be present in the case of the post-focal questions (30b.) is the elicitation of a late positive ERP deflection. This could be driven by the update of the mental model. This update could be caused by the reorienting of attention induced by the presence of a rather prominent element, featuring a rising movement in pitch (see for example Fraundorf et al., 2010; Hsu et al., 2015; see also 3.7.2 and 3.8 and Polich, 2007 for a review of the P3 family connected to attentional mechanisms and Coulson et al., 1998, and Sassenhagen et al., 2014 for interpreting the P600 as part of the P3 family). Indeed, late positive deflections are connected to the update of the mental model (Brouwer & Hoeks, 2013; Burkhardt, 2006; Schumacher et al., 2015; Wang & Schumacher, 2013), which in this case can be caused by the fact that the post-focal pitch accent encodes the speech act information. This could result in the later update of the mental model, which occurs in order to include illocutionary information in the mental representation (see Gisladottir, Chwilla, & Levinson, 2015 for the processing of speech acts, Schumacher & Weiland, 2011 for processes of pragmatic enrichment and Coulson & van Petten, 2002, de Grauwe et al., 2010, Weiland, Bambini & Schumacher, 2014 for the processing of figurative language). This late positivity might also reflect the initiation of the action of answering in the mind of the listener (Gisladottir et al., 2015).

The present experiment aims to further investigate language-specific expectations as to how prosodically prominent elements occurring in the upcoming part of the utterance after the focus are processed. The expectations concerning the occurrence of prominent information that have been discussed above regarding the experiment of Wang et al. (2011) can not only be derived by the context, but can also be inferred by the probabilistic distribution of discourse prominent and prosodic prominent elements within the utterances of a specific language (see 3.5). In the rating experiment Bari listeners seem to expect prominence in post-focal position and to rate the prominence degree of constituents occurring in this position accordingly. Thus, the findings of the previous experiment suggest that Bari listeners already allocate attentional resources to post-focal position. However, this does not rule out the possibility that high prominence at the beginning of the utterance, attracting a large amount of attentional resources, could cause the semantic processing in post-focal position to be shallow. The different prosodic realisation of the post-focal position in questions and statements allows us to test whether attentional resources are allocated in post-focal position by default and are not diverted from it by the early focal accents. The possible scenarios are the following: (i) if attention is already oriented towards this position, neither statements nor questions should undergo shallow processing, even if the amplitude of the two N400 effects elicited could be different (higher for questions compared to statements); (ii) if attention is attracted away from the post-focal position (similarly to what has been reported in previous studies on other languages), the presence of fine-grained cues to prominence (pitch accent to signal sentence modality) could play a role in re-orienting attention towards this part of the utterance.

The effect of the interplay between top-down and bottom-up processes guided by prosodic prominence on semantic processing will be here investigated.

7.2 Method

7.2.1 Participants

Thirty-two right-handed, monolingual native speakers of Italian participated in this ERP experiment after giving written informed consent in accordance with the Declaration of Helsinki. Participants were students of the University of Bari. The participation in this experiment was conditional on subjects speaking the same variety of Italian spoken in the recorded stimuli, the variety of Italian spoken in Bari. Therefore, to participate, subjects needed to come from and to live in the area around Bari. Participants (25 female, 7 male) were aged between 19 and 32 years (mean-age = 22.72 years, SD = 2.79). None of them reported any auditory, visual or neurological impairment.

7.2.2 Material

The experiment had a 2 x 2 x 2 design (three factors with two levels each): (i) the factor prosody with 2 levels: narrow contrastive focus (NF) and post-focal position (PF); (ii) the factor semantic congruence: semantically congruent (C+) or incongruent (C-) with the context; (iii) sentence modality: realisation of the utterances as statements (S) or as questions (Q). An example of the stimuli in the different conditions can be seen in (30)-(31), where the target word is indicated in bold and the contrastive narrow focus in capital letters. For the interested reader, the list of all the stimuli is provided in Appendix A4.

All stimuli were recorded by a trained female phonetician native speaker of the Bari variety of Italian in a sound-attenuated cabin (44 100 Hz sampling rate, 16-bit resolution). In order to ensure segmental comparability of the 120 critical words, they are composed of three syllables with primary stress on the second syllable. Figure 94 shows the mean F0 contours of all trials (time window of the whole utterance) as well as the individual F0 contours of all critical words superimposed on each other. Means are calculated aggregating the factors prosody and sentence modality. Each mean F0 contour for one condition has two lines, one indicating congruent stimuli and the other indicating incongruent stimuli. Figure 95 is similar to Figure 94, but shows only the time window of the target.

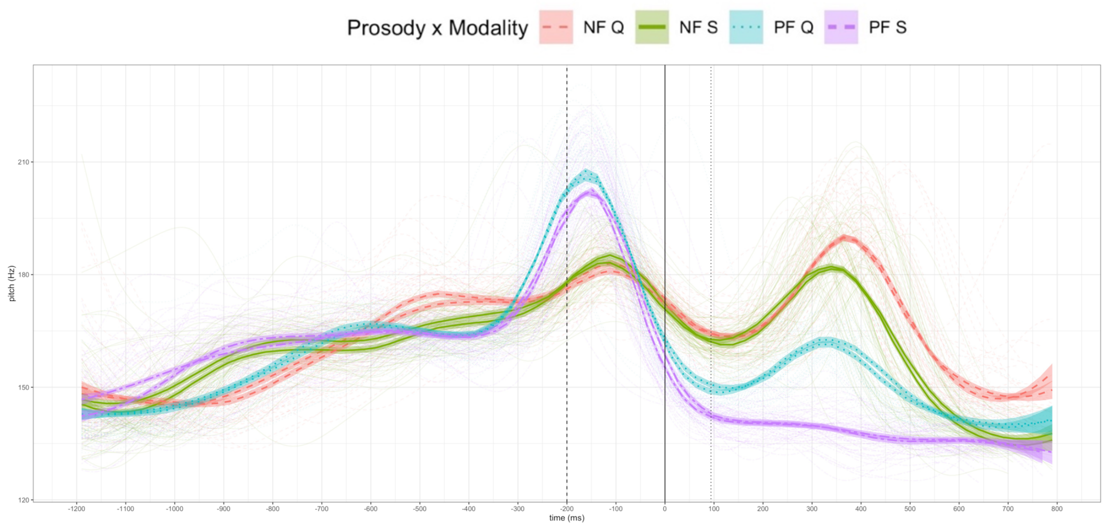


Figure 94. F0 contours of the stimuli form the beginning to the end of the utterances. The thicker lines represent the means of the contours aggregated by the factors prosody and sentence modality. The solid vertical line indicates the onset of the determiner, the dotted vertical line at 100 ms indicates the onset of the target, while the dashed vertical line indicates the 200 ms before the onset of the determiner.

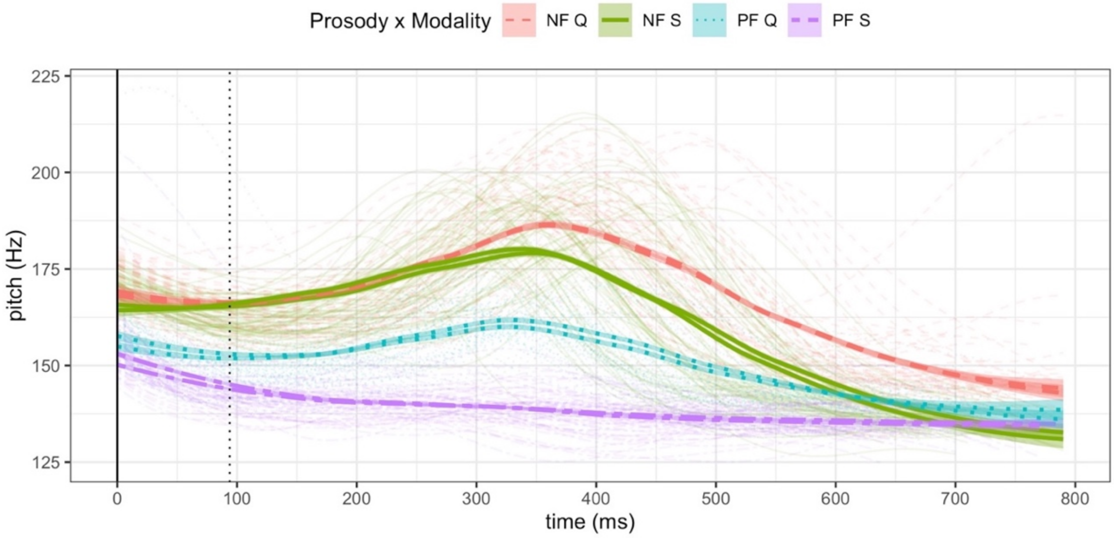


Figure 95. F0 contours of critical words. The means of the contours are represented by thicker lines. Time (in milliseconds) starts at the onset of the determiner (la) of the target word (indicated by the solid vertical line). The dotted vertical line at 100 ms indicates the onset of the target.

Table 41 reports the mean of the maximum values of F0 and the mean of periodic energy for the target and the mean duration of the determiner and of the target. Figure 96 shows the relative values of Periodic Energy Mass (PEM) for the target word in all the prosody × modality conditions. Results of the mixed analysis showed that narrow focus question and statements had the effect of increasing PEM values compared to post-focal questions (NF Q: $\beta = 14.98, \pm 1.53$ $p < 0.0001$; NF S: $\beta = 12.62, \pm 2.3$ $p < 0.0001$). By contrast, post-focal statements had the effect of decreasing PEM values compared to post-focal questions (PF S: $\beta = -8.68, \pm 2.3$ $p < 0.001$).

condition	**F0 [Hz]**	**periodic energy [dB]**	**determiner duration [ms]**	**target duration [ms]**
NF Q	197.4 (7.62)	0.57 (0.36)	95.4 (21.7)	764.41 (58.68)
NF S	191.31 (8.74)	0.54 (0.36)	97.2 (19.7)	754.08 (61.27)
PF Q	166.84 (6.29)	0.53 (0.34)	93.1 (25.1)	683.85 (67.54)
PF S	142.42 (3.98)	0.44 (0.31)	89.9 (22.6)	683.55 (57.79)

Table 41. Means and standard deviations (in parentheses) of maximum F0 in the target, maximum periodic energy in the target, determiner duration and target duration, for each prosody × modality (from the top: narrow focus questions, narrow focus statements, post-focal questions, post-focal statements).

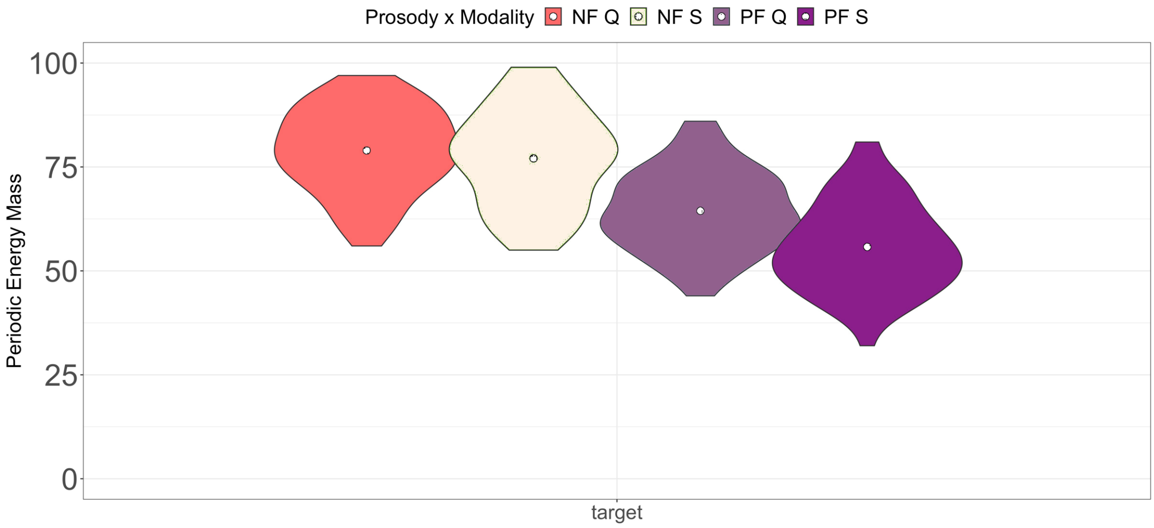


Figure 96. PEM of the target word in all the Prosody × Modality conditions (from the left: narrow focus questions, narrow focus statements, post-focal questions and post-focal statements). White dots indicate mean values.

Fillers were also recorded and were realised either all in broad focus (all new condition, default for hearing utterances without context) or had a narrow contrastive focus in a position that differed from the one used for the trials.

Since not only prosody but also cloze probability of the target influences the N400 (see 3.5.1 and Kutas & Hillyard, 1980), the following measures were taken. For reasons intrinsic to the experimental design, in the C- condition the cloze probability of the target (being semantically incongruent with the context) would be very low, always differing from the C+ condition. To control for this variable would not make sense in the present study. Therefore, instead, targets were controlled for the frequency of occurrence, to prevent that the modulations of the N400 between congruent and incongruent words was further modulated by differences in their frequency index (Rugg, 1990). The calculation of frequency was carried out using as a reference corpus itTenTen16 (Jakubíček et al., 2013) and calculating the value on the Zipf's scale (see van Heuven, Mandera, Keuleers & Brysbaert, 2014). Figure 97 shows the results for all the stimuli in the different conditions. Results of the type II Wald test run on the mixed analysis with FREQUENCY as the dependent variable, CONGRUENCE and MODALITY as fixed effects, and random intercepts for ITEM as random effects, did not show an interaction of the fixed effects [$\chi2(1) = 0.14$ $p= 0.71$].

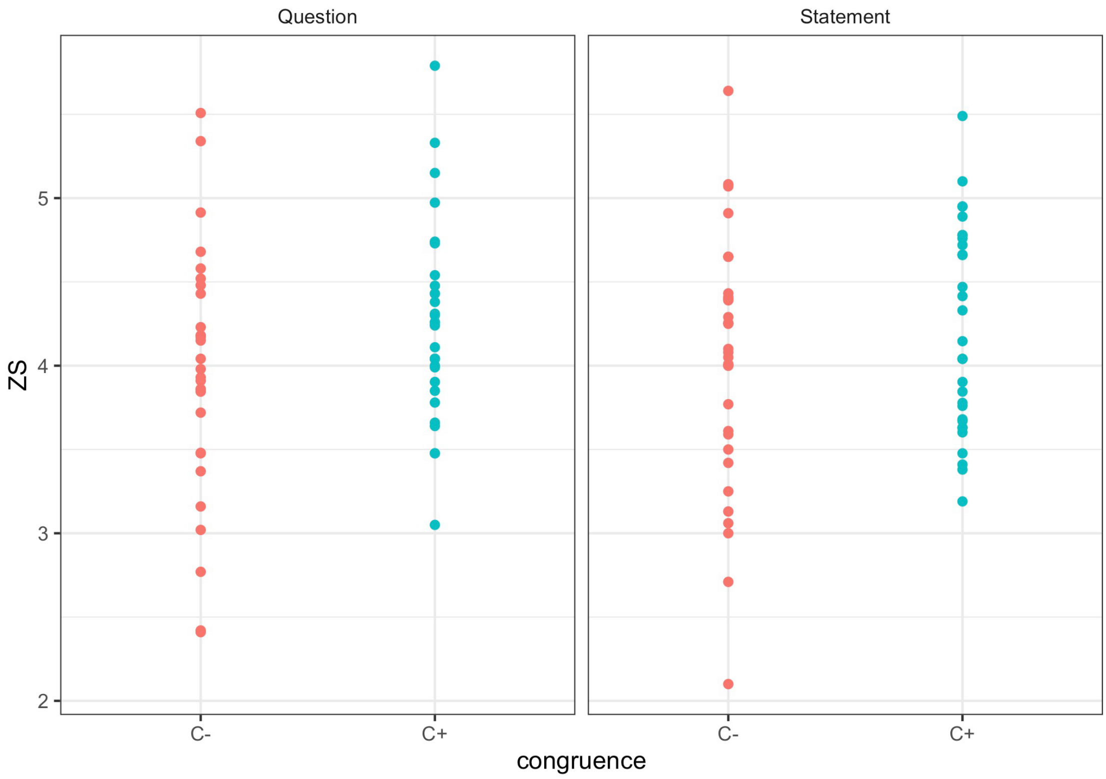


Figure 97. Frequency values measured in terms of Zipf's scale. On the left panel, values for questions, on the right panel values for statements. On each panel, left values relate to incongruent words (C-) and right values to congruent words (C+).

Moreover, the incongruent targets were controlled for semantic congruence with the preceding verb and the frequency of the target occurring in the context of the verb was calculated (again using the itTenTen16 corpus). Note that the lexical items used in the two modalities were different, in order to prevent effects of repetition of the words on the N400 (van Petten et al., 1991) between the modalities. Figure 98 shows the results. The figure shows that the incongruent targets have lower values than the congruent ones, given by the difficulty of at the same time controlling number of syllables, stress patterns and incongruence with the preceding context. However, the type II Wald test ran on the mixed analysis with FREQUENCY as the dependent variable, CONGRUENCE and MODALITY as fixed effects, and random intercepts for ITEM as random effects, also in this case did not show an interaction of the fixed effects [$\chi 2(1) = 0.08$ $p= 0.78$].

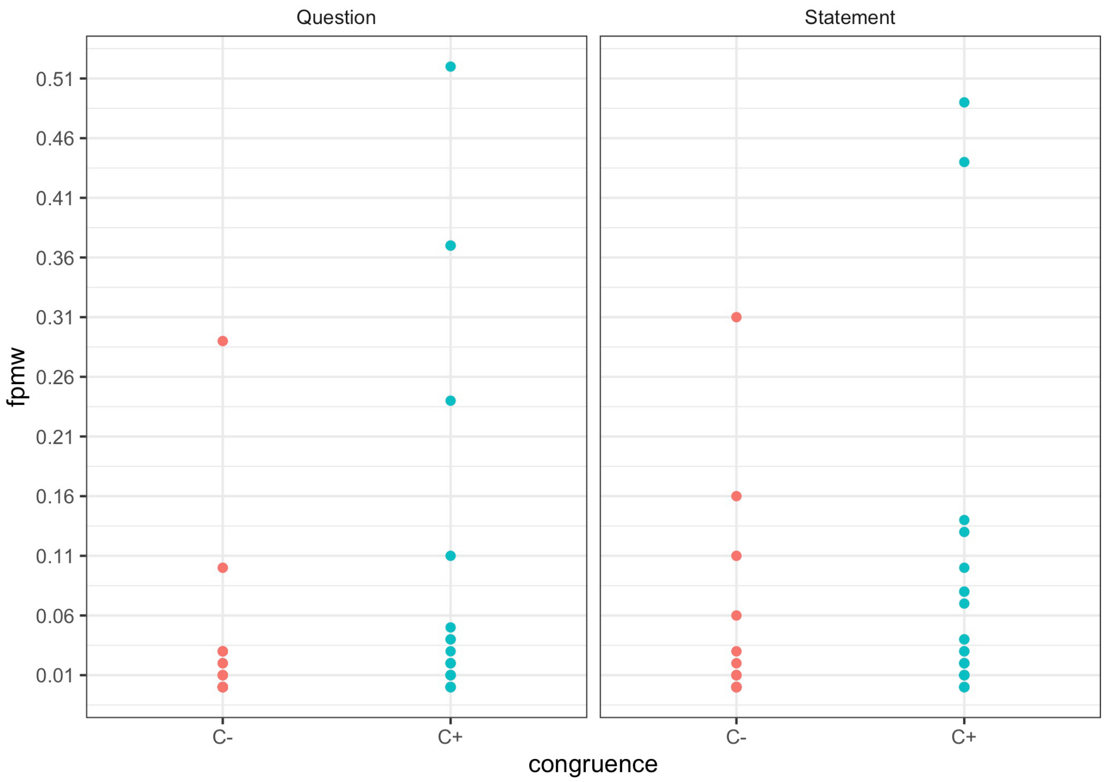


Figure 98. Frequency per million words (fpmw) of the occurrence of the target word with the word preceding it. On the left panel, values for questions, on the right panel values for statements. On each panel, left values relate to incongruent words (C-) and right values to congruent words (C+).

7.2.3 Procedure

Each experimental session contained 360 trials presented to participants without context. They involved 240 critical items (60 lexically different sentences × 2 intonation contours × 2 sentence modalities) plus 120 filler items. Critical and filler items were pseudo-randomized.

During EEG recordings, after each auditory stimulus participants performed a word recognition task. Recognition words were equally distributed to address the first part of the sentence, the servile verb *bisogna* (you need) or the infinite verb (e.g., *girare*, to turn). Recognition of the critical word occurred only when it was congruent within the utterance; incongruities were never addressed directly. The expected yes/no responses were equally distributed across the materials and conditions. Each experimental session contained all 360 trials, involving 240 critical items plus 120 filler items. Critical and filler items were pseudo-randomized. Three lists with different randomisations were used. Each participant saw only one of the three lists of the 360 items which were presented in eight blocks with pauses in between. Each block was made up either of questions or of statements, in order to prevent participants to focus on sentence modality. In order to avoid repetition effects, test sentences with the same lexical material were assigned to different experimental blocks. Furthermore, in order to prevent systematic order effects in the exposure to the stimuli with the same lexical material were presented in different condition sequences across the blocks. After electrode application, participants were instructed to look at the computer monitor in front of them and to focus on a fixation star while the auditory stimuli were presented over loudspeakers.

At the beginning of the recording session participants were familiarised with the experimental procedure by means of a short practice block (six stimuli). Participants were instructed to look at the monitor in front of them and to focus on a fixation star while the auditory stimuli were presented over loudspeakers. The electroencephalogram (EEG) was recorded and digitised (500 Hz) using 24 Ag/AgCl electrodes placed on the scalp according to the standard 10-20 system (*BrainVision Brain-Amp amplifier*). EEGs were referenced online to the left mastoid. The ground electrode was placed at AFz. Electrooculogram (EOG) was recorded by two pairs of electrodes to control for eye-movement artefacts. For horizontal eye movements, the electrodes were placed at the outer canthus of each eye, and for vertical eye movements, they were placed above and below the left eye. The impedances of the electrodes were kept below 5 kΩ.

7.2.4 Analysis

Data were analysed using a Python3 implementation of MNE python version 0.19 (Gramfort et al., 2013). Data were rereferenced offline to linked mastoids, eye artefacts were automatically detected. Because of the presence of differences in the auditory signal prior to the critical word onset, the EEG was filtered with a 0.3 - 45 Hz filter to counter pre-stimulus evoked activity (Maess, Schröger & Widmann, 2016). The identified portions of raw data containing blinks were excluded from further analysis. The data were epoched from -200 to 1000 ms post onset of the determiner of the critical word and resampled to 100 Hz for further analysis because of computational limitations. Trials with false or time-out responses to the comprehension question were excluded from the analysis. This resulted in the rejection of 7% of the data points over all conditions.

Due to differences in the acoustic properties across the critical conditions immanent to the design, a regression-based ERP (rERP) analysis was performed (Hauk et al., 2006; Smith

& Kutas, 2015) using the lm() function in R (R Core Team, 2019). Linear models by subject, channel and sample (i.e. for time points in 10ms steps) with factors PROSODY (NF, PF), CONGRUENCE (C+, C-) and MODALITY (Q, S) as well as PITCH (Hz, continuous) and PERIODIC ENERGY (dB, continuous; Albert et al., 2018). Pitch and periodic energy were extracted for periods of 10ms from the audiofiles using PRAAT and R (Boersma & Weenink, 2020; R Core Team, 2019) from -200 to 1000ms post determiner-onset (the sentence-final critical words ended at 800ms; the interval from 800-1000ms was filled with silence). Linear mixed-effect models were calculated using the lmer() function from the "lme4" package (Bates, Maechler, Bolker & Walker, 2014) for R with mean fitted values in the windows 400-600ms and 600-800ms. The models included three fixed factors PROSODY, CONGRUENCE and MODALITY, as well as two continuous factors SAGGITALITY and LATERALITY based on the planar (x and y) coordinates of the standard BESA coordinate system. Random intercepts for subjects as wee as by subject random slopes for the effect of PROSODY were fitted into the model.

Stimuli, data and the scripts used for the analysis are retrievable at https://osf.io/zepfa/.

7.3 Results

The grand-avaraged rERPs (fitted microvolt values) for the eight experimental conditions are shown in Figure 99. The rERPs were time-locked to the onset of the determiner of the critical word (at 0 ms). Given the specific hypothesis concerning the N400, the time window from 400 to 600 ms after the onset of the determiner was analysed statistically. The reason for choosing the determiner as point to time-lock the rERP is that determiner + noun constitutes a prosodic word. However, the determiner does not contain information about the semantic congruence (or incongruence) of the target. Subtracting the duration of the determiner, the effect that is observed can be shifted to the 300-500 ms time window. This negativity was considered as the N400, since the time window in which it is observed corresponds to the time window around 300-500 ms after the onset of the critical word without the determiner (see Figure 99 and Figure 100). Statistical analysis for this time window (400-600 ms) registered significant effects of the interactions among saggitality, prosody, modality and congruence [$\chi2 = 21.85$, $p < .0001$]. Contrast obtained with emmeans() function (Hauk et al., 2006) show that in the posterior regions for incongruent critical word in narrow focus for both questions and statements rERPs deviate from expected words in a negative direction (prosody = NF: C-

Q posterior – C+ Q posterior, β = -0.42, p <.0001, C- S posterior – C+ S posterior, β = -0.61, p <.0001). The same also holds for post-focal condition (prosody = PF: C- Q posterior – C+ Q posterior, β = -0.23, p <.0001, C- S posterior – C+ S posterior, β = -0.56, p <.0001). The pairwise comparison (Figure 100) shows that the effect of congruence is higher in statements than in questions, for both prosodic conditions. In questions, the effect of congruence is higher for NF than PF whereas in statements the effect of congruence does not differ between the conditions. In posterior regions, NF Q increase the negativity associated to C+ in comparison to NF S (C+ Q posterior – C+ S posterior, β -0.19082, p <.0001), whereas the negativity of C- remains the same between the different illocutions (C- Q posterior – C- S posterior, β = -0.001, p = 1). For PF the negativity of C- increases in statements in comparison to questions (C- Q posterior – C- Q posterior, β = 0.29, p <.0001), whereas the negativity for C+ remains the same between the different illocutions (C+ Q posterior – C+ S posterior, β = -0.03, p = 0.99).

In addition, in the time-window of 600–800 ms, the statistical analysis registered an interaction of prosody, modality, congruence and sagittality [$\chi 2$ = 16.47, P < 0.001]. A further inspection of the interaction, obtained with emmeans() function, revealed a more pronounced positivity over anterior sites for incongruent critical word in questions in post-focal position (β = 0.36, P < 0.0001; see Figure 100).

Looking at the single effects of prominence on the components, in the 600-800 ms time-window a higher positivity is registered for questions over statements in NF for incongruent and congruent stimuli (congruence = C-: Q anterior – S anterior, β = 0.23, p =.003; congruence = C+: Q anterior – S anterior, β = 0.29, p <.0001). The same holds for the incongruent stimulus in PF (congruence = C-: Q anterior – S anterior, β = 0.66, p <.0001). By contrast, the congruent stimulus yields a higher late positivity for statements compared to questions (congruence = C+: Q anterior – S anterior, β = - 0.28, p =.0002).

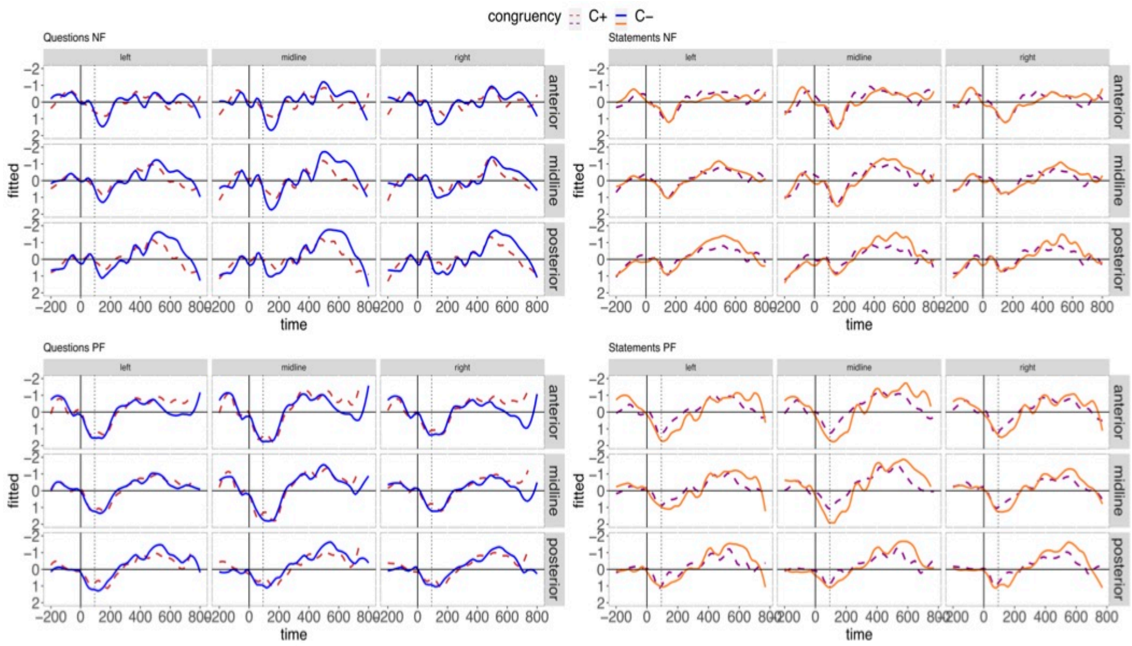


Figure 99. Grand-average rERPs for the four experimental conditions. The negativity is plotted upwards. The first line shows the results for NF, the second line shows the results for PF. Dashed lines indicate C+ condition, solid line C-. Electrodes are grouped by centrality, laterality and saggitality. Time course on horizontal axis spans from 200 ms before until 800 ms after the onset of the determiner of the critical word (= vertical solid bar). Vertical dotted bar indicates the onset of the target.

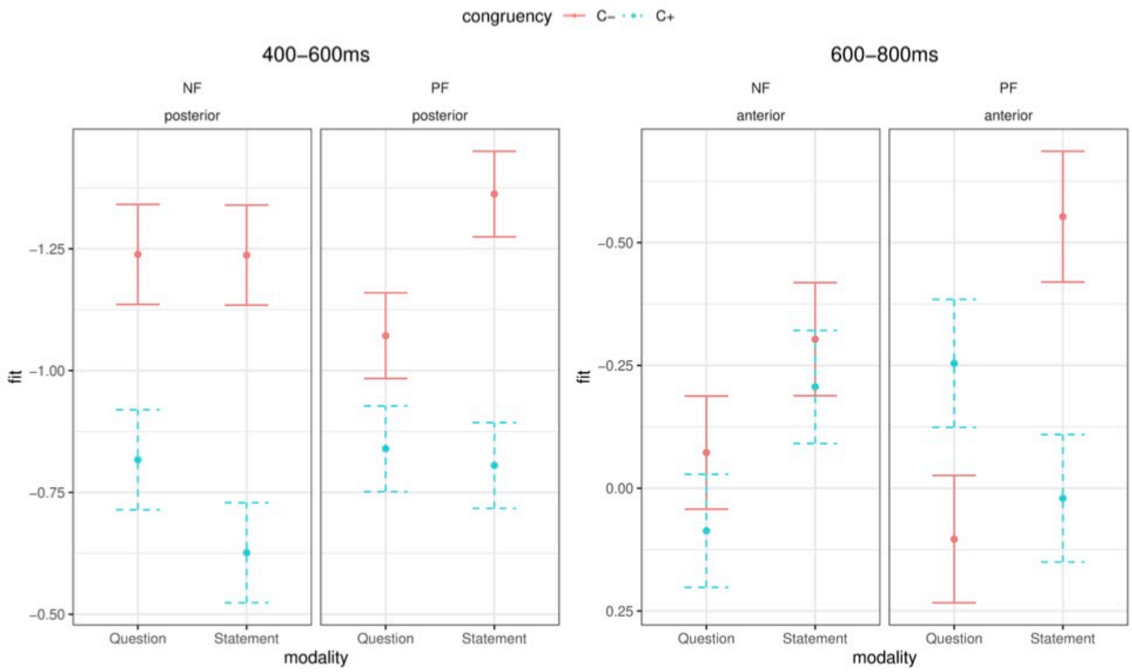


Figure 100. Significant congruence effects of rERPs for the time window from 400-600ms (left panel) and from 600-800ms (right panel). Error bars represent 83 % confidence intervals. Negativity is plotted upwards. C+ is represented with dotted lines, C- with solid lines.

Given that the full model of the 400-600 ms time-window registered a significant interaction of laterality, prosody, modality and congruence [$\chi 2 = 7.31$, $p < .03$], further comparisons among conditions of ERPs modulations in the semantically congruent condition were performed. Contrast obtained with emmeans() show that in the left region there is a tendency for statements in PF to elicit a more pronounced N400 compared to NF (modality = S: NF left – PF left, β = -0.29, p =.05). This tendency becomes significant in the midline region (modality = S: NF left – PF left, β = -0.41, p =.001). By contrast, the effect of the contrast between NF and PF in questions is not registered.

In the comparison between modalities in the congruent targets (Figure 101), a reduced N400 amplitude is registered on the left region for questions compared to statements in the same condition (prosody = NF: Q left – S left, β = 0.12, p < .0001). A reduced N400 effect is further registered for questions in post-focal position compared to statements in the same position (prosody = PF: Q left – S left, β = 0.24, p < .0001; prosody = PF: Q midline – S midline, β = 0.32, p < .0001; prosody = PF: Q right – S right, β = 0.17, p < .0001). Given that these effects are more present and more widely distributed for PF, this confirms that in post-focal position of questions, the presence of the pitch accent signalling sentence modality, causes some attentional resources in questions to be diverted from the semantic processing towards the processing of the speech act. This happens to a lesser extent in narrow focus, where more attentional recourses are allocated.

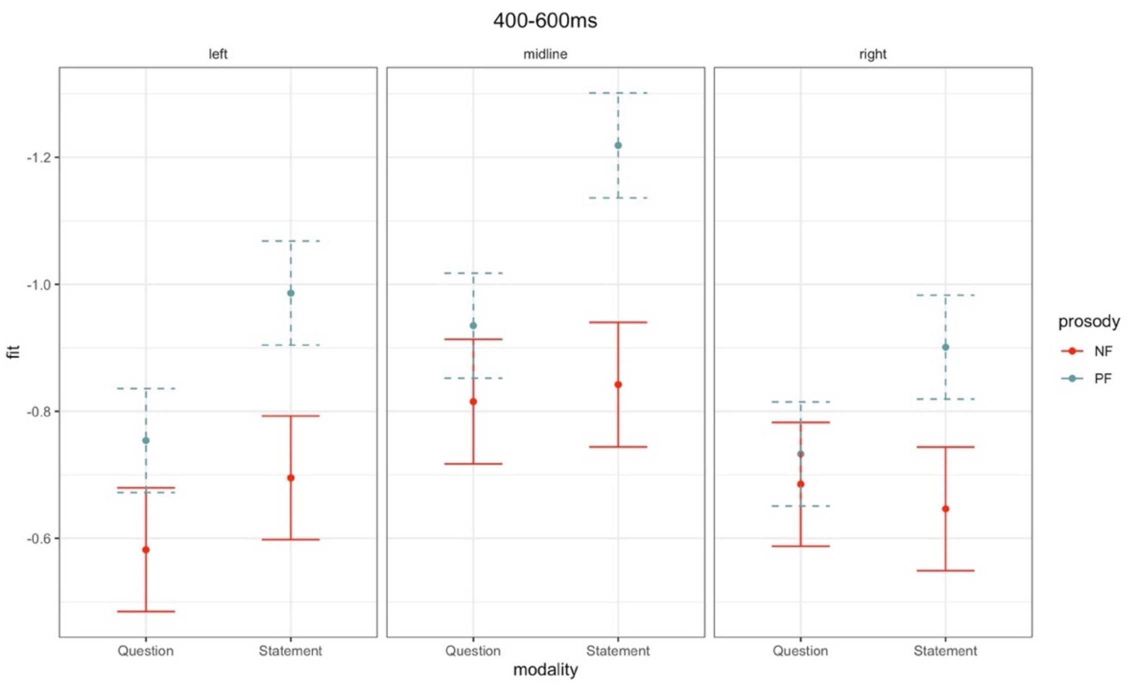


Figure 101. Effects of rERPs in the processing of congruent targets for the comparison between focal conditions in questions and statements for the time window from 400-600ms. Error bars represent 83 % confidence intervals. Negativity is plotted upwards. PF is represented with dotted lines, NF with solid lines. Left panel reports effects in the left lateral region, right panel effects in the midline lateral region.

7.4 Discussion and conclusion

In this EEG study, the hypothesis on whether language-specific allocation of resources prompted by post-focal F0 movement modulates the depth of processing was examined. The semantic processing of words that are incongruent with the context of the utterance and that are presented with different degrees of prominence was investigated. The aim was to analyse the contribution of signal and expectation-driven inferences on the online processing.

General results revealed that semantically incongruent words elicited an N400 effect for all conditions (NF Q, PF Q, NF S and PF S). Thus, differently from what has been reported in the literature (concerning other languages), the present experiment did not register a shallow processing of the semantic information of words occurring in the background (i.e. post-focal position) of statements. Differently for what was expected, the N400 effect registered in statements was higher compared to questions. Nonetheless, results show that the processing magnitude is modulated by prosody, this time in line with what has been observed in previous studies (Kristensen et al., 2012; Li & Ren, 2012).

Evidence for the prominence-driven modulation of processing is the larger N400 effect, registered in questions for the contrastively accented condition (NF Q) compared to the post-focal condition (PF Q). This finding supports previous studies considering prominence as an attentional orienting device (Kristensen et al., 2012; Li & Ren, 2012; Magne et al., 2005; Röhr et al., 2020; Sanford et al., 2006; Schumacher, Backhaus & Dangl, 2015). In particular, it presents further evidence for the findings from Chinese of Li and Ren (2012), indicating that the gradual enhancement of the acoustic parameters cueing prominence is processed in real time and contributes to the orienting of attention. Indeed, in Li and Ren's study, a word marked with a greater pitch excursion and realised with longer duration induced deeper semantic processing of the stimulus compared to the same word realised with lower excursion and reduced duration. Similarly, in the current experiment the greater prominence of the contrastive accent compared to the post-focal accent, the former presenting more F0 excursion and corresponding to enhanced values of PEM, engendered an increased allocation of attentional resources compared to the less prominent accent. In turn, the attentional resources devoted to the early narrow focus are reflected in a deeper processing of the incongruence.

Surprisingly, a different result is found for statements. In this case, the target occurring in post-focal position (PF S), despite lacking F0 movement and presenting the lowest values of PEM, still registered a pronounced N400 effect. Even more surprising is that the effect of the incongruence was not modulated by prominence: the same depth of processing was registered for both the presence of a very prominent accent (NF S) and for the lack of prominence cues (PF S; no F0 movement and low values of PEM). In addition, the incongruence of the post-focal position in statements (PF S) seems to be deeper processed than in questions occurring in the same position (PF Q), as shown by the more negative deflection elicited by the former condition compared to the latter. This also contradicts the predictions made for the comparison between statements and questions, since the target in questions showed higher values of pitch and of PEM compared to the ones in statements and this was expected to yield differences in processing in the opposite direction.

Therefore, at a first superficial look these results seem to contradict previously accumulated evidence that the post-focal position is only processed in a shallow way and that the lack of cues to prominence impedes the orienting of attention (Kristensen et al., 2012; Li & Ren, 2012; Wang et al., 2011). More importantly they seem to challenge to some extent the conception of the prominence lending function of prosody as an attentional orienting device. Nonetheless, an explanation regarding this effect was already part of the predictions delineated in the introduction of this chapter (7.1) and shows that the results are indeed in line with the previous literature. The interpretation of these results needs to consider top-down expectations. Indeed, these results can be explained by the fact that Bari listeners have expectations of finding prominence cues in the post-focal position and that attention in the post-focal position is present by default, as suggested also by results of the rating experiment in Chapter 6. These expectations can be interpreted by the inferences driven from the probabilistic distribution of prominence in the post-focal position and to the reportedly high tendency to place prominent words in final position (see 2.3.3. and 2.3.5). Another possible explanation is that, given that in the stimuli used in this experiment questions and statements could only be distinguished by prosody, attentional resources could already be oriented towards this position to distinguish between modalities. However, the fact that the experimental blocks were divided per sentence modality (each block contained either questions or statements), rules out this explanation. However, future research employing native speakers of a different variety (or language) might provide additional evidence for the fact that default attention in this variety of Italian is guided towards the post-focal position by top-down expectations. In any case, results of the pronounced N400 effect in the target in post-focal statement seem to suggest that the attentional resources are entirely oriented towards the target.

Nevertheless, the hypothesis regarding the present experiment implied that the signal-based factors would still enhance the depth of semantic processing. Looking only at the N400 effect, this does not seem to be the case, since questions in post-focal position elicited a reduced N400 effect compared to statements. If attention is considered to be present by default in this position, the presence of reduced attentional resources should not be attested. Even more so, this should not be the case, given the relatively high prosodic prominence characterising questions. Yet, the pitch accent in questions signals sentence modality. Therefore, the cues to modality may have drawn processing resources towards the modality and its corresponding speech act rather than towards the semantic congruence. This hypothesis is confirmed by the presence of a prosody-related modulation which occurred in a later time window (emerging between 600-800ms), where a positive deflection connected to the incongruent stimulus was engendered by questions and not by statements. Since the Late Positivity has been found elsewhere as a marker of mental model updating (see e.g., Baumann & Schumacher, 2012; Schumacher & Baumann, 2010), the present findings suggest that attentional resources are indeed allocated to this condition. The orienting of attentional resources towards this part of the utterance is modulated by the effect of congruence, suggesting that these prosodic cues that are modality-specific lead to signal-driven attention allocation.

These results suggest that the increased prominence in the post-focal constituent prioritise the request speech act at an earlier stage, consuming the semantic processing resources. Indeed, the interrogative modality can initiate the action of answering (Gisladottir et al., 2015), reflected by the Late Positivity. Therefore, the presence of these cues to modality creates an interference, leading to a less elaborate processing of semantic information, reflected by the decreased N400 effect found in the post-focal position of questions. Thus, the semantic incongruence in the post-focal position of statements was more fully processed compared to questions. However, the presence of the prosodic cues induced the allocation of additional resources towards a later update of the discourse representation.

This update comprised both the mental representation of the illocutionary information, which needs to be encoded in the discourse model, and the integration in the discourse representation of the incongruence. Indeed, the effect of the prosodic cue to modality leads to a short-term less deep processing of the incongruity. This incongruity is however fully resolved in a later stage, during the update of the mental representation (see e.g., Yang et al., 2013). The presence of this effect is also confirmed by the fact that congruent targets elicit a larger N400 when they occur in the post-focal position of statements compared to questions. In the latter modality, some resources are devoted to the processing of sentence modality conveyed by the prosodic cue, at the expenses of semantic processing. The presence of a biphasic N400-Late Positivity pattern indicates that prosodic prominence can modulate the initial as well as the later stage of the discourse processing. This effect, which is registered in post-focal position, emerges to a lesser extent in narrow focus, since more attentional resources are allocated towards the target occurring in this condition.

Interestingly, in the congruent condition, an enhanced N400 is registered for post-focal statements over narrow focus statements. This effect could be interpreted in line with Baumann and Schumacher (2012), where the inherent function of the lack of pitch movement induced the search for a proper previous referent, which results in increased processing demands. By contrast, in questions the difference between post-focal and narrow focus is not enhanced as in statements, because the higher level of prominence registered in post-focal questions engenders a more similar processing compared to narrow focus. However, this hypothesis needs to be confirmed by further studies deepening the relation between prosodic marking of information status in Italian and its consequences in processing. A further possibility to consider, is that, since the congruent target words in the stimuli represented accessible and not given referents, the presence of pitch movement associated with them, even if related to modality, was perceived as more appropriate for these types of referents compared to the complete lack of movement. The present interpretation of the results is also in line with the higher Late Positivity engendered by post-focal questions over post-focal statements in the congruent. The positivity for congruent post-focal statements would be elicited by the increased effort to search for a given entity, since this given information has not been established in the preceding part of the utterance. Both these interpretations for the N400 and the Late

Positivity are in line with the findings by Schumacher and Baumann (2010) on German, but again, they need to be confirmed by further research focussing on Italian.

To summarise, the findings of this experiment are in line with the previously established effect that focal exponents which are marked with an appropriate accent orient the attentional system towards their processing (see e.g., Wang et al., 2011). Hereby, additional contribution is provided to literature assessing the role of signal-based processing in real-time processing of information (Kristensen et al., 2012; Li & Ren, 2012): the current results confirm that signal-based cues to prominence trigger attention orienting and that these cues to prominence can also be fine-grained in order to modulate attentional resources, as previously attested by Li and Ren (2012). In addition, this study has revealed that in a language (i.e., the variety of Italian spoken in Bari) in which the distribution of prominences is such that in some cases they also occur post-focally, attention can be drawn both to focal and post-focal information. Despite the fact that the N400 congruence effects are attenuated by the demands of modality-specific processing in post-focal position, effects on processing of modality-induced mechanisms give rise to a Late Positivity. In focal position the effect of congruence is pronounced, meaning that attention is oriented to the target. The same applies to words in post-focal position. This default allocation of attention can be induced by the inferences driven from the probabilistic distribution of prominences in the Bari variety of Italian (characterised by more paradigmatic choice in prominences in post-focal position), and possibly, also by the fact that the end of the utterance is the reportedly preferred location for prominent information in Italian (see 2.3.5). Crucially, in questions, post-focal information presents cues that convey a specific speech act function, modulating the processing of the target and reducing the attention allocated to semantics. Thus, the findings reveal that prosodic cues which are different (either signalling focus or signalling modality), have an influence on selective attention.

Chapter 8

General discussion and conclusion

This thesis was concerned with the prosodic marking of information structure, the distribution of prominence within an utterance and its effects on perception. In particular, it presented a first investigation on the prosodic marking of information structure in the Northern variety of Italian spoken in Udine, including the implementation of categorical analysis with the continuous prosodic parameters related to F0 and periodic energy. The same continuous parameters used in the analyses of the variety spoken in Udine were also used to investigate the prosodic marking of information structure of another variety of Italian, the one spoken in Bari. This variety is particularly interesting (among other Southern varieties) for its greater paradigmatic choice on the distribution of prominence in the post-focal region. The aim of these two investigations was to assess the degree of prosodic prominence in the realisation of the different focal structures, with a particular interest in the prosodic realisation of the post-focal region. The novelty of this production study is that it combined the categorical analysis in terms of pitch accent types of two varieties of Italian with continuous measures related to periodic energy. An additional innovation was the analysis of the acoustic correlates of prominence (i.e., energy and duration) of constituents which only gradually diverged in their degree of prominence as determined by the focus structure they occurred in (i.e., from background, to broad focus, to contrastive focus; previous studies on Italian have compared only background with contrastive focus, see Avesani, Vayra & Zmarich, 2007).

The degree of prominence of words occurring in different focal structures was not only investigated in production but also in perception in both of the above-mentioned varieties of Italian (Chapter 6). The aim was first to prove that participants could perceive the differences in the degrees of prominence, and second, to test whether the different paradigmatic choices in prominence realisation in post-focal position could lead to differences in the ratings of these two varieties. The study is thus the first to systematically explore the gradual perception of the degree of prominence of constituents in different focal structures, which represents research that was still missing in the landscape of the studies on Italian. Both varieties of Italian were subject to the perceptual evaluation of German native speakers (learners of Italian), with the aim of investigating the effect of differences in prominence perception specific to the native language of the listener. A general aim of the conducted rating study was to test the effect of expectation-based and signal-based inferences in the perception of prominence. Finally, the influence of prosodic prominence, and in particular its attention orienting capacity, was investigated during real-time language processing, capitalizing on semantic incongruity effects. The variety taken into consideration was the one spoken in Bari, providing the interesting case of a rather high pitch excursion registered in the post-focal region of questions. This latter investigation had the objective of further deepening the understanding of the role of expectation-based and signal-based factors related to prosodic prominence, this time focusing on real-time integration of prosodic information. The novelty of this latter study was to investigate the semantic processing of post-focal region when it entails movement in pitch that may be interpreted as cue to prominence.

Chapters 1 and 2 provided the necessary theoretical background for an understanding of the investigations conducted in the experimental chapters. This entails the basic functions and features of prosody and prominence and their relation to information structure and information status. The picture that emerged was that of a probabilistic relation between information status and prosodic marking, the details of which are language-specific. The reported differences in this probabilistic mapping between Italian and German was of particular interest. Chapter 1 also deepened the discussion of the parameters related to periodic energy that were later used in the experimental section. Subsequently, Chapter 3 introduced the relation between prominence and its perception, outlining the effects on perception of prominence distribution and the mapping between form and function, together with the relation between prominence and online processing. Studies conducted mostly on West-Germanic languages seem to indicate that listeners benefit from prosodic information to decode the upcoming message and that this decoding is possibly due to the probabilistic inference processes that listeners make, given the knowledge of the distribution of acoustic cues. These inferences also play a role in the perception of the degree of prominence and are presumably language-specific and create different top-down expectations, depending on the native language (or variety) of the listener. However, the literature does not provide clear explanations, neither for when the expectations from the native language meet the signal of the non-native language nor for the consequences on the probabilistic distribution of fine-grained cues to prominence in a certain utterance position. Therefore, to substantiate our insights into the perception of fine-grained cues to prominence and the understanding of the interplay between expectation-driven and signal-driven inferences, Chapter 5, Chapter 6 and Chapter 7 empirically explore the different aspects of the production of prominence patterns in different focal structures and their perception (both behavioural and online).

8.1 Summary and discussion of the experimental results

The experimental part of this thesis has first dealt with the prosodic marking of information structure in one variety of Italian, the one spoken in Udine. The corresponding production experiment investigated the prosodic realisation of words in three different positions in utterances realised in three different focal structures. These focal structures were elicited by different questions providing different discourse contexts, which placed the target word in broad focus, narrow contrastive focus and in post-focal position. The experiment provided a comparison between categorical analysis and continuous measures relative to periodic energy (synchrony, scaling and Periodic Energy Mass/PEM). These continuous measures proved useful for analysing more fine-grained differences between the conditions and aided the interpretation of speaker-specific productions. Results of this experiment have provided further evidence that the mapping between focus types and pitch accents types is not one-to-one, as already reported in previous studies on English and German (Grice et al., 2017; Ito et al., 2004; Mücke & Grice, 2014; Röhr & Baumann, 2010; Roessig & Mücke, 2019, among others). In addition, these results have shown that prominence in this Italian variety is conveyed by both a categorical three-way distinction and a gradual modulation of all the parameters related to prominence that have been investigated: absence or presence of pitch movement for the comparison between background (post-focal position) and focal conditions and gradual modification of the continuous parameters that have been considered to be connected to prominence (energy and duration; similarly to what was found for German by Roessig & Mücke, 2019). Pronounced F0 dynamics marked the focal exponent of contrastive focus and the words occurring in the broad focus domain (mostly rising pitch for narrow contrastive focus, falling pitch for broad focus), while PEM values decreased from contrastive focus to broad focus. The major focus of the production study was to assess whether post-focal elements were indeed produced as less prominent than elements in broad focus. This has been shown to be the case. Even though the post-focal word was in a structurally prominent position, the acoustic cues to prominence were low: low PEM values and no pitch movement.

A secondary result of the production study was concerned with the comparison between two datasets. The two datasets were composed of similar sentences, with one dataset consisting of answers which were shorter (short dataset, in 5.3) compared to the other dataset (long dataset, in 5.2). The comparison showed that the different conditions in the short dataset were realised with a lower degree of overlap of the continuous parameters. Thus, when reading the short dataset (with less redundancy), speakers conveyed the intended information structure more clearly. Results supported previous findings showing that intonation is sensitive to the task and the setting of the experiment (Niebuhr & Michaud, 2015).

While the production study focussed on the variety of Italian spoken in Udine, the introduction to the perception study additionally presented a smaller amount of production data on another variety of Italian, the one spoken in Bari. This variety has been more thoroughly investigated in the literature compared to the one spoken in Udine, and the realisation of the post-focal position in statements with a flat and low contour has already been attested (Grice et al., 2005). Results of this small corpus confirmed that the post-focal region of statements was indeed realised with the previously described F0 contour. By contrast, issues concerning the perception of prominence conveyed by features other than F0 movement had not been analysed before. The results observed for this variety were similar to the ones found for the Udine variety: broad focus and contrastive narrow focus were marked by (rising-)falling and falling accents, respectively. The PEM of narrow focus was higher than that of broad focus, which in turn was higher than that of post-focal position. Furthermore, the post-focal position was not marked by a dynamic F0, and had instead a flat and low contour.

The rating study (Chapter 6) analysed the perception of prominence in the two varieties. Ratings were collected from native speakers of Italian rating their own variety, and from learners of Italian. The productions of two speakers, one per variety, were chosen as stimuli for the prominence rating task. Reportedly, in the two varieties there is a different probabilistic distribution of prominence features in the post-focal position, with the variety of Bari presenting a higher probability to find F0 movement in the post-focal position (even if compressed) compared to the Udine variety. Crucially, in post-focal position in questions the Bari variety presents F0 movements attributable to pitch accents. Importantly, acoustic features related to the accents occurring in post-focal position of questions are connected with high levels of periodic energy, as shown in the example provided in Chapter 2 (2.3.3) and in the study reported in Chapter 7 (where also high values of PEM are registered). The goal of the prominence rating task and of the comparison of the perceived prominence between the two groups, Udine and Bari, was to test whether these differences in distribution that are found in questions had an influence in the perceptual domain when rating statements. Furthermore, assuming an impact of probabilistic distributional patterns of prominence cues, the ratings of German learners of Italian should pattern with the ones of Udine. However, the assumed different mapping between information status and prominence, and the higher tendency to have prominent elements at the end of the utterance could also lead to differences between these two groups.

In Bari stimuli, the broad focus condition presented F0 movement in a wide range, while the target in post-focal position was characterised by a flat and low contour. In addition, the target in broad focus presented higher values of PEM compared to the target in post-focal position. If the perception of prominence only depended on the acoustic features, results expected for the group of listeners from Bari would be to find low ratings for the target words occurring in post-focal position, specifically lower than words occurring in broad focus. Again, if perception depended only on acoustic features, the expected results would be to find no difference between the Bari group and the learners' group. By contrast, in the stimuli chosen for Udine, targets in broad focus and targets in post-focal position differed only in the F0 movement (F0 excursion was considerably higher in broad focus than in post-focal position) and not in the distribution of PEM values. This circumstance was particularly relevant to disentangle the effects of pitch movement and of cues to prominence not strictly dependent on pitch. If only PEM played a role in perception, there would be no differentiation between broad focus and post-focal conditions in the Udine variety.

Contrary to the predictions made based only on acoustics, results showed that in rating the degree of prominence in the Bari variety, the native Bari listeners differed from the group of learners and from the group of native Udine listeners. Native Bari listeners showed a lack of significance in the difference between prominence assigned to the target in broad focus compared to that assigned to the target in post-focal position, and the learners, by contrast, showed a significant distinction across the conditions. The Bari group also differed compared with the Udine group (where both groups rated their own variety): again, ratings of the Udine group showed a difference in the prominence given to the post-focal position compared to broad focus, while ratings of the Bari group did not. The ratings of the Udine group for the target matched the ones of the learners' group, showing that a same probable distribution of prominences in production yields similar results in perception.

Overall, results were interpreted following the idea that the distribution of prominence features (considering both statement and question modalities) influenced the ratings of Bari participants in building their expectations to find prominence in post-focal position. The fact that Bari listeners perceived the post-focal position with a higher degree of prominence than expected from the characteristics of the acoustic signal, proves the effect of the language-specific distribution of prominence features on the perception of prominence. By contrast, the listeners belonging to the Udine group presumably had a higher expectation of finding no prominence in the post-focal position, because of the more probable mapping of the post-focal region with attenuation, given the flat and low contour characterising the post-focal position both in statements and in questions in this variety. Moreover, the usual distribution of prominence cues found in the production experiment on the Udine variety (Chapter 5), was such that PEM values were lower in the post-focal target than in the target in broad focus. The stimuli listeners were presented with were characterised by a distribution of PEM values in post-focal position that was different from the usual, in that it showed a higher degree of prominence. Nonetheless, participants rated words occurring in this position as lower in prominence than the ones occurring in broad focus. In doing this, participants might either have followed their expectations on the degree of prominence in the post-focal position (i.e. of a low degree of prominence) or followed only the cues to prominence conveyed by F0 movement, or used both strategies together. The same pattern in the rating of prominence was also found for the German learners rating this variety. Although both groups could have followed expectations together with acoustics (top-down and bottom-up inferences respectively), for both groups the hypothesis that the expectations could have played a more consistent role in rating the target word was supported by the low correlation between ratings and PEM found in the last noun of the utterances.

Considering the difference between the groups of the two varieties of Italian, the differences in the comparison between the group of Bari and the group of learners, and the lack of difference between the group of Udine and the group of learners, results seem to undermine the hypothesis that the reported higher tendency of Italian to place prominent information at the end of the phrase compared to West-Germanic languages (see Chapter 2), plays a role in the building of expectations for prominence in the post-focal region. However, it is possible that the impact of this tendency on prominence perception cannot be generalized for Italian as a whole (as shown by the fact that the Udine group does not exhibit high expectations of prominence in the post-focal position, whereas the Bari group appears to do so). Whether expectations can possibly still be affecting Bari Italian (and maybe other varieties as well) needs to be tested by further studies which, in addition, need to involve a systematic and scientifically rigorous analysis of the preferred location of prominent information in Italian, which so far is still missing from the literature. An additional finding of the rating study was that Germans did not perceive differences across the two varieties (Udine and Bari), despite the above-mentioned differences in the relation between PEM and post-focal position. Therefore, they either followed the expectations deriving from their native language regarding the level of prominence in the post-focal position, or were less sensitive to prominence features that did not involve F0 movement.

The ERP study further investigated the influence of top-down and bottom-up inferences in the perception of prominence and in online processing. Moreover, it tried to deepen the understanding of the role of fine-grained cues to prominence on the language processing system. The experiment addressed these issues by testing the online processing of statements and questions in the Bari variety of Italian. Thereby, it investigated the real-time effect of the presence of post-focal cues to prominence that are expressed through pitch movement and rather high levels of PEM. Considering PEM and F0 movement together, the stimuli of the experiment presented a three-way differentiation of acoustic prominence of the target: (i) very high prominence in the contrastive narrow focus condition, (ii) an intermediate level of prominence on the target occurring in post-focal position of questions and (ii) low prominence in the target occurring in the post-focal position of statements.

The processing of the stimuli used in the experiment was tested using the ERP method, which registers the changes in brain activity due to sensory or cognitive processes and measures their variation in time. These changes yield different ERP components (deflections from the base activity). Previous studies have investigated the processing of semantically incongruent stimuli, showing that focally accented words trigger the allocation of attentional resources, consequently engendering a deep processing of the incongruence of the stimulus with the context (Kristensen et al., 2013 for Dutch and Li & Ren, 2012 for Chinese). The same studies have shown that this allocation of attentional resources triggered by focus and accentuation hinders the processing of incongruence in post-focal position (when part of the background and, in the case of Dutch, deaccented). Following these findings, the ERP study reported in this thesis asked whether prominence in post-focal position could redirect attentional resources towards it, and prevent words in this position from being processed shallowly. The effect of the intermediate level of prominence in post-focal position of questions in the Bari variety was, therefore, investigated. The low acoustic prominence of this accent compared to the accents marking focus allowed a second question to be investigated. This second question was related to the effect of more gradual changes in prominence on attention orienting and depth of processing.

General findings confirmed that changes in acoustic prominence, even if fine-grained as in the case of the investigated post-focal prominences, are processed in real time. The main finding indicates that for listeners from Bari, attention is allocated to the post-focal position by default. In fact, the incongruence of targets occurring in this position in statements (characterised by the absence of pitch accent and by low values of PEM), is processed deeply. This was taken as evidence that the distribution of accents in the Bari variety, comprising the high probability of finding features related to a rather high prominence in the post-focal position, may attract attention in this region by default, differently from what is reported for Dutch and Chinese, where the post-focal position is processed more shallowly. Moreover, an additional possibility is that, since the end of the utterance is reported to be a particularly prominent position for Italian, reportedly more prominent that in West-Germanic languages, attention is positioned there by default. These two interpretations can work together in impacting the top-down orienting of attention.

The results of the ERP study support the idea that perception of prominence and processing of prominence are largely influenced by top-down expectations that can be derived by the probabilistic distribution of prominence in the utterance and by the mapping of prominence with information structure. Indeed, the inference arising from where prominence is probabilistically located in the context in a specific language, can ease the processing of semantic information. In the case of Bari Italian, this leads to an ease in the semantic processing of words occurring at the end of the utterance independently from the signal (words in post-focal position as deeply processed as words in narrow focus). The signal and the increased prominence in questions, by contrast, modulates attentional resources, consuming them to update information about sentence modality. Thus, acoustic prominence in questions in the form of a compressed pitch movement associated with the stressed syllable does not lead to deeper semantic processing in this variety. Nonetheless, the attentional resources drawn to the post-focal position of questions were enhanced as a consequence of the modality-specific prosodic cues, as shown by the Late Positivity effects.

Findings of previous studies imply that the interplay between information structure and prosodic marking helps in language processing, in that it facilitates creating inferences about where the important part of the message lies (Akker & Cutler, 2003; Cutler & Fodor, 1979; Wang et al., 2009, among others). Here, the position at the end of the sentence already facilitates the processing of the incongruence, and the higher prominence of the narrow focus condition was not necessary to help in semantic processing. This may indicate that the signal conveying the high prominence of narrow contrastive focus provides extra cues in case of a change in the location of the prominence from the preferred position. Therefore, it would be interesting to investigate stimuli that have the target not at the end of the utterance, but in a preceding position, and see whether an increase of the N400 in narrow focus statements compared to post-focal position (again in statements) is registered. Actually, it could be hypothesised that the same ease in semantic processing of the incongruence found in Chapter 7 on the Bari data could occur in West-Germanic languages, since the end of the utterance is always a prominent position, also in these languages (see the notion of edge placement in Himmelmann & Primus, 2015, and note that the target in previous studies never occurred at the very end of the utterance). However, the ease in the semantic processing of the final element occurring in post-focal position could be smaller for these languages compared to

Bari Italian for two reasons: (i) the reported higher tendency for Italian to move prominent elements at the end of the utterance and (ii) the higher probability of finding post-focal prominences in Bari Italian.

In conclusion, this thesis attempted to address the complexity of gradual changes in the production of prominence, of the interplay between expectation-driven and signal-driven factors in the perception of these gradual changes, and of the effects of these factors on online processing. Taken together, the results of the experiments presented in this thesis (production and perception) indicate that probabilistic mapping has a consistent influence on the perception of prominence (as shown by the high default attention allocated to the post-focal position by Bari listeners) and that acoustic prominence further modulates attention.

This thesis has generated a first set of findings to understand prominence and its mechanisms, and the conclusions drawn are useful to support previous research and to outline a more complete picture on the building up of expectations. However, future research is required to fully understand the interplay of top-down and bottom-up inferences. The results of this thesis should encourage other research on the topic presented here, which could partially overcome the limitations of the present one and deepen other aspects. Questions that could be addressed by future studies are delineated in the next section of this chapter (8.2).

8.2 Future directions

Results of the experiments presented in this thesis constitute a solid basis for future research investigating the production, perception and online processing of post-focal elements.

Future research should focus on the investigation of the distribution of prominence in different focal structures in Italian using (semi-)spontaneous speech instead of read speech (as done in the present thesis). Again, measures referring to periodic energy would prove very useful in the analysis of (semi-)spontaneous speech and this type of speech could foster our understanding of the production of post-focal constituents, both investigating the productions of isolated NPs (along the lines of Swerts et al., 2002) and the production of sentence-long utterances. A collection of a large amount of data could allow researchers to find (semi-)spontaneous productions of words in post-focal position, occurring in different positions in the metrical structure and to have a large enough sample to draw further conclusions on the production of the post-focal position. A similar investigation of (semi-)spontaneous speech using periodic-energy-related measures could be conducted on the production of German, in order to have comparable designs and to extract the same parameters from the data. This would improve the picture on the comparison of the two languages and make a very valuable contribution to the discussion around the difference (or lack of difference) of the mapping between prosody and functions in the two languages.

Since the discussion of probabilistic distributions of prominences assigned a pivotal role to questions in Bari, the perception of the degree of prominence of such questions is in itself interesting. Prominence ratings of words in the post-focal position of questions (performed using a continuous scale) would add to the description of prominence perception in Bari Italian. This investigation would test whether the prominence characteristics found in this signal are indeed prominence lending as claimed in the experiment in Chapter 7. Moreover, it would be interesting to have Udine participants rate the post-focal position in questions realised in the Bari variety, in order to see the effect that a different probabilistic distribution of prominence in the Udine variety has on the perception of post-focal prominence. In addition, the investigation of the online processing of stimuli collected in the Bari variety (similar to the ones of Chapter 7) by listeners form Udine, could deepen the understanding of the results reported in Chapter 7. This would help to more clearly disentangle the effects of expectation-based prominence from signal-based effects. Indeed, listeners coming from Bari have stronger expectations regarding the possibility of enhanced prominence connected to sentence modality in the post-focal position, compared to listeners from Udine. Similarly, it would be interesting to test German learners listening to the variety spoken in Bari. Also in this case German participants would not be familiar with the presence of post-focal rising-falling accents and the unexpected prominence could facilitate the processing of the incongruence, more effectively than reported for Bari participants, who already expect this distribution of prominences. However, the difficulty in this case would be to use words that learners could very easily understand, in order to prevent effects guided by the semantic processing of the non-native language rather than by prosody.

An interesting finding of the ERP study (Chapter 7) was that fine-grained cues to prominence in the post-focal position of questions are perceived and impact language processing. However, these cues are also cues to modality and, in the experiment, they occurred early in the target word (on the second syllable) and not at the end of the word, as it happens when the question exponent is in post-focal position in other languages and varieties (see 2.3.3.1 and 2.3.3.2). Therefore, in the ERP experiment presented in this thesis, the processing is partly diverted from semantics towards modality. It could be hypothesised that if the cue to modality would occur at the end of the word and the prominence features connected to the stressed syllable would be only cues to prominence, resources would not be consumed by modality. In this case the processing of incongruence in this modality could be higher than the one in statements. Future research could deepen this issue. For example, it could adopt stimuli of the Bari variety, in order to have the prominence features on the stressed syllable, but implementing them in order to have at the same time an additional final rise (possible especially in read speech, see Savino, 2012). This would allow listeners to have further cues to modality at the end of the word. In this case, using Udine native speakers could prevent participants from recognising the rising-falling accent as a cue to modality and therefore help disentangle the two types of processing: the effect of prominence on the semantic processing could be registered (N400 effects), while effects connected to modality could be processed only later (registered by a Late Positivity).

Overall, the relation between information status and prosodic realisation and the differences between the languages and varieties presented in the current research should also be investigated further, from both production and perception perspectives. In this thesis measures such as synchrony, scaling and PEM (Albert et al., 2018; Cangemi et al., 2019) have proven to be appropriate for these purposes and should continue to be used in future research. This could aim at using continuous parameters in order to conceptualise differences in prominence relations between the languages and varieties in terms of dynamic systems (see Roessig & Mücke, 2019 and Roessig, Mücke & Grice, 2019 for an example of a dynamical system applied to German data; see Nava, 2010 for an example of a dynamical system applied to the comparison between Spanish and English).

To conclude, this thesis provides a systematic picture on the production, perception and real-time processing of the post-focal position in Italian (focusing on two varieties).

Firstly, it has shown that in Italian speakers use a variety of cues related to F0 changes and gradual changes in energy and duration in order to mark subtle differences in focus structure, which involve gradual changes in prominence from background (i.e. post-focal position), to broad focus, to contrastive focus. Particularly interesting is the finding that the post-focal position of statements is subject to reduction not only of the F0 range, where pitch is characterised by (near) absence of movement, but also of energy and duration in comparison to focal positions (broad focus and contrastive focus). This provides additional evidence to the theoretical view that both categorical and gradual changes are involved in the marking of prominence relations in information structure. Secondly, the present work has shown that inferences are built upon the knowledge of the probabilistic distribution of prominence and that these inferences impact the degree of prominence perceived in the utterance: when words in a language have some probability to be produced as prominent in a particular position in the utterance, the expectation for prominence in that position increases and overrides the signal. In this case, words are perceived as having a higher degree of prominence than the signal conveys. This finding provides further evidence for the theoretical framework of the inference under uncertainty, which implies that listeners infer from the probabilistic distribution of acoustic cues in a language the likelihood of these cues to occur in a given context. Thirdly, this thesis has shown that these kinds of expectation-based inferences have a consequence for online processing: expectation-based factors prevent the listener from missing the part of the utterance in which the important message is thought to occur. At the same time, the processing of the acoustic cues of the signal triggers (re)orienting of attention, which involves the updating of the mental model. Overall, results support the view of acoustic prosodic prominence as characterised by a bundle of cues whose different possible values are probabilistically distributed in the listeners' perceptual space. Top-down expectations corresponding to this perceptual space are then considered to play a crucial role in perception and to modulate general cognitive processes. Both perception and online processing are then further modulated by signal-based factors, which can, however, be overridden by expectations.

Bibliography

Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition*, *6*(2), 81-96. doi: 10.1017/S1366728903001056.

Albert, A. (2018a). Periogramic analysis: On the benefits of periodicity-informed F0 representations (Poster Presentation, 19-22 June). *Laboratory Phonology XVI*, 203-204. doi: 10.13140/RG.2.2.13435.62246.

Albert, A. (2018b). A tonal conspiracy: A perceptually-motivated acoustic model of prosodic prominence (Oral Presentation). *Second International Conference Prominence in Language* (11-13 July), 26-27.

Albert, A., & Cangemi, F. (2020). Periogram Projekt, GitHub repository, https://github.com/finkelbert/Periogram-Projekt. (Last access date August 2020).

Albert, A., Cangemi, F. & Grice, M. (2018). Using periodic energy to enrich acoustic representations of pitch in speech: A demonstration. In Klessa, K., Bachan, J., Wagner, A., Karpiński, M. & Śledziński, D. (Eds.), *Proceedings of the 9th International Conference on Speech Prosody* (pp. 804-808). University Poznań.

Anderson, A., Bader, M., Bard, E., G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., MacAllister, J., Miller, J., Sotillo, C., Thompson, H., Weinert, R. (1991). The HCRC Map Task Corpus, *Language and Speech*, 34 (4):351-366. doi: 10.1177/002383099103400404.

Andreeva, B., Barry, W. J. & Koreman, J. (2014). A cross-language corpus for studying the phonetics and phonology of prominence. In Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J. & Piperidis, S. (Eds.), *Proceedings of the 9th International Conference on Language Resources and Evaluation* (pp. 326-330). European Language Resources Association.

Antinucci, F. & Cinque, G. (1977). Sull'ordine delle parole in italiano: l'emarginazione. *Studi di Grammatica Italiana*, 6, 121-146.

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and cognitive processes*, *23*(4), 495-527. doi: 10.1080/01690960801920099.

Arnold, J. E., Kaiser, E., Kahn, J. M., & Kim, L. K. (2013). Information structure: linguistic, cognitive, and processing approaches. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(4), 403-413. doi: 10.1002/wcs.1234.

Arnold, D., Wagner, P., & Baayen, R. H. (2013). Using generalized additive models and random forests to model prosodic prominence in German. In Bimbot, F. (Ed.), *Proceedings of the 14th Annual Conference of the International Speech Communication Association* (pp. 272-276). International Speech Communication Association.

Arnold, D., Wagner, P., Möbius, B. (2010). The effect of priming on the correlations between prominence ratings and acoustic features. In M. Hasegawa-Johnson (Ed.), *Proceedings of the 5th International Conference of Speech Prosody* (paper 2003). International Speech Communication Association.

Arnold, D., Wagner, P., & Möbius, B. (2011). Evaluating different rating scales for obtaining judgments of syllable prominence from naïve listeners. In W. Lee & E. Zee (Eds.), *Proceedings of the17th International Congress of Phonetic Sciences* (pp. 252-255). City University of Hong Kong.

Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics*, (26), 3-25. doi: 10.1006/jpho.1997.0063.

Arvaniti, A., Ladd, D. R., & Mennen, I. (2000). What is a starred tone? Evidence from Greek. *Papers in laboratory phonology V: Acquisition and the lexicon*, 119-131.

Arvaniti, A. (2011). Segment-to-Tone Association. In Cohn, A., Fougeron, C., & Huffman, M. (Eds.), *Handbook of Laboratory Phonology* (pp. 265-274). Oxford University Press.

Astruc, L., & Prieto, P. (2006). Acoustic cues of stress and accent in Catalan. In R. Hoffmann & H. Mixdorff (Eds.), *Proceedings of the 3rd International Conference on Speech Prosody* (pp. 337-340). Verlag der Wissenschaften GmbH. doi: 10.36505/ExLing-2006/01/0012/000012.

Avesani, C. (1990). A contribution to the synthesis of italian intonation, *Proceedings of the 1st International Conference on Spoken Language Processing*, (1), 833-836.

Avesani, C. (1995). ToBIt: un sistema di trascrizione per l'intonazione italiana. *Atti delle 5e Giornate di Studio del Gruppo di Fonetica Sperimentale (AIA)*, 85-98.

Avesani, C. (1997). I toni della RAI. Un esercizio di lettura intonativa. In Accademia delle Crusca (Ed.), *Gli italiani trasmessi: la radio* (pp. 659-727). Accademia della Crusca.

Avesani, C., Bocci, G., Vayra, M., Zappoli, A. (2015). Prosody and information status in Italian and German L2 intonation. In M., Chini (Ed.), *Il parlato in [italiano] L2: aspetti pragmatici e prosodici /[Italian] L2 Spoken Discourse: Pragmatic and Prosodic Aspects* (pp. 93-116). Franco Angeli.

Avesani, C., Hirschberg, J., & Prieto, P. (1995). The intonational disambiguation of potentially ambiguous utterances in English, Italian and Spanish. In K. Elenius & P. Branderud (Eds.), *Proceedings of the 13th International Congress of Phonetic Sciences* (pp. 174-177). Stockholm University. doi: 10.1007/978-94-011-4317-2_4.

Avesani, C., & Vayra, M. (2003). Broad, narrow and contrastive focus in Florentine Italian. In M. J. Solé, D. Recasens & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (Volume 2, pp. 1803-1806). Causal Productions.

Avesani, C., & Vayra, M. (2005). Accenting, deaccenting and information structure in Italian dialogue. In L. Dybkjær & W. Minker (Eds.), *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue* (pp. 19-24). Pennsylvania State University.

Avesani, C., Vayra, M., & Zmarich, C. (2007). On the articulatory bases of prominence in Italian. In J. Trouvain (Ed.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 981-984). Universität des Saarlandes.

Ayers, G. M. (1996). *Nuclear accent types and prominence: Some psycholinguistic experiments* (PhD Thesis). The Ohio State University.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31-56. doi: 10.1177/00238309040470010201.

Baker, R. E. (2010). Non-native perception of native English prominence. *Speech Prosody.*

Bard, E., & Aylett, M. (1999). The dissociation of deaccenting, givenness, and syntactic role in spontaneous speech. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. & A.C. Bailey (Eds.), *Proceedings of the 14th Iternational Congress of Phonetic Sciences* (Volume 3, pp. 1753–1756). University of California.

Barnes, J., Brugos, A., Shattuck-Hufnagel, S. & Veilleux, N. (2011). Voiceless intervals and perceptual completion in F0 contours: Evidence from scaling perception in American English. In W. Lee & E. Zee (Eds.), *Proceedings of the17th International Congress of Phonetic Sciences* (Volume 3, pp. 108-111). City University of Hong Kong.

Barnes, J., Veilleux, N. Brugos, A., Shattuck-Hufnagel, S. (2012). Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology,* 3(2), 337-383. doi: 10.1515/lp-2012-0017.

Bates, D.M., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Batliner, A., Nöth, E., Buckow, J., Huber, R., Warnke, V., & Niemann, H. (2001). Duration features in prosodic classification: why normalization comes second, and what they really encode. In M. Bacchiani, J. Hirschberg, D. Litman, M. Ostendorf (Eds.), *Proceedings of the Workshop on Prosody and Speech Recognition* (pp. 23-28).

Baumann, S. (2006). *The Intonation of Givenness - Evidence from German* (PhD thesis) *Linguistische Arbeiten* 508. Niemeyer.

Baumann, S. (2014). The importance of tonal cues for untrained listeners in judging prominence. In S. Fuchs, M. Grice, L. Lancia & D. Mücke (Eds.), *Proceedings of the 10th International Seminar on Speech Production* (pp. 21-24). Universität zu Köln.

Baumann, S. (2016). Second occurrence focus. In C. Féry & S. Ischihara (Eds.), *The Oxford handbook of information structure* (pp. 483-502). Oxford University Press. doi: 10.1093/oxfordhb/9780199642670.001.0001.

Baumann, S., Becker, J., Grice, M., & Mücke, D. (2007). Tonal and articulatory marking of focus in German. In J. Trouvain (Ed.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1029-1032). Universität des Saarlandes.Baumann, S. & Grice, M. (2004). Accenting Accessible Information. In B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody 2004* (pp. 21-24). Speech Prosody Special Interest Group.

Baumann, S., & Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics*, *38*(10), 1636-1657. doi: 10.1016/j.pragma.2005.03.017.

Baumann, S., Grice, M., & Steindamm, S. (2006). Prosodic marking of focus domains-categorical or gradient. In R. Hoffmann & H. Mixdorff (Eds.), *Proceedings of the International Conference on Speech Prosody* (pp. 301-304). TUD press.

Baumann, S., & Hadelich, K. (2003). Accent type and givenness: an experiment with auditory and visual priming. In M. J. Solé, D. Recasens & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1811-1814). Causal Productions.

Baumann, S., Mücke, D., & Becker, J. (2010). Expression of second occurrence focus in German. *Linguistische Berichte*, *(221)*, 61-78.

Baumann, S., & Riester, A. (2012). Referential and Lexical Givenness: Semantic, Prosodic and Cognitive Aspects, In G. Elordieta & P. Prieto (Eds.), *Prosody and Meaning* (Volume 25, pp. 119-162). De Gruyter. doi: 10.1515/9783110261790.119.

Baumann, S., & Riester, A. (2013). Coreference, lexical givenness and prosody in German. *Lingua*, (*136*), 16-37. doi: 10.1016/j.lingua.2013.07.012.

Baumann, S., & Röhr, C. (2015). The perceptual prominence of pitch accent types in German. In H. Little (Ed.), *Proceedings of the 18th International Congress of the Phonetic Sciences* (pp. 298-303). University of Glasgow.

Baumann, S., & Schumacher, P. B. (2012). (De-) accentuation and the processing of information status: evidence from event-related brain potentials. *Language and Speech*, *55*(3), 361-381.

Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics*, *70*, 20-38.

Beaver, D. I., Clark, B., Flemming, E. S., Jaeger, T. F., & Wolters, M. (2007). When semantics meets phonetics: Acoustical studies of second-occurrence focus. *Language*, *83*(2), 245-276.

Beckman, M. E. (1986). *Stress and Non-Stress Accent*. De Gruyter. doi: 10.1515/9783110874020.

Beckman, M. E. (1996). The parsing of prosody. *Language and cognitive processes*, *11*(1-2), 17-68.

Beckman, M. E., & Ayers G. (1997). *Guidelines for ToBI labelling*. The Ohio State University.

Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. *Papers in Laboratory Phonology*, *3*, 7-33.

Beckman, M., & Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. *Papers in Laboratory Phonology II: Segment, Gesture, and Prosody*, 68-86.

Beckman, M. E. & Hirschberg, J. (1994). *The ToBI Annotation Conventions*. Ohio State University.

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

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., & Gildea, D. (1999). Forms Of English Function Words--Effects Of Disfluencies, Turn Position, Age And Sex, And Predictability. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 395-398). University of California.

Benincà P., Salvi, G. & Frison, L. (1988). L'ordine degli elementi della frase e le costruzioni marcate. In L. Renzi, (Ed.), *Grande Grammatica Italiana di Consultazione*. Il Mulino.

van Berkum, J. J., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*(2), 147-182.

van Berkum, J. J., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, *17*(3), 701-718. doi: 10.1006/jmla.1999.2641.

Berruto, G. (2007). Miserie e grandezze dello standard. La nozione di standard, non standard, substandard in linguistica e sociolinguistica. In P. Molinelli, (Ed.), *Standard e non standard tra scelta e norma*. *Atti Convegno della SIG* (pp. 13-41). Il Calamo.

Besson, M., Faita, F., Czternasty, C., & Kutas, M. (1997). What's in a pause: event-related potential analysis of temporal disruptions in written and spoken sentences. *Biological Psychology*, *46*(1), 3-23.

Birch, S., & Clifton, C. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech*, *38*(4), 365-391. doi: 10.1177/002383099503800403.

Birch, S., & Rayner, K. (1997). Linguistic focus affects eye movements during reading. *Memory & Cognition*, *25*(5), 653-660. doi: 10.3758/BF03211306.

Bird, S., & Ladd, D. R. (1991). Presenting autosegmental phonology. *Journal of Linguistics*, *27*(1). doi: 193-210, 10.1017/S0022226700012469.

Bishop, J. (2012). Information structural expectations in the perception of prosodic prominence. In G. Elordieta & P. Prieto (Eds.), *Prosody and Meaning* (Volume 25, pp. 239-270). doi: 10.1515/9783110261790.239.

Bishop, J. (2016). Individual differences in top-down and bottom-up prominence perception. In J. Barnes, A. Burgos, S. Shattuck-Hufnagel & N. Veilleux (Eds.), *Proceedings of the 8th International Conference of Speech Prosody* (pp. 668-672). International Speech Communication Association. doi: 10.21437/speechprosody.2016-137.

Bocci, G. (2013). *The Syntax Prosody Interface: A cartographic perspective with evidence from Italian* (Vol. 204). John Benjamins Publishing.

Bocci, G., & Avesani, C. (2011). Phrasal prominences do not need pitch movements: postfocal phrasal heads in Italian. In P. Cosi, R. De Mori, G. Di Fabrizio & R. Pieraccini (Eds.), *Proceedings of the 12th Annual Conference of the International Speech Communication Association* (pp. 1357–1360). Curran Associates.

Bock, J. K., & Mazzella, J. R. (1983). Intonational marking of given and new information: Some consequences for comprehension. *Memory & Cognition*, *11*(1), 64-76. doi: 10.3758/BF03197663.

Boersma, P., & Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Version 6.1.16, retrieved 6 June 2020 from http://www.praat.org/.

Bolinger, D. L. (1954). English prosodic stress and Spanish sentence order. *Hispania*, *37*(2). doi: 152-156, 10.2307/335628.

Bolinger, D. L. (1958). A theory of pitch accent in English. *Word*, *14*(2-3), 109-149.

Bornkessel, I., Schlesewsky, M., & Friederici, A. D. (2003). Contextual information modulates initial processes of syntactic integration: The role of inter-versus intrasentential predictions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 871. doi: 10.1037/0278-7393.29.5.871.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology*, *10*, Article 298. doi: 10.3389/fpsyg.2019.00298.

Bornkessel-Schlesewsky, I. & Schumacher, P. B. (2016). Towards a Neurobiology of Information Structure. In C. Féry, & S. Ishihara, (Eds.) *The Oxford Handbook of Information Structure* (pp. 581-598). Oxford University Press.

Braun, B. (2006). Phonetics and phonology of thematic contrast in German. *Language and Speech*, *49*(4), 451-493, https://doi.org/10.1177/00238309060490040201.

Brazil, D., Coulthard, M. & Johns, C. (1980). *Discourse Intonation and Language Teaching*. Longman.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7-9), 1044-1098, https://doi.org/10.1080/01690965.2010.504378.

Bredart, S. & Modolo, K. (1998). Moses strikes again: focalization effects on a semantic illusion, *Acta Psychologica*, *67*(2), 135-144.

Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound* (PhD Thesis). MIT Press.

Brouwer, H., & Hoeks, J. C. (2013). A time and place for language comprehension: mapping the N400 and the P600 to a minimal cortical network. *Frontiers in Human Neuroscience*, *7*, Article 758. doi: 10.3389/fnhum.2013.00758.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*(1), 34-44. doi: 10.1162/jocn.1993.5.1.34.

Bruce, G. (1977). *Swedish word accents in sentence perspective* (Vol. 12). Liber Läromedel/Gleerup.

Bruggeman, A., Cangemi, F., Wehrle, S., El Zarka, D., & Grice, M. (2017). Unifying speaker variability with the Tonal Centre of Gravity. In M. Belz, C. Mooshammer, S. Fuchs, S. Jannedy, O. Rasskazova & M. Žygis (Eds.), *Proceedings of Phonetik und Phonologie im deutschsprachigen Raum* (Volume 13, pp. 21-24). Humboldt-Universität zu Berlin. doi: 10.18452/18805.

Brugos, A., Veilleux, N., Breen, M., & Shattuck-Hufnagel, S. (2008). The Alternatives (Alt) tier for ToBI: advantages of capturing prosodic ambiguity. In P. Barbosa, S. Madureira & C. Reis (Eds.), *Proceedings of the 4th International Conference on Speech Prosody* (pp. 273-276). International Speech Communication Association.

Büring, D. (2007). Semantics, intonation and information structure. In Ramchand, G. & Reiss, C. (Eds.) *The Oxford handbook of linguistic interfaces* (pp. 1-32), Oxford University Press. doi: 10.1093/oxfordhb/9780199247455.013.0015.

Büring, D. (2009). Towards a typology of focus realization. In M. Zimmermann & C. Féry (Eds.), *Information structure: Theoretical, Typological, and Experimental Perspectives* (pp. 177-205). Oxford University Press. doi: 10.1093/acprof:oso/9780199570959.001.0001.

Büring, D. (2016). *Intonation and Meaning*. Oxford University Press. doi: 10.1093/acprof:oso/9780199226269.001.0001.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, *98*(2), 159-168. doi: 10.1016/j.bandl.2006.04.005.

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., & Keller, P. (1996). Perception of lexical tone across languages: Evidence for a linguistic mode of processing. In H. T. Bunnell & W. Idsardi (Eds.), *Proceeding of the Fourth International Conference on Spoken Language Processing* (pp. 2514-2517). Institute of Electrical and Electronic Engineers.

Baumann, S., & Cangemi, F. (2020). Integrating phonetics and phonology in the study of linguistic prominence. *Journal of Phonetics*, *81*, 1-6. doi: 10.1016/j.wocn.2020.100993.

Calhoun, S. (2007). *Information structure and the prosodic structure of English: A probabilistic relationship* (Doctoral dissertation, University of Edinburgh).

Calhoun, S., Nissim, M., Steedman, M., & Brenier, J. (2005). A framework for annotating information structure in discourse. In A. Meyers (Ed.), *Proceedings of the workshop on Frontiers in Corpus Annotations II: Pie in the Sky* (pp. 45-52). https://doi.org/10.3115/1608829.1608836.

Campbell, W. N. (1993). Durational cues to prominence and grouping. In D. House & P. Touati (Eds.), *Proceedings of an European Speech Communication Association Workshop on Prosody* (pp. 38-41). European Speech Communication Association.

Campbell, N. (1995). Loudness, spectral tilt, and perceived prominence in dialogues. In K. Elenius & P. Branderud (Eds.), *Proceedings of the 13th International Congress of Phonetic Sciences* (pp. 676-679). Stockholm University.

Canepari, L. (1980). *Italiano standard e pronunce regionali*. Cooperativa libraria editrice degli studenti dell'Università di Padova.

Cangemi, F. (2014). *Prosodic detail in Neapolitan Italian*. Language Science Press. doi: 10.26530/oapen_533874.

Cangemi, F., Albert, A., & Grice, M. (2019). Modelling intonation: Beyond segments and tonal targets. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 572-576). Australasian Speech Science and Technology Association Inc.

Cangemi, F. & D'Imperio, M. (2013). Tempo and the perception of sentence modality. *Journal of the Association for Laboratory Phonology, 4 (1),* 191-219. doi: 10.1515/lp-2013-0008.

Cangemi, F., El Zarka, D., Wehrle, S., Baumann, S., & Grice, M. (2016). Speaker-specific intonational marking of narrow focus in Egyptian Arabic. In J. Barnes, A. Burgos, S. Shattuck-Hufnagel & N. Veilleux (Eds.), *Proceedings of the 8th International Conference on Speech Prosody* (pp. 335-339), International Speech Communication Association. doi: 10.21437/speechprosody.2016-69.

Cangemi, F., Krüger, M., & Grice, M. (2015). Listener-specific perception of speaker-specific production in intonation. In S. Fuchs, D. Pape, C. Petrone & P. Perrier (Eds.), *Individual differences in speech production and perception* (pp. 123-145). Peter Lang.

Cangemi, F., & Grice, M. (2016). The importance of a distributional approach to categoriality in autosegmental-metrical accounts of intonation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *7*(1), Article 9. doi: 10.5334/labphon.28.

Caputo, M., R. (1997). *Le domande in un corpus di italiano parlato. Analisi prosodica e pragmatica* (PhD Thesis). University of Naples "Federico II".

Caputo, M. R., & D'Imperio, M. (1995). Verso un possibile sistema di trascrizione prosodica dell'italiano: cenni preliminari. *Atti delle IV Giornate di Studio del Gruppo di Fonetica Sperimentale*, 71-83.

Chafe, W. L. (1974). Language and consciousness. *Language*, 111-133.

Chafe, W. L. (1976). Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View. In C. Li (Ed.), *Subject and Topic* (pp. 25-56), Academic Press.

Chen, A., Rietveld, T., & Gussenhoven, C. (2001). Language-specific effects of pitch range on the perception of universal intonational meaning. In P. Dalsgaard, B. Lindberg, H. Brenner & Z. Tan (Eds.), *Proceedings of the 7th European Conference on Speech Communication and Technology* (pp. 1403-1406), International Speech Communication Association.

Chodroff, E., Arthurs, A., Kurian, P., Pazol, J., & Cole, J. (2019). Categorical and gradient effects of information structure on nuclear prominence in American English, (Presentation, January 3-6, 2019. Linguistic Society of America.

Chodroff, E., & Cole, J. S. (2018). Information structure, affect, and prenuclear prominence in American English. *Proceedings of the Annual Conference of the International Speech Communication Association,*1848-1852.

Chodroff, E. R., & Cole, J. (2019). The phonological and phonetic encoding of information status in American English nuclear accents. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 1570-1574). The University Of New York.

Chong, A. J., & German, J. S. (2017). The accentual phrase in Singapore English. *Phonetica*, *74*(2), 63-80. doi: 10.1159/000447429.

Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, *32*(3), 274-285. doi: 10.1111/j.1469-8986.1995.tb02956.x.

Clark, H. H. & Haviland S. E. (1977). Comprehension and the given-new contract. In R. Freedle (Ed.), *Discourse Production and Comprehension* (pp. 1-40). Ablex.

Clark, H. H. & Marshall, C. R. (1981). Definite reference and mutual knowledge. In J. Aravind, B. Webber & I. Sag (Eds.), *Elements of Discourse Understanding* (pp. 10-63), Cambridge University Press.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804-809.

CLIPS (2004), Corpora e Lessici di Italiano Parlato e Scritto, http://www.clips.unina.it/it/ (Last access date 10.10.2017).

Cole, J., Goldstein, L., Katsika, A., Mo, Y., Nava, E., & Tiede, M. (2008). Perceived prosody: Phonetic bases of prominence and boundaries. *The Journal of the Acoustical Society of America*, *124*(4), 2496-2496.

Cole, J., Hualde, J. I., Smith, C. L., Eager, C., Mahrt, T., & de Souza, R. N. (2019). Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics*, *75*, 113-147.

Cole, J., Kim, H., Choi, H., & Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *Journal of Phonetics*, *35*(2), 180-209.

Cole, J., Mahrt, T., & Hualde, J. I. (2014). Listening for sound, listening for meaning: Task effects on prosodic transcription. In N. Campbell, D. Gibbon & D. Hirst (Eds.), *Proceedings of the 7th International Conference on Speech Prosody* (pp. 859-863), International Speech Communication Association. doi:10.21437/speechprosody.2014-160.

Cole, J., Mo, Y., & Baek, S. (2010a). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes*, *25*(7-9), 1141-1177. doi: 10.1080/01690960903525507.

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010b). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, *1*(2), 425-452. doi:10.1515/labphon.2010.022.

Cole, J., & Shattuck-Hufnagel, S. (2016). New methods for prosodic transcription: Capturing variability as a source of information. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *7*(1). doi:10.5334/labphon.29.

Cole, R. A., Jakimik, J., & Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *The Journal of the Acoustical Society of America*, *64*(1), 44-56.

Connolly, J. F., Stewart, S. H., & Phillips, N. A. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language*, *39*(2), 302-318. doi:10.1016/0093-934X(90)90016-A.

Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language*, *43*(1), 1-18.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6*(3), 256-266.

Contreras, H. (1980), Sentential stress, word order and the notion of subject in Spanish. In. Waugh, L., R. & van Schoonevelds, C., H. (Eds.), *The Melody of Language* (pp. 45-53). University Park Press.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, *3*(3), 201-215. doi:10.1038/nrn755.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*(1), 21-58.

Coulson, S., & van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory & Cognition*, *30*(6), 958-968.

Cruttenden, A. (2006). The de-accenting of given information: A cognitive universal. In Bernini, G., & Schwartz, M. L. (Eds.), *Pragmatic Organization of Discourse in the Languages of Europe* (pp. 311-355). Mouton de Gruyter.

Cruz-Ferreira, M. (1984). Perception and interpretation of non-native intonation patterns. In A. Cohen & M. P. R. van den Broecke (Eds.), *Proceedings of the 10th International Congress of Phonetic Sciences* (pp. 565-569). Mouton de Gruyter. doi: 10.1515/9783110884685.

Culicover, P. W., & Rochemont, M. (1983). Stress and focus in English. *Language*, 123-165. doi: 10.2307/414063.

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics*, *20*(1), 55-60.

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*(2), 141-201. doi: 10.1177/002383099704000203.

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, *7*(1), 49-59.

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech*, *20*(1), 1-10.

Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge University Press.

Crystal, D. (1970). Prosodic Systems and Language Acquisition. In P. R. Léon, G. Faure, & A. Rigault (Eds.), *Prosodic Feature Analysis* (pp. 77–90). Didier.

D'agostin, R., & Romano, A. (2007). Analisi dell'intonazione delle varietà friulana e italiana parlate a Codroipo (Udine). *La prosodia en el ámbito lingüístico románico,* 101-121. La página.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*(2), 292-314. doi:10.1016/S0749-596X(02)00001-3.

Dahmen, S. (2013). *Prosodie oder Segmente? Phonetische Untersuchungen zu Trainingseffekten bei italienischen Deutschlernenden*, (PhD Thesis). University of Cologne.

Del Giudice, A., Shosted, R., Davidson, K., Salihie, M., & Arvaniti, A. (2007). Comparing methods for locating pitch "elbows". In J. Trouvain (Ed.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1117-1120). Universität des Saarlandes.

De Grauwe, S., Swain, A., Holcomb, P. J., Ditman, T., & Kuperberg, G. R. (2010). Electrophysiological insights into the processing of nominal metaphors. *Neuropsychologia*, *48*(7), 1965-1984. doi: 10.1016/j.neuropsychologia.2010.03.017.

De Ruiter, L. E. (2015). Information status marking in spontaneous vs. read speech in story-telling tasks–Evidence from intonation analysis using GToBI. *Journal of Phonetics*, *48*, 29-44. doi: 10.1016/j.wocn.2014.10.008.

Dilley, L. & Heffner, C. (2013): The role of f0 alignment in distinguishing intonation categories. evidence from American English. *Journal of Speech Sciences, 3*, 3-67.

Dimitrova, D. V., Stowe, L. A., Redeker, G., & Hoeks, J. C. (2012). Less is not more: Neural responses to missing and superfluous accents in context. *Journal of cognitive neuroscience*, *24*(12), 2400-2418. doi: 10.1162/jocn_a_00302.

D'Imperio, M. (1995). Timing differences between prenuclear and nuclear pitch accents in Italian. *Journal of the Acoustical Society of America*, 98(5), 2894–2894. doi:10.1121/1.414392.

D'Imperio, M. (1997). Breadth of focus, modality and prominence perception in Neapolitan Italian. In A. Kim, & M. D'Imperio (Eds.), *The Ohio State University Working Papers in Linguistics – Papers from the Linguistics Laboratory* (Volume 50, pp. 19–39). Ohio State University.

D'Imperio, M. (1998). Acoustic-perceptual correlates of sentence prominence in Italian questions and statements. *The Journal of the Acoustical Society of America*, *104*(3), 1779-1779. doi: 10.1121/1.424133.

D'Imperio, M. (2000a). Acoustic-perceptual correlates of sentence prominence in Italian, *The Ohio State University Working Papers in Linguistics – Papers from the Linguistics Laboratory*, (52), 59-79.

D’Imperio, M. (2000b). *The role of perception in defining tonal targets and their alignment* (PhD Thesis). The Ohio State University.

D’Imperio, M. (2001). Focus and tonal structure in Neapolitan Italian. *Speech Communication*, *33*(4), 339-356. doi: 10.1016/S0167-6393(00)00064-9.

D’Imperio, M. (2002). Italian intonation: An overview and some questions. *Probus*, *14*(1), 37-69. doi:10.1515/prbs.2002.005.

D’Imperio, M. (2003). Tonal structure and pitch targets in Italian focus constituents. *Catalan Journal of Linguistics*, *2*, 55-65. doi: 10.5565/rev/catjl.43.

D’Imperio, M. (2011). Tonal Alignment. In A. Cohn, C. Fougeron, & M. Huffman (Eds.), *Handbook of Laboratory Phonology* (pp. 275-287). Oxford University Press.

Endo, R., & Bertinetto, P. M. (1997). Aspetti dell'intonazione in alcune varietà dell'italiano. In F. Cutugno (Ed.): *Fonetica e fonologia degli stili dell'italiano parlato. Atti delle VII Giornate di Studio del GFS* (pp. 27-49). Esagrafica.

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, 20(5), 540-551. doi: 10.1016/S0022-5371(81)90165-1.

Eriksson, A., Grabe, E., & Traunmüller, H. (2002). Perception of syllable prominence by listeners with and without competence in the tested language. In B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody* (pp. 1-4). Laboratoire Parole et Langage.

Eriksson, A., Thunberg, G. C., & Traunmüller, H. (2001). Syllable prominence: A matter of vocal effort, phonetic distinct-ness and top-down processing. In P. Dalsgaard, B. Lindberg, H. Brenner & Z. Tan (Eds.), *Proceedings of the 7th European Conference on Speech Communication and Technology* (pp. 399-402). International Speech Communication Association.

Face T., L. (2000). Prosodic manifestations of focus in Spanish. *Southwest Journal of Linguistics* 19, 45-62.

Face, T. L., & D’Imperio, M. (2005). Reconsidering a focal typology: Evidence from Spanish and Italian. *Italian Journal of Linguistics*, *17*(2), 271-289.

Fanselow, G., & Lenertová, D. (2011). Left peripheral focus: Mismatches between syntax and information structure. *Natural Language & Linguistic Theory*, *29*(1), 169-209. doi: 10.1007/s11049-010-9109-x.

Farnetani, E., & Zmarich, C. (1997). Prominence patterns in Italian: an analysis of F0 and duration. *Proceedings of an European Speech Communication Association Workshop on Intonation* (pp. 115-118). European Speech Communication Association.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. *Current Directions in Psychological Science*, *11*(1), 11-15. doi: 10.1111/1467-8721.00158.

Féry, C. (1993). *German Intonational Patterns*. Niemeyer.

Féry, C., & Krifka, M. (2008). Information structure: Notional distinctions, ways of expression. In P. van Sterkenburg (Ed.), *Unity and diversity of languages* (pp. 123-136). John Benjamins. doi:10.1075/z.141.13kri.

Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, *36*(4), 680-703. doi:10.1016/j.wocn.2008.05.001.

Finco, F. (2007). Fonetiche e fonologjie. In F. Fabbro (Ed.), *Manuâl di lenghistiche furlane* (pp. 47-82). Forum Edizioni.

Fox J., & Weisberg, S. (2019). *CAR - An R Companion to Applied Regression*. Sage publications.

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language*, *63*(3), 367-386. doi: 10.1016/j.jml.2010.06.004.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, *1*(3), 183-192. doi:10.1016/0926-6410(93)90026-2.

Friedrich, C. K., Alter, K., & Kotz, S. A. (2001). An electrophysiological response to different pitch contours in words. *NeuroReport*, *12*(15), 3189-3191. doi: 10.1097/00001756-200110290-00009.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, *1*(2), 126-152. doi: 10.1177/002383095800100207.

Gagliardi, G., Vallauri, E. L., & Tamburini, F. (2012). La prominenza in Italiano: demarcazione piu' che culminazione. In M. Falcone, & A. Paoloni (Eds.) *La voce nelle applicazioni: VIII Convegno Nazionale dell'Associazione Italiana di Scienze della Voce* (pp. 255-270), Bulzoni.

Gili-Fivela, B. (2002). Tonal alignment in two Pisa Italian peak accents. In B. Bel, & I. Marlien, (Eds.), *Proceedings of the Speech Prosody 2002 Conference* (pp. 339-342), Laboratoire Parole et Langage.

Gili-Fivela, B. (2003). *The Phonetics and Phonology of Intonation: the Case of Pisa Italian*. (PhD Thesis). Scuola Normale Superiore di Pisa).

Gili-Fivela, B., Avesani, C., Barone, M., Bocci, G., Crocco, C., D'Imperio, M., Giordano, R., Marotta, G., Savino, M., & Sorianello, P. (2015). Intonational phonology of the regional varieties of Italian. In S. Frota & P. Prieto (Eds.), *Intonation in Romance* (pp. 140-197), Oxford University Press. doi:10.1093/acprof:oso/9780199685332.003.0005

Gisladottir, R. S., Chwilla, D. J., Levinson, S. C. (2015). Conversation Electrified: ERP Correlates of Speech Act Recognition in Underspecified Utterances. *PLoS ONE*, *10*(3). doi: 10.1371/journal.pone.0120068.

Goldsmith, J. A. (1976). An Overview of Autosegmental Phonology. *Linguistic Inquiry 2*, 23-68.

Goldstein, E. B. (2010). *Encyclopedia of Perception*. SAGE. doi:10.4135/9781412972000.

Grabe, E., Rosner, B. S., García-Albea, J. E., & Zhou, X. (2003). Perception of English intonation by English, Spanish, and Chinese listeners. *Language and Speech*, *46*(4), 375-401. doi:10.1177/00238309030460040201.

Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*(267), 1-13. doi:10.3389/fnins.2013.00267.

Gray, H. M., Ambady, N., Lowenthal, W. T., & Deldin, P. (2004). P300 as an index of attention to self-relevant stimuli. *Journal of Experimental Social Psychology*, *40*(2), 216-224. doi: 10.1016/S0022-1031(03)00092-1.

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society*, *35*, 151-166.

Grice, M. (1995). *The intonation of Interrogation in Palermo Italian. Implications for intonation theory*. Mouton De Gruyter. doi:10.1515/9783110932454.

Grice, M. (2006). *Intonation*. In K. Brown (Ed.). *Encyclopedia of Language and Linguistics* (pp. 778-788), Elsevier.

Grice, M. & Baumann, S. (2000). Deutsche Intonation und GToBI. *Linguistische Berichte* 181, 267-298.

Grice, M., & Baumann, S. (2007). An introduction to intonation – functions and models. In J. Trouvain & U. Gut (Eds.), *Non-Native Prosody: Phonetic description and teaching practice* (pp. 25-52), De Gruyter Mouton. doi:10.1515/9783110198751.

Grice, M, Baumann, S. & Benzmüller, R. (2005). German Intonation in Autosegmental-metrical Phonology. In S. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 55-83), Oxford University Press. doi: 10.1093/acprof:oso/9780199249633.003.0003.

Grice, M., D'imperio, M., Savino, M., & Avesani, C. (2005). Strategies for intonation labelling across varieties of Italian. In S. Jun (Ed.), *Prosodic typology: The Phonology of Intonation and Phrasing* (pp. 362-389), Oxford University Press. doi:10.1093/acprof:oso/9780199249633.003.0013.

Grice, M., Ladd, D. R., & Arvaniti, A. (2000). On the place of phrase accents in intonational phonology. *Phonology*, *17*(2), 143-185. doi:10.1017/S0952675700003924.

Grice, M., Ritter, S., Niemann, H., & Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics*, *64*, 90-107. doi:10.1016/j.wocn.2017.03.003.

Grice, M., & Savino, M. (2003). Map tasks in Italian: Asking questions about given, accessible and new information. *Catalan Journal of Linguistics*, *2*, 153-180. doi:10.5565/rev/catjl.48.

Grice, M., Savino, M., & Roettger, T. B. (2018). Word final schwa is driven by intonation—The case of Bari Italian. *The Journal of the Acoustical Society of America*, *143*(4), 2474-2486. doi:10.1121/1.5030923.

Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational linguistics*, *12*(3), 175-204.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69(2), 274-307. doi:10.2307/416535.

Gussenhoven, C. (1983). Testing the reality of focus domains. *Language and Speech*, 26(1), 61–80. doi:10.1177/002383098302600104.

Gussenhoven, C. (2002). Phonology of intonation. *Glot International*, *6*(9/10), 271-284.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge University Press. doi:10.1017/CBO9780511616983.

Gussenhoven, C. & Rietveld, A., C., M. (1988). Fundamental Frequency Declination in Dutch: Testing Three Hypotheses. *Journal of Phonetics*, *16*, 335-369. doi: 10.1016/s0095-4470(19)30509-1.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, *38*(11), 1531-1549. doi:10.1016/S0028-3932(00)00053-1.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438-441. doi:10.1126/science.1095455.

Halliday, M. (1967). *Intonation and grammar in British English*. The Hague. doi:10.2307/411669.

Harrington, J., Beckman, M. E., Fletcher, J., & Palethorpe, S. (1998). An electropalatographic, kinematic, and acoustic analysis of supralaryngeal correlates of word-level prominence contrasts in English. In R. Mannell & J. Robert-Ribes (Eds.), *5th International Conference on Spoken Language Processing* (pp. 1851-1858), Australian Speech Science and Techology Association.

Hajicova, E. (1993). *Issues of Sentence Structure and Discourse Patterns*. Charles University.

Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F. & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *NeuroImage*, *30*,1383–1400. doi:10.1016/j.neuroimage.2005.11.048.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. University of Chicago Press.

Heim, S., & Alter, K. (2006). Prosodic pitch accents in language comprehension and production: ERP data and acoustic analyses. *Acta Neurobiologiae Experimentalis*, *66*(1), 55-68.

Heldner, M., & Strangert, E. (1997). To what extent is perceived focus determined by F0-cues? In G. Kokkinakis, N. Fakotakis & E. Dermatas (Eds.), *Proceedings of the 5th European Conference on Speech Communication and Technology* (pp. 875– 877), International Speech Communication Association.

Hermes, D. J., & Rump, H. H. (1994). Perception of prominence in speech intonation induced by rising and falling pitch movements. *The Journal of the Acoustical Society of America*, *96*(1), 83-92. doi:10.1121/1.410377.

Heselwood, B. (1998). An unusual kind of sonority and its implications for phonetic theory, *Working Papers in Linguistics and Phonetics*, *6*, 68-80.

von Heusinger, K. (1999). *Intonation and information structure* (PhD Thesis).

von Heusinger, K. & Schumacher, P. B. (2019). Discourse Prominence: Definition and Application. *Journal of Pragmatics*, *154*, 117-127. doi:10.1016/j.pragma.2019.07.025.

van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176-1190. doi:10.1080/17470218.2013.850521.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393-402. doi:10.1038/nrn2113.

Himmelmann, N. P., & Primus, B. (2015). Prominence beyond prosody: A first approximation. In A. De Dominicis (Ed.), *pS-prominenceS: Prominences in Linguistics. Proceedings of the International Conference* (pp. 38-58), Disucom Press.

Hirschberg, J., & Avesani, C. (1997). The role of prosody in disambiguating potentially ambiguous utterances in English and Italian. In A. Botinis, G. Kouroupetroglou & G. Karayiannis (Eds.) *Intonation: Theory, Models and Applications* (pp. 189-192). Dordrecht: Kluwer Academic Publishers.

Hirschberg, J., & Pierrehumbert, J. (1986). The intonational structuring of discourse. In A. W. Biermann (Ed.), *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics* (pp. 136-144). Association of Computational Linguistics.

Holcomb, P. J., & Neville, H. J. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology*, *19*(4), 286-300.

Honorof, D. N., & Whalen, D. H. (2005). Perception of pitch location within a speaker's F0 range. *The Journal of the Acoustical Society of America*, *117*(4), 2193-2200.

Howell, J. (2016). Revisiting English word stress and rhythm in the post-nuclear domain. *The Journal of the Acoustical Society of America*, *140*(4), 3399-3399. doi: 10.1121/1.4970899.

Hruska, C., & Alter, K. (2004). Prosody in Dialogues and Single Sentences: How Prosody can influence Speech Perception. In A. Steube (Ed.), *Information structure: Theoretical and empirical aspects* (pp. 221-226). De Gruyter. doi: 10.1515/9783110905892.

Hruska, C., Alter, K., Steinhauer, K., & Steube, A. (2001). Misleading dialogues: Human's brain reaction to prosodic information. In Ch. Cavé, I. Guaïtella & S. Santi (Eds.), *Proceedings of the Orality and Gesture. Interactions et comportements multimodaux dans la communication* (pp. 425-430). *l'Harmattan*.

Hsu, C. H., Evans, J. P., & Lee, C. Y. (2015). Brain responses to spoken F0 changes: Is H special? *Journal of Phonetics*, *51*, 82-92. doi:10.1016/j.wocn.2015.02.003.

Hualde, J. I., Cole, J., Smith, C., Eager, C., Mahrt, T., & Napoleão de Souza, R. (2016). The perception of phrasal prominence in English, Spanish and French conversational speech. In J. Barnes, A. Burgos, S. Shattuck-Hufnagel & N. Veilleux (Eds.), *Proceedings of the 8th International Conference of Speech Prosody* (pp. 459-463), International Speech Communication Association. doi:10.21437/speechprosody.2016-94.

Huang, T. (2004). *Language-specificity in auditory perception of Chinese tones* (PhD Thesis). Ohio State University Columbus.

Huss, V. (1978). English word stress in the post-nuclear position. *Phonetica*, *35*(2), 86-105. doi:10.1159/000259924.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of memory and language*, *58*(2), 541-573. doi: 10.1016/j.jml.2007.06.013.

Ito, K., Speer, S., & Beckman, M. (2004). Informational status and pitch accent distribution in spontaneous dialogues in English. In B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody* (pp. 279–282). Speech Prosody Special Interest Group.

Jagdfeld, N., & Baumann, S. (2011). Order Effects on the Perception of Relative Prominence. In W. Lee & E. Zee (Eds.), Proceedings of the*17th International Congress of Phonetic Sciences* (pp. 958-961). City University of Hong Kong.

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., Suchomel, V. (2013). The TenTen Corpus Family. In A. Hardie & R. Love (Eds.), *Proceedings of the 7th International*

Corpus Linguistics Conference (pp. 125–127). University Centre for Computer Corpus Research on Language.

de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, *97*(1), 491-504. doi: 10.1121/1.412275.

Jun, S. A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology*, *15*(2), 189-226.

Jun, S. (2005). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press. doi: 10.1093/acprof:oso/9780199249633.001.0001.

Jun, S. A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus*, *14*(1), 147-172. doi: 10.1515/prbs.2002.002.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. *Typological Studies in Language*, *45*, 229-254.

Kaland, C.C.L., Krahmer, E.J., & Swerts, M.G.J. (2014). White bear effect in language production: evidence from the prosodic realisation of adjectives. *Language and Speech*, *57*(4), 470-486.

Karns, C. M., Isbell, E., Giuliano, R. J., & Neville, H. J. (2015). Auditory attention in childhood and adolescence: An event-related potential study of spatial selective attention to one of two simultaneous stories. *Developmental Cognitive Neuroscience*, *13*, 53-67. doi: 10.1016/j.dcn.2015.03.001.

Keating, P., Cho, T., Fougeron, C., & Hsu, C. S. (2003). Domain-initial articulatory strengthening in four languages. *Phonetic Interpretation: Papers in Laboratory Phonology VI*, 143-161. doi: 10.1017/cbo9780511486425.009.

Kim, S., & Cho, T. (2013). Prosodic boundary information modulates phonetic categorization. *The Journal of the Acoustical Society of America*, *134*(1), EL19-EL25. doi: 10.1121/1.4807431.

Kishon-Rabin, L., Roth, D. A. E., Van Dijk, B., Yinon, T., & Amir, O. (2004). Frequency discrimination thresholds: the effect of increment versus decrement detection of frequency. *Journal of Cclinical Physiology and Pharmacology*, *15*(1-2), 29-40. doi: 10.1515/JBCPP.2004.15.1-2.29.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*(2), 153-194. doi: 10.1006/jmla.1998.2620.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148. doi: 10.1037/a0038695.

Kleinschmidt, D. F., Weatherholtz, K., & Jaeger, T. F. (2018). Sociolinguistic perception as inference under uncertainty. *Topics in Cognitive Science*, *10*(4), 818-834.

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*(3), 223-244. doi: 10.1177/0023830908098541.

Krahmer, E., & Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech Communication*, *34*(4), 391-405. doi: 10.1016/S0167-6393(00)00058-3.

Krahmer, E., & Swerts, M. (2006). Perceiving focus. In Lee, C.-M. (Ed.), *Topic and focus: a cross-linguistic perspective* (pp. 121-137). Kluwer Academic Publishers. doi: 10.1007/978-1-4020-4796-1_7.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language,* 57(3), 396-414. doi: 10.1016/j.jml.2007.06.005.

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology*, *50*(2), 194-231. doi: 10.1016/j.cogpsych.2004.08.002.

Krifka, M. (2007). Basic notions of information structure. In Féry, C. & Krifka, M. (Eds.), *Interdisciplinary Studies of Information Structure 6* (pp. 13-55). Universitätsverlag Potsdam. doi: 10.1556/ALing.55.2008.3-4.2.

Kristensen, L. B., Wang, L., Petersson, K. M., & Hagoort, P. (2012). The interface between language and attention: prosodic focus marking recruits a general attention network in spoken language comprehension. *Cerebral Cortex*, *23*(8), 1836-1848. doi: 10.1093/cercor/bhs164.

Kruijff-Korbayová, I., & Steedman, M. (2003). Discourse and information structure. *Journal of Logic, Language and Information*, *12*(3), 249-259.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, *118*(2), 1038-1054. doi: 10.1121/1.1923349.

Kohler, K. J. (1987). Categorical pitch perception. In *Proceedings of the 11th International Congress of Phonetic Sciences* (pp. 331-333). Academy of Sciences of the Estonian S.S.S.R.

Kohler, K. J. (2005). Form and function of non-pitch accents. *Prosodic Patterns of German Spontaneous Speech,* 97-123.

Kohler, K. J. (2006). What is emphasis and how is it coded. In R. Hoffmann & H. Mixdorff (Eds.), *Proceedings of the International Conference on Speech Prosody 2006* (pp. 748-751). Thelem Universitätsverlag und Buchhandel Dresden.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*(3), 557-577. doi: 10.1017/S0048577201990559.

Koster, J. (2000). Pied piping and the word orders of English and Dutch. In M. Hirotani, A. Coetzee, N. Hall & J. Y. Kim (Eds.), *Proceedings of the North East Linguistic Society* (Volume 30, pp. 415-426). Graduate Linguistics Student Association.

Kügler, F. (2003). Do we know the answer? Variation in yes-no-question intonation. *Linguistics in Potsdam*, *21*, 9-29.

Kügler, F., & Féry, C. (2017). Postfocal downstep in German. *Language and Speech*, *60*(2), 260-288. doi: 10.1177/0023830916647204.

Kügler, F., & Gollrad, A. (2015). Production and perception of contrast: The case of the fall-rise contour in German. *Frontiers in Psychology*, *6*, Article 1254. doi: 10.3389/fpsyg.2015.01254.

Kuno, S. (1972). Functional sentence perspective: A case study from Japanese and English. *Linguistic Inquiry* 3: 269-320.

Kutas, M., & Federmeier, K.D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621-647. doi: 10.1146/annurev.psych.093008.131123.

Kutas, M., & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*, *11*(2), 99-116. doi: 10.1016/0301-0511(80)90046-0.

Ladd, D. R. (1980). *The Structure of Intonational Meaning*. Indiana University Press.

Ladd, D. R. (1996, 2008). *Intonational phonology*. Cambridge University Press.

Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics*, *25*(3), 313-342.

Ladefoged, P. (1997). Linguistic phonetic descriptions. In W. J. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (pp. 589-618). Blackwell.

Lambrecht, K. (1994). *Information Structure and Sentence Form*. Cambridge University Press.

Lambrecht, K. (1996). Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents. Language, *75*(3), 129-152.

Leben, W. (1971). Suprasegmental and segmental representation of tone. *Studies in African Linguistics*, 183-200.

Lehiste, I. (1970). *Suprasegmentals*. MIT Press.

Lehiste, I., & Fox, R. A. (1992). Perception of prominence by Estonian and English listeners. *Language and Speech*, *35*(4), 419-434.

Leiner, D. J. (2014). SoSci Survey (Version 2.4.00-i) [Computer Software]. Available from https://www.soscisurvey.de.

Lepschy, A. L., & Lepschy, G. (1998). *The Italian language today*. New Amsterdam Books. doi: 10.4324/9781315003214.

Li, W., Deng, N., Yang, Y., & Wang, L. (2018). Process focus and accentuation at different positions in dialogues: An ERP study. *Language, Cognition and Neuroscience, 33*(2), 255–274. doi: 10.1080/23273798.2017.1387278.

Li, X. Q., & Ren, G. Q. (2012). How and when accentuation influences temporally selective attention and subsequent semantic processing during on-line spoken language comprehension: An ERP study. *Neuropsychologia*, *50*(8), 1882-1894. doi: 10.1016/j.neuropsychologia.2012.04.013.

Liberman, M. Y. (1975). *The intonational system of English* (PhD Thesis), Massachusetts Institute of Technology.

Liberman, M. Y., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In Aronoff, M. & Oerhle, R. (Eds.), *Language Sound Structure* (pp. 157–233), MIT Press.

Liberman, M. Y. & Prince, A. (1977). On Stress and Linguistic Rhythm. *Linguistic Inquiry, 8*, 249-336. doi: 10.2307/4177987.

Luchkina, T. V. (2016). *Prosodic and structural variability in free word order language discourse* (PhD Thesis). University of Illinois at Urbana-Champaign.

Luchkina, T., & Cole, J. (2014). Structural and prosodic correlates of prominence in free word order language discourse. In N. Campbell, D. Gibbon & D. Hirst (Eds.), *Proceedings of the 7th International Conference on Speech Prosody* (pp. 1119-1123). doi: 10.21437/speechprosody.2014-213.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. MIT Press. doi: 10.1118/1.4736938.

Luck, S. J., & Kappenman, E. S. (2012). ERP components and selective attention. In E. S. Kappenman & S. J. Luck (Eds.), *The Oxford Handbook of Event-Related Potential Components* (pp. 295-327), Oxford University Press. doi: 10.1093/oxfordhb/9780195374148.013.0144.

Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, *4*(11), 432-440. doi: 10.1016/S1364-6613(00)01545-X.

van Maastricht, L. J., Swerts, M. G. J., & Krahmer, E. J. (2013). Acquiring native-like intonation in Dutch and Spanish: Comparing the L1 and L2 of native speakers and second language learners (Poster Presentation, 21-23 August). In *Phonetics, Phonology, and Languages in contact 13*.

van Maastricht, L. J., Swerts, M. G. J., & Krahmer, E. J. (2014). *"Would you like RED or WHITE wine? -I would like red WINE": Native speaker perception of (non-) native intonation patterns* (Poster Presentation, 10-12 September). Tone and Intonation in Europe.

Maess, B., Schröger, E., & Widmann, A. (2016). High-pass filters and baseline correction in M/EEG analysis. Commentary on: "How inappropriate high-pass filters can produce artefacts and incorrect conclusions in ERP studies of language and cognition". *Journal of Neuroscience Methods*, *266*, 164-165. doi: 10.1016/j.jneumeth.2015.12.003.

Magne, C., Astésano, C., Lacheret-Dujour, A., Morel, M., Alter, K., & Besson, M. (2005). On-line processing of "pop-out" words in spoken French dialogues. *Journal of Cognitive Neuroscience*, *17*(5), 740-756. doi: 10.1162/0898929053747667.

Mahrt, T., Cole, J., Fleck, M., & Hasegawa-Johnson, M. (2012). F0 and the perception of prominence. In *Proceedings of the Thirteenth annual conference of the international speech communication association* (pp. 2422-2425). International Speech and Communication Association.

Marotta, G., & Sorianello, P. (1999). Question intonation in Sienese Italian. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1161-1164).

Marotta, G., & Sorianello, P. (2001). La teoria autosegmentale nell'analisi dell'intonazione interrogativa in due varietà di italiano toscano (Lucca e Siena). In F. A. Leoni, R. Sornicola, E. S. Krosbakken & C. Stromboli (Eds.), *Atti del XXXIII Congresso della Società di Linguistica Italiana* (Vol. 1, pp. 177-204). Bulzoni Editore.

Martin, P. (2012). The Autosegmental-Metrical Prosodic Structure: Not Fit for French? In *Proceedings of the 6th International Conference on Speech Prosody* (Vol. 1, pp. 131-134). Tongji University Press.

McCallum, W. C., Farmer, S. F., & Pocock, P. V. (1984). The effects of physical and semantic incongruites on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *59*(6), 477-488. doi: 10.1016/0168-5597(84)90006-6.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1-86. doi: 10.1016/0010-0285(86)90015-0.

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, *32*(4), 543-563. doi: 10.1016/j.wocn.2004.02.002.

Mo, Y. (2008). Duration and intensity as perceptual cues for naïve listeners' prominence and boundary perception. In P. Barbosa, S. Madureira & C. Reis (Eds.), *Proceedings of the 4th International Conference on Speech Prosody* (pp. 739-742). International Speech Communication Association.

Mo, Y. (2011). *Prosody production and perception with conversational speech* (PhD Thesis), University of Illinois at Urbana-Champaign.

Mo, Y., Cole, J., & Lee, E. (2008). Naive listeners' prominence and boundary perception. In P. Barbosa, S. Madureira & C. Reis (Eds.), *Proceedings of the 4th International Conference on Speech Prosody* (pp. 735-738). International Speech Communication Association.

Mücke, D., & Grice, M. (2014). The effect of focus marking on supralaryngeal articulation–Is it mediated by accentuation? *Journal of Phonetics*, *44*, 47-61. doi: 10.1016/j.wocn.2014.02.003.

Nakatani, C. H. (1997). The computational processing of intonational prominence: A functional prosody perspective (PhD Thesis). Harvard University.

Nava, E. A. (2010). *Connecting phrasal and rhythmic events: Evidence from second language speech*. University of Southern California.

Neppert, J. M. H. (1998). *Elemente einer Akustischen Phonetik*. Helmut Buske Verlag.

Nespor, I. & Vogel, M. (1986). *Prosodic phonology*. Foris.

Niebuhr, O. (2009). F_0-Based Rhythm Effects on the Perception of Local Syllable Prominence. *Phonetica*, *66*(1-2), 95-112. doi: 10.1159/000208933.

Niebuhr, O., Bergherr, J., Huth, S., Lill, C., & Neuschulz, J. (2010). Intonationsfragen hinterfragt-Die Vielschichtigkeit der prosodischen Unterschiede zwischen Aussage-und Fragesätzen mit deklarativer Syntax. *Zeitschrift für Dialektologie und Linguistik*, *77*(3), 304-346.

Niebuhr, O., D'Imperio, M., Fivela, B. G., & Cangemi, F. (2011). Are There" Shapers" and" Aligners"? Individual Differences in Signalling Pitch Accent Category. In W. S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Conference on Phonetic Sciences* (pp. 120-123). City University of Hong Kong.

Niebuhr, O., & Michaud, A. (2015). Speech data acquisition - The underestimated challenge. *Kieler Arbeiten in Linguistik und Phonetik*, (3), 1-42.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus--norepinephrine system. *Psychological Bulletin*, *131*(4), 510. doi: 10.1037/0033-2909.131.4.510.

Nolan, F. (2006). Intonation. In B. Aarts & A. McMahon (Eds.), *Handbook of English Linguistics* (pp. 433–457). Oxford: Blackwell. doi: 10.1002/9780470753002.ch19.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204-238. doi: 10.3758/APP.71.6.1207.

Ortega-Llebaria, M., Gu, H., & Fan, J. (2013). English speakers' perception of Spanish lexical stress: Context-driven L2 stress perception. *Journal of Phonetics*, *41*(3-4), 186-197. doi: 10.1016/j.wocn.2013.01.006.

Oxenham, A.J. (2012). Pitch perception. *The Journal of Neuroscience*, *32*(39), 13335-13338. doi: 10.1523/JNEUROSCI.3815-12.2012.

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, *33*(2), 107-120. doi: 10.1007/s10919-008-0065-7.

Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, *38*(3), 422–430. doi: 10.1016/j.wocn.2010.04.001.

Peters, J., Hanssen, J., & Gussenhoven, C. (2015). The timing of nuclear falls: Evidence from Dutch, West Frisian, Dutch Low Saxon, German Low Saxon, and High German. *Laboratory Phonology*, *6*(1), 1-52. doi: 10.1515/lp-2015-0004.

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, *35*, 73-89. doi: 10.1146/annurev-neuro-062111-150525.

van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 394-417. doi: 10.1037/0278-7393.25.2.394.

van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience*, *3*(2), 131-150. doi: 10.1162/jocn.1991.3.2.131.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation* (PhD Thesis). MIT Press.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, *2*, 33-52. doi: 10.1146/annurev-linguistics-030514-125050.

Pierrehumbert, J. B., & Beckman, M. E. (1988). Japanese tone structure. *Linguistic inquiry monographs*, (15), 1-282.

Pierrehumbert, J., & Hirschberg, J. (1990). The Meaning of Intonational Contours in the Interpretation of Discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.). *Intentions in Communication* (pp. 271-311). The MIT Press. doi: 10.7551/mitpress/3839.003.0016.

Pierrehumbert, J. B., & Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica*, *46*(4), 181-196. doi: 10.1159/000261842.

Pintér, G., Mizuguchi, S., & Tateishi, K. (2014). Perception of prosodic prominence and boundaries by L1 and L2 speakers of English. In H. Li, H. Meng, B. Ma, E. Chng & L. Xie (Eds.), *Proceedings of the 15th Annual Conference of the International Speech Communication Association* (pp. 544–547). International Speech and Communication Association.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, *118*(10), 2128-2148. doi: 10.1016/j.clinph.2007.04.019.

Prieto, P. (2011). Tonal alignment. In M. Van Oostendorp, C. Ewen, B. Hume & K. Rice (Eds.), *Companion to Phonology* (pp. 1185-1203). Wiley-Blackwell.

Prince, E. F. (1981). Towards a taxonomy of given-new information. In P. Cole (Ed.) *Radical Pragmatics* (pp. 223-254). New York: Academic Press.

Qin, Z., & Mok, P. (2012). The perception of speech and non-speech tones by tone and non-tone language listeners. *In Proceedings of the 6th International Conference on Speech Prosody* (Vol. 1, pp. 366-369). Tongji University Press.

R Core Team. (2019). R: A language and environment for statistical computing. *R Foundation for Statistical Computing* (version 3.6.2). Retrieved from https://www.r-project.org/.

Riesberg, S., Kalbertodt, J., Baumann, S., & Himmelmann, N. P. (2018). On the perception of prosodic prominences and boundaries in Papuan Malay. In S. Riesberg, A. Shiohara & A. Utzumi (Eds.) *Perspectives on information structure in Austronesian languages* (pp. 389-414). Language Science Press.

Riester, A. (2008). *The components of focus and their use in annotating information structure*. (PhD Thesis). University of Stuttgart.

Riester, A., & Piontek, J. (2015). Anarchy in the NP. When new nouns get deaccented and given nouns don't. *Lingua*, *165*, 230-253. doi: 10.1016/j.lingua.2015.03.006.

Rietveld, A., & Gussenhoven, C. (1985). On the relation between pitch excursion size and prominence. *Journal of Phonetics*, *13*(3), 299-308. doi: 10.1016/s0095-4470(19)30761-2.

Ritter, S., & Grice, M. (2015). The role of tonal onglides in German nuclear pitch accents. *Language and speech*, *58*(1), 114-128. doi: 10.1177/0023830914565688.

Roca, I., & Johnson, W. (1999). *A Course in Phonology*. Oxford, UK: Blackwell.

Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlesewsky, M. (2007). To predict or not to predict: Influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience*, *19*(8), 1259-1274. doi: 10.1162/jocn.2007.19.8.1259.

Röhr, C. T. (2016). *The Information Status of Nominal and Verbal Expressions: Intonational Evidence from Production and Perception in German* (PhD Thesis). Universität zu Köln.

Röhr, C. T., & Baumann, S. (2010). Prosodic marking of information status in German. In M. Hasegawa-Johnson (Ed.), *Proceedings of the 5th International Conference of Speech Prosody* (paper 19). International Speech Communication Association.

Röhr, C. T. & Baumann, S. (2011). Decoding Information Status by Type and Position of Accent in German. In W. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (Volume 3, pp. 1706-1709). City University of Hong Kong.

Röhr, C. T., Brilmayer, I., Baumann, S., Grice, M. & Schumacher, P. B. (2020). Signal-Driven and Expectation-Driven Processing of Accent Types. *Language, Cognition & Neuroscience*, *36*(1), 33-59. doi: 10.1080/23273798.2020.1779324.

Röhr, C. T., Thies, T., Baumann, S. & Grice, M. (2016). Prosodic Marking of Information Status in Task-Oriented Dialogues. In C. Draxler & F. Kleber (Eds.), *Tagungsband der 12. Tagung Phonetik und Phonologie im deutschsprachigen Raum* (pp. 168-171).

Roessig, S., & Mücke, D. (2019). Modelling dimensions of prosodic prominence. *Frontiers in Communication*, *4*, 44. doi: 10.3389/fcomm.2019.00044.

Roessig, S., Mücke, D. & Grice, M. (2019). The dynamics of intonation: Categorical and continuous variation in an attractor-based model. *PLoS One*, 14(5). doi: 10.31234/osf.io/qkm6b.

Roettger, T. B. (2017). *Tonal placement in Tashlhiyt: How an intonation system accommodates to adverse phonological environments* (Vol. 3). Language Science Press. doi: 10.31730/osf.io/46q5j.

Roettger, T. B., & Grice, M. (2019). The tune drives the text: Competing information channels of speech shape phonological systems. *Language Dynamics and Change*, *9*(2), 265-298. doi: 10.31234/osf.io/5bm8r.

Roettger, T. B., Mahrt, T., & Cole, J. (2019). Mapping prosody onto meaning–the case of information structure in American English. *Language, Cognition and Neuroscience*, 1-20. doi: 10.31234/osf.io/tevd6.

Roettger, T. B., & Stoeber, M. (2017). Manual Response Dynamics Reflect Rapid Integration of Intonational Information during Reference Resolution. *Proceedings of 39th Annual Conference of the Cognitive Science Society*. Austin, Texas: Cognitive Science Society.

Romano, A. & Miotti R. (2008). Distancias prosódicas entre variedades románicas. In Turculet, A. (Ed.), *La variation diatopique de l'intonation dans le domaine roumain et roman* (pp. 231-249). Editura Universitatii.

Rooth, M. (1985). *Association with Focus* (PhD Thesis). University of Massachusetts.

Rooth, M. (1992). A theory of focus interpretation. *Natural language semantics*, *1*(1), 75-116. doi: 10.1007/bf02342617.

Roseano, P., Vanrell, M. D. M., & Prieto, P. (2015). Intonational phonology of Friulian and its dialects. In Frota, S. & Prieto, P. (Eds.) *Intonational variation in Romance* (pp. 101-139). Oxford University Press. doi: 10.1093/acprof:oso/9780199685332.003.0004.

Rosenberg, A. Cooper, E., Levitan, R. & Hirschberg, J. (2012). Cross-language prominence detection. *Speech Prosody*, 278-281. doi: 10.1109/icassp.2013.6639316.

Roy, J., Cole, J., & Mahrt, T. (2017). Individual differences and patterns of convergence in prosody perception. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1). doi: 10.5334/labphon.108.

Roye, A., Jacobsen, T., & Schröger, E. (2007). Personal significance is encoded automatically by the human brain: an event-related potential study with ringtones. *European Journal of Neuroscience*, *26*(3), 784-790. doi: 10.1111/j.1460-9568.2007.05685.x.

Ruchkin, D. S., Johnson Jr, R., Canoune, H. L., Ritter, W., & Hammer, M. (1990). Multiple sources of P3b associated with different types of information. *Psychophysiology*, *27*(2), 157-176. doi: 10.1111/j.1469-8986.1990.tb00367.x.

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Memory & cognition*, *18*(4), 367-379. doi: 10.3758/bf03197126.

Sanford, A. J. (2002). Context, Attention and Depth of Processing During Interpretation. Mind & Language, 17(1&2), 188-206. doi: 10.1111/1468-0017.00195.

Sanford, A. J., & Graesser, A. C. (2006). Shallow processing and underspecification. *Discourse Processes*, *42*(2), 99-108. doi: 10.1207/s15326950dp4202_1.

Sanford, A. J., Sanford, A. J., Molle, J., & Emmott, C. (2006). Shallow processing and attention capture in written and spoken discourse. *Discourse Processes*, *42*(2), 109-130. doi: 10.1207/s15326950dp4202_2.

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and language*, *137*, 29-39. doi: 10.1016/j.bandl.2014.07.010.

Savino, M. (2012). The intonation of polar questions in Italian: Where is the rise? *Journal of the International Phonetic Association,* *42*(1), 23-48. doi:10.1017/S002510031100048X.

Savino, M. (2013). Exploring Regional Variation in Italian Question Intonation: A Corpus-Based Study. In Bamford, J., Cavalieri, S. & Diani, G. (Eds.), *Variation and change in spoken and written discourse: perspectives from corpus linguistics* (pp. 79-108). John Benjamins. doi: 10.1075/ds.21.11sav.

Savino, M., & Grice, M. (2011). The perception of negative bias in Bari Italian questions. In *Prosodic categories: Production, perception and comprehension*. Springer, Dordrecht, pp. 187-206. doi: 10.1007/978-94-007-0137-3_8.

Schröder, M. (2001). Emotional speech synthesis: A review. In *Proceedings of the 7th European conference on speech communication and technology* (Volume 1, pp. 561–564). Aalborg University.

Schütze, C., & Sprouse, J. (2014). Judgment data. In R. Podesva & D. Sharma (Eds.), *Research Methods in Linguistics* (pp. 27-50). Cambridge University Press. doi: 10.1017/cbo9781139013734.004.

Schumacher, P. B., Backhaus, J., & Dangl, M. (2015). Backward-and forward-looking potential of anaphors. *Frontiers in Psychology*, *6*, 1746. doi: 10.3389/fpsyg.2015.01746.

Schumacher, P. B., & Baumann, S. (2010). Pitch accent type affects the N400 during referential processing. *NeuroReport*, *21*(9), 618-622. doi: 10.1097/wnr.0b013e328339874a.

Schumacher, P. B., & Hung, Y. C. (2012). Positional influences on information packaging: Insights from topological fields in German. *Journal of Memory and Language*, *67*(2), 295-310. doi: 10.1016/j.jml.2012.05.006.

Schumacher, P. B., & Weiland, H. (2011). Reading Brecht and talking to the espresso: Electrophysiological investigations of conventional and novel metonymy. In *Proceedings of the Metonymy 2011 workshop*, pp. 23-28.

Schweitzer, A. (2019). Exemplar-theoretic integration of phonetics and phonology: Detecting prominence categories in phonetic space. *Journal of Phonetics*, *77*, 100915. doi: 10.1016/j.wocn.2019.100915.

Schweitzer, K., Riester, A., Walsh, M. & Dogil, G. (2009). Pitch accents and information status in a German radio news corpus. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association* (pp. 2415-2418). International Speech Communication Association.

Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. *The handbook of phonological theory*, *1*, 550-569.

Selting, M. (1995). *Prosodie im Gespräch*. Niemeyer. doi: 10.1515/9783110934717.

Shattuck-Hufnagel, S., Ostendorf, M. & Ross, K. (1994). Stress Shift and Early Pitch Accent Placement in Lexical Items in American English. *Journal of Phonetics 22*, 357-388. doi: 10.1016/s0095-4470(19)30291-8.

Shport, I. A. (2015). Perception of acoustic cues to Tokyo Japanese pitch-accent contrasts in native Japanese and naive English listeners. *The Journal of the Acoustical Society of America*, *138*(1), 307-318. doi: 10.1121/1.4922468.

Shuai, L., & Gong, T. (2014). Temporal relation between top-down and bottom-up processing in lexical tone perception. *Frontiers in behavioral neuroscience*, *8*, 97. doi: 10.3389/fnbeh.2014.00097.

Silverman, K. E., & Pierrehumbert, J. B. (1990). The timing of prenuclear high accents in English. In Kingston, J. & Beckman, M. (Eds.) *Papers in Laboratory Phonology I: Between the Grammar on Physics of Speech*, 72-106. doi: 10.1017/cbo9780511627736.005.

Singmann, H., Bolker, B., Westfall, J. & Aust, F. (2018). afex: Analysis of Factorial Experiments. R package version 0.21-2. https://CRAN.R-project.org/package=afex

Sluijter, A. M., & Van Heuven, V. J. (1995). Effects of focus distribution, pitch accent and lexical stress on the temporal organization of syllables in Dutch. Phonetica, 52 (2):71–89. doi: 10.1159/000262061.

Sluijter, A. M., & Van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical society of America*, *100*(4), 2471-2485.

Smith, C. (2011). Perception of prominence and boundaries by naïve French listeners. In W. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1874–1877). City University of Hong Kong.

Smith, N. J., & Kutas, M. (2015). Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology*, *52*(2), 157-168. doi: 10.1111/psyp.12317.

Smith, R., & Rathcke, T. (2020). Dialectal phonology constrains the phonetics of prominence. *Journal of Phonetics*, *78*, 100934. doi: 10.1016/j.wocn.2019.100934.

Spivey, M.J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, 102, 10393–10398. doi: 10.1073pnas.0503903102.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and clinical neurophysiology*, *38*(4), 387-401. doi: 10.1016/0013-4694(75)90263-1.

Streefkerk, B. M. (2002). *Prominence. Acoustic and lexical/syntactic correlates.* (PhD Thesis). LOT.

Swerts, M., Krahmer, E., & Avesani, C. (2002). Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics*, *30*(4), 629-654. doi: 10.1006/jpho.2002.0178.

Swerts, M., & Vroomen, J. (2015). Accent shifts in spoken noun phrases affect verification latencies of listeners in Dutch but not Canadian French. *Journal of Phonetics*, *52*, 170-182. doi: 10.1016/j.wocn.2015.07.005.

Swerts, M., & Zerbian, S. (2010). Intonational differences between L1 and L2 English in South Africa. *Phonetica*, *67*(3), 127-146. doi: 10.1159/000321052.

Szendröi, K. (2002). Stress–focus correspondence in Italian. In Beyssade, C., Bok-Bennema, R., Drijkoningen, F. & Monachesi, P. (Eds.), *Romance Languages and Linguistic Theory 2000: selected papers from "Going Romance" 2000* (pp. 287-303), John Benjamins Amsterdam. doi: 10.1075/cilt.232.16sze.

Terken, J. (1991). Fundamental frequency and perceived prominence of accented syllables. *The Journal of the Acoustical Society of America*, *89*(4), 1768-1776. doi: 10.1121/1.401019.

Terken J. (1997). Variation of accent prominence within the phrase: models and spontaneous speech data. In Sagisaka, Y., Campbell, N., & Higuchi, N. (Eds.), *Computing Prosody*, pp. 95-115. doi: 10.1007/978-1-4612-2258-3_8.

Terken, J., & Hirschberg, J. (1994). Deaccentuation of words representing "given" information: Effects of persistence of grammatical function and surface position. *Language and Speech*, 37 (2), 125–145. doi: 10.1177/002383099403700202.

Terken, J., & Hermes, D. (2000). The perception of prosodic prominence. In Horne, M. (Ed.), *Prosody: Theory and experiment* (pp. 89-127). Springer. doi: 10.1007/978-94-015-9413-4_5.

Terken, J., & Nooteboom, S. G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and cognitive processes*, *2*(3-4), 145-163. doi: 10.1080/01690968708406928.

Thies, T., Röhr, C. T., Baumann, S., & Grice, M. (2018). Prosodic Marking of Information Status in Picture Story Descriptions. *Proceedings of the Conference on Phonetics & Phonology in German-speaking countries (P&P 13),* 13, 193-196.

Toepel, U., & Alter, K. (2004). On the independence of information structural processing from prosody. *Information structure: Theoretical and empirical aspects*, 227-240. doi: 10.1515/9783110905892.227.

Toepel, U., Pannekamp, A., & Alter, K. (2007). Catching the news: Processing strategies in listening to dialogs as measured by ERPs. *Behavioral and Brain Functions*, *3*(1), 53. doi: 10.1186/1744-9081-3-53.

Turk, A. E., & Sawusch, J. R. (1996). The processing of duration and intensity cues to prominence. *The Journal of the Acoustical Society of America*, *99*(6), 3782-3790. doi: 10.1121/1.414995.

Turnbull, R. (2017). The role of predictability in intonational variability. *Language and speech*, *60*(1), 123-153. doi: 10.1177/0023830916647079.

Uhmann, S. (1991). On the tonal disambiguation of focus structures. *Journal of Semantics* 8, (3), 219-238. doi: 10.1093/jos/8.3.219.

Vallduví E. (1991). The role of plasticity in the association of focus and prominence, *Proceedings of the eastern states conference on linguistics* (ESCOL) 7, pp. 295-306.

Vallduví, E. (1992). *The Informational Component*. Garland.

Vanderslice, R., & Ladefoged, P. (1972). Binary suprasegmental features and transformational word-accentuation rules. *Language*, 819-838. doi: 10.2307/411990.

Ventura, C., Grice, M., Savino, M., Kolev, D., Brilmayer, I. & Schumacher P. B. (2020). Attention allocation in a language with post-focal prominences. *NeuroReport* 31(8), 624–62. doi: 10.1097/wnr.0000000000001453.

Wagner, P. (2005). Great expectations-introspective vs. perceptual prominence ratings and their acoustic correlates. In *Proceedings of the 9th European Conference on Speech Communication and Technology* (pp. 2381-2384). International Speech Communication Association.

Wagner, M., Breen, M., Flemming, E., Shattuck-Hufnagel, S., & Gibson, E. (2010). Prosodic effects of discourse salience and association with focus. In M. Hasegawa-Johnson (Ed.), *Proceedings of the 5th International Conference on Speech Prosody* (Paper 239). International Speech Communication Association.

Wagner, P., Ćwiek, A., & Samlowski, B. (2019). Exploiting the speech-gesture link to capture fine-grained prosodic prominence impressions and listening strategies. *Journal of Phonetics*, *76*, 100911. doi: 10.1016/j.wocn.2019.07.001.

Wagner, P., Origlia, A., Avesani, C., Christodoulides, G., Cutugno, F., D'Imperio, M., Mancebo, D. E., Gili-Fivela, B., Lacheret, A., Ludusan, B., Moniz, H., Chasaide, A., N., Niebuhr, O., Rousier-Vercruyssen, L., Simon, A., Šimko, J., Tesser, F. & Vainio, M. (2015). Different Parts of the same Elephant: a Roadmap to Disentangle and Connect Different Perspectives of Prosodic Prominence. In Procedings of the 18th *International Congress of Phonetic Sciences*. International Phonetic Association.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and cognitive processes,* 25(7-9), 905–945. doi: 10.1080/01690961003589492.

Wang, L., Bastiaansen, M., Yang, Y., & Hagoort, P. (2011). The influence of information structure on the depth of semantic processing: How focus and pitch accent determine the size of the N400 effect. *Neuropsychologia*, *49*(5), 813-820. doi: 10.1016/j.neuropsychologia.2010.12.035.

Wang, L., Hagoort, P., & Yang, Y. (2009). Semantic illusion depends on information structure: ERP evidence. *Brain research*, *1282*, 50-56. doi: 10.1016/j.brainres.2009.05.069.

Wang, Q. (2008). L2 stress perception: The reliance on different acoustic cues. In *Proceedings of Speech Prosody* (pp. 135-138).

Ward, P., & Sturt, P. (2007). Linguistic focus and memory: An eye movement study, *Memory & cognition*, *35*(1), 73-86. doi: 10.3758/bf03195944.

Watson, D. G. (2010). The many roads to prominence: Understanding emphasis in conversation. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (Volume 52, pp. 163-183). Elsevier. doi: 10.1016/S0079-7421(10)52004-8.

Watson, D. G., Arnold, J. E., & Tanenhouse, M. K. (2008). Tic Tac TOE: Effective of predictability and importance on acoustic prominence in language production. Cognition 106(3), 1548-1557. doi: 10.1016/j.cognition.2007.06.009.

Watson, D. G., Tanenhaus, M. K., & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. L+ H. *Cognitive science*, *32*(7), 1232-1244. doi: 10.1080/03640210802138755.

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*, *49*(3), 367-392. doi: 10.1177/00238309060490030301.

Weiland, H., Bambini, V., & Schumacher, P. B. (2014). The role of literal meaning in figurative language comprehension: Evidence from masked priming ERP. *Frontiers in Human Neuroscience*, *8*, 583. doi: 10.3389/fnhum.2014.00583.

Welby, P. (2003). Effects of pitch accent position, type, and status on focus projection. *Language and Speech*, *46*(1), 53-81. doi: 10.1177/00238309030460010401.

Weinreich, U. (1954). Stress and Word Structure in Yiddish. T*he Field of Yiddish: Studies in Yiddish Language, Folklore and Literature*, 1-27.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, *91*(3), 1707-1717. doi: 10.1121/1.402450.

Wochner, D., Schlegel, J., Dehé, N., & Braun, B. (2015). The prosodic marking of rhetorical questions in German. *Interspeech 2015: 16th Annual Conference of the International Speech Communication Association*, pp. 987-991.

Xu, Y. (2011). Post-focus Compression: Cross-linguistic Distribution and Historical Origin. *ICPhS*, 152-155.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*(2), 159-197. doi: 10.1016/j.wocn.2004.11.001.

Yamane, N., Yoshimura, N., & Fujimori, A. (2016). Prosodic transfer from Japanese to English: Pitch in focus marking. *Phonological Studies*, *19*, 97-105.

Yang, C. L., Perfetti, C. A., & Schmalhofer, F. (2007). Event-related potential indicators of text integration across sentence boundaries. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 55. doi: 10.1037/0278-7393.33.1.55.

Yang, X., Chen, X., Chen, S., Xu, X., & Yang, Y. (2013). Topic Structure Affects Semantic Integration: Evidence from Event-Related Potentials. *PLoS ONE 8(12)*: e79734. doi: 10.1371/journal.pone.0079734.

You, H. J. (2012). Determining prominence and prosodic boundaries in Korean by non-expert rapid prosody transcription. In *Proceedings of the 6th International Conference on Speech Prosody,* 318-321.

Zappoli, A., Vespignani, F., Baumann, S., Grice, M., & Schumacher, P. B. (2018). The processing of German Pitch Accents by Italian learners of German (Oral presentation, March). *The relation between prosodic and referential structure DGfS Conference*.

Zimmermann, M., & Féry, C. (2010). Introduction. In Zimmermann, M., & Féry, C. (Eds.), *Information structure: Theoretical, typological, and experimental perspectives* (pp. 1-11). Oxford University Press.

Zubizarreta M., L. (1998). *Prosody, focus, and word order*. MIT Press.

Appendix

A.1 Test Material Long dataset

Complete reading material (question-answers pairs) for the Long dataset. The critical words in the answers are underlined (the target words are printed in bold face). Note that the answers were presented to participants without any word being underlined or bolded. Items and conditions are here presented in order, but were randomised for the experiment.

Item	Condition	Questions	Answers
1	BF	Che cosa bisogna fare per amministrare la città?	Per amministrare la città, bisogna votare il **partito** alle elezioni
1	NF	Che cosa bisogna votare alle elezioni per amministrare la città? Il partito o il candidato?	
1	PF	Per amministrare la città bisogna votare o boicottare il partito alle elezioni?	
2	BF	Che cosa bisogna fare quando si frigge?	Quando si frigge, bisogna proteggere i **vestiti** dagli schizzi
2	NF	Che cosa bisogna proteggere dagli schizzi quando si frigge? I vestiti o i capelli?	
2	PF	Quando si frigge bisogna proteggere o smacchiare i vestiti dagli schizzi?	
3	BF	Che cosa bisogna fare per suonare nell'orchestra?	Per suonare nell'orchestra, bisogna imparare lo **spartito** perfettamente
3	NF	Che cosa bisogna imparare perfettamente per suonare nell'orchestra? Lo spartito o il ritornello?	
3	PF	Per suonare nell'orchestra bisogna imparare o copiare lo spartito perfettamente?	
4	BF	Che cosa bisogna fare durante l'inverno?	Durante l'inverno, bisogna scaldare il **locale** per i clienti
4	NF	Che cosa bisogna scaldare per i clienti durante l'inverno? Il locale o il terrazzo?	
4	PF	Durante l'inverno bisogna scaldare o arieggiare il locale per i clienti?	
5	BF	Che cosa bisogna fare per avere una bocca sana?	Per avere una bocca sana, bisogna disinfettare le **gengive** con il collutorio
5	NF	Che cosa bisogna disinfettare con il collutorio per avere una bocca sana? Le gengive o i denti?	
5	PF	Per avere una bocca sana bisogna disinfettare o spazzolare le gengive con il collutorio?	

6	BF	Che cosa bisogna fare per sedersi comodamente?	Per sedersi comodamente, bisogna mettere i **cuscini** sugli sgabelli
6	NF	Che cosa bisogna mettere sugli sgabelli per sedersi comodamente? I cuscini o la coperta?	
6	PF	Per sedersi comodamente bisogna mettere o togliere i cuscini sugli sgabelli?	
7	BF	Che cosa bisogna fare per arredare la casa?	Quando si arreda la casa, bisogna scegliere il **divano** per il soggiorno
7	NF	Che cosa bisogna scegliere per il soggiorno quando si arreda la casa? Il divano o il letto?	
7	PF	Quando si arreda la casa bisogna scegliere o noleggiare il divano per il soggiorno?	
8	BF	Cosa bisogna fare per andare a caccia?	Per andare a caccia, bisogna caricare il **fucile** con i proiettili
8	NF	Cosa bisogna caricare con i proiettili per andare a caccia? Il fucile o la pistola?	
8	PF	Per andare a caccia bisogna caricare o sigillare il fucile con i proiettili?	
9	BF	Cosa si fa per allenare i polpacci?	Per allenare i polpacci, è preferibile pedalare in **salita** con la bicicletta
9	NF	Dove bisogna pedalare con la bicicletta per allenare I polpacci? In salita o in discesa?	
9	PF	Per allenare i polpacci è meglio pedalare o camminare in salita con la bicicletta?	
10	BF	Che cosa bisogna fare per non perdersi?	Per non perdersi, bisogna consultare la **cartina** nel dettaglio
10	NF	Che cosa bisogna consultare nel dettaglio per non perdersi? La cartina o il volantino?	
10	PF	Per non perdersi bisogna consultare o memorizzare la cartina nel dettaglio?	
11	BF	Che cosa si fa quando si va in gita?	Quando si va in gita, bisogna preparare un **panino** per il pranzo
11	NF	Che cosa si prepara per il pranzo quando si va in gita? Un panino o una minestra?	
11	PF	Quando si va in gita bisogna preparare o comprare un panino per il pranzo?	
12	BF	Cosa bisogna fare durante una camminata?	Durante una camminata, bisogna evitare le **ortiche** lungo il sentiero
12	NF	Che cosa bisogna evitare lungo il sentiero durante una camminata? Le ortiche o le margherite?	
12	PF	Durante una camminata bisogna evitare o raccogliere le ortiche lungo il sentiero?	

13	BF	Che cosa bisogna fare per avere un mobile lucido?	Per avere un mobile lucido, bisogna <u>stendere</u> la **<u>vernice</u>** sulla <u>superficie</u>
13	NF	Che cosa bisogna stendere sulla superficie per avere un mobile lucido? La vernice o la tempera?	
13	PF	Per avere un mobile lucido bisogna stendere o rovesciare la vernice sulla superficie?	
14	BF	Che cosa bisogna fare per una crescita rigogliosa?	Per una crescita rigogliosa, bisogna <u>fornire</u> il **<u>concime</u>** alle <u>piante</u>
14	NF	Che cosa bisogna fornire alle piante per una crescita rigogliosa? Il concime o il sale?	
14	PF	Per una crescita rigogliosa bisogna fornire o togliere il concime alle piante?	
15	BF	Che cosa bisogna fare per preparare un buon tè?	Per preparare un buon tè, bisogna <u>utilizzare</u> la **<u>bustina</u>** una sola <u>volta</u>
15	NF	Che cosa bisogna utilizzare una volta sola per preparare un buon tè? La bustina o il cucchiaino?	
15	PF	Per preparare un buon tè bisogna utilizzare o immergere la bustina una sola volta?	
16	BF	Che cosa bisogna fare per usare il fornello a gas?	Quando si usa il fornello a gas, bisogna <u>pagare</u> il **<u>metano</u>** ogni <u>mese</u>
16	NF	Che cosa si deve pagare ogni mese per utilizzare il fornello a gas? Il metano o etano?	
16	PF	Per usare il fornello a gas bisogna pagare o riallacciare il metano ogni mese?	
17	BF	Che cosa bisogna fare per eseguire la sinfonia?	Per eseguire la sinfonia, bisogna <u>includere</u> il **<u>violino</u>** tra gli <u>strumenti</u>
17	NF	Che cosa bisogna includere tra gli strumenti per eseguire la sinfonia? Il violino o il mandolino?	
17	PF	Per eseguire la sinfonia bisogna includere o escludere il violino tra gli strumenti?	
18	BF	Che cosa bisogna fare per irrigare?	Per irrigare, bisogna <u>riempire</u> i **<u>canali</u>** lungo i <u>campi</u>
18	NF	Che cosa bisogna riempire lungo i campi per irrigare? I canali o le buche?	
18	PF	Per irrigare bisogna riempire o svuotare i canali lungo i campi?	
19	BF	Come si fa il pieno alla macchina?	Per fare il pieno, bisogna <u>riempire</u> di **<u>benzina</u>** il <u>serbatoio</u>
19	NF	Di che cosa bisogna riempire il serbatoio? Di benzina o di naftalina?	

19	PF	Per fare il pieno bisogna riempire o ripulire dalla benzina il serbatoio?	
20	BF	Che cosa bisogna fare per allestire la mostra?	Quando si allestisce la mostra, bisogna appendere i **dipinti** alle pareti
20	NF	Che cosa bisogna appendere alle pareti per allestire la mostra? I dipinti o i poster?	
20	PF	Per allestire la mostra bisogna appendere o appoggiare i dipinti alle pareti?	

Table A1. Sentences (questions and answers) used for the reading task. Questions elicited different information structures for the answers, thus for each item (firs column on the left) target words (displayed in bold) occurred in BF (broad focus), NF (narrow contrastive focus) and PF (post-focal position).

A.2 Test Material Short dataset (Udine variety) and Bari dataset (whole dataset).

Complete reading material (question-answers pairs) for the Short dataset and for the dataset collected for the Bari variety. The critical words in the answers are underlined (the target words are printed in bold face). Note that the answers were presented to participants without any word being neither underlined nor bolded. Items and conditions were randomised for the experiment.

Item	Condition	Questions	Answers
1	BF	Cosa bisogna fare per arieggiare?	Bisogna aprire la **finestra** nella stanza
1	NF	Per arieggiare bisogna aprire la finestra o la porta nella stanza?	
1	PF	Per arieggiare bisogna aprire o chiudere la finestra nella stanza?	
2	BF	Cosa bisogna fare quando piove?	Bisogna prendere gli **stivali** per il fango
2	NF	Per il fango bisogna prendere gli stivali o i sandali?	
2	PF	Quando piove bisogna prendere o buttare gli stivali per il fango?	
3	BF	Cosa bisogna fare per entrare nella stanza?	Bisogna girare la **maniglia** sulla porta
3	NF	Per entrare nella stanza bisogna girare la maniglia o ill chiavistello sulla porta?	
3	PF	Per entrare nella stanza bisogna girare o spingere la maniglia sulla porta?	
4	BF	Cosa bisogna fare per non avere freddo?	Bisogna portare il **maglione** nella borsa
4	NF	Quando fa freddo bisogna portare il maglione	

		o il cappello nella borsa?	
4	PF	Per non avere freddo bisogna portare o lasciare il maglione nella borsa?	
5	BF	Cosa bisogna fare per preparare la marmellata?	Bisogna cuocere le **ciliegie** nella pentola
5	NF	Per preparare la marmellata bisogna cuocere le ciliegie o i semi nella pentola?	
5	PF	Per preparare la marmellata bisogna cuocere o schiacciare le ciliegie nella pentola?	
6	BF	Cosa bisogna fare per preparare il tè?	Bisogna tagliare il **limone** a fettine
6	NF	Per preparare il tè bisogna tagliare il limone o il mandarino a fettine?	
6	PF	Per preparare il tè bisogna tagliare o dividere il limone a fettine?	
7	BF	Cosa bisogna fare per produrre il vino?	Bisogna avere la **cantina** con le botti
7	NF	Per produrre il vino bisogna avere la cantina o la cucina con le botti?	
7	PF	Per produrre il vino bisogna avere o noleggiare la cantina con le botti?	
8	BF	Cosa bisogna fare per imparare a suonare?	Bisogna usare la **pianola** con frequenza
8	NF	Pr imparare a suonare bisogna usare la pianola o il computer con frequenza?	
8	PF	Per imparare a suonare bisogna usare o osservare la pianola con frequenza?	
9	BF	Cosa bisogna fare per allenare i polpacci?	Bisogna correre in **salita** ogni giorno
9	NF	Per allenare i polpacci bisogna correre in salita o in discesa ogni giorno?	
9	PF	Per allenare i polpacci bisogna correre o camminare in salita ogni giorno?	
10	BF	Cosa bisogna fare per ridurre l'umidità?	Bisogna risanare le **paludi** in pianura
10	NF	Per ridurre l'umidità bisogna risanare le paludi o i fiumi in pianura?	
10	PF	Per ridurre l'umidità bisogna risanare o irrigare le paludi in pianura?	
11	BF	Cosa bisogna fare quando si va in gita?	Bisogna preparare un **panino** per la merenda
11	NF	Quando si va in gita bisogna preparare un panino o un caffè per la merenda?	

11	PF	Quando si va in gita bisogna preparare o comprare un panino per la merenda?	
12	BF	Cosa bisogna fare per preparare il minestrone?	
12	NF	Per fare il minestrone bisogna comprare la verdura o il pane al mercato?	Bisogna comprare la **verdura** al mercato
12	PF	Per fare il minestrone bisogna comparare o valutare la verdura al mercato?	
13	BF	Cosa bisogna fare per preparare la zuppa?	
13	NF	Bisogna tagliare le carote o i fagioli a dadini?	Bisogna tagliare le **carote** a dadini
13	PF	Per fare la zuppa bisogna tagliare o comprare le carote a dadini?	
14	BF	Cosa bisogna fare quando ci si taglia?	
14	NF	Bisogna coprire la ferita o la garza con la pomata?	Bisogna coprire la **ferita** con la pomata
14	PF	Biaogna coprire o grattare la ferita con la pomata?	
15	BF	Cosa bisogna fare a Natale?	
15	NF	Bisogna comprare un regalo o una pianta per gli amici?	Bisogna comprare un **regalo** per gli amici
15	PF	Bisogna comprare o riciclare un regalo per gli amici?	
16	BF	Cosa bisogna fare per andare in vacanza?	
16	NF	Per andare in vacanza cosa bisogna riempire la valigia o la macchina con i vestiti?	Bisogna riempire la **valigia** con i vestiti
16	PF	Bisogna riempire o svuotare la valigia con i vestiti?	
17	BF	Cosa bisogna fare per entrare nell'orchestra?	
17	NF	Per entrare nell'orchestra bisogna suonare il violino o il mandolino da tanti anni?	Bisogna suonare il **violino** da tanti anni
17	PF	Per entrare nell’orchestra bisogna suonare o possedere il violino da tanti anni?	
18	BF	Cosa bisogna fare per irrigare?	
18	NF	Per irrigare bisogna riempire il canale o le buche tra i campi?	Bisogna riempire il **canale** tra i campi
18	PF	Per irrigare bisogna riempire o svuotare il canale tra i campi?	

19	BF	Cosa bisogna fare per preparare la torta?	Bisogna pesare la **farina** con la bilancia
19	NF	Bisogna pesare la farina o la vanillina con la bilancia?	
19	PF	Per fare una torta bisogna pesare o filtrare la farina con la bilancia?	
20	BF	Cosa bisogna fare dopo il pranzo?	Bisogna sbattere la **tovaglia** dalle briciole
20	NF	Bisogna sbattere la tovaglia o i tovaglioli dalle briciole?	
20	PF	Bisogna sbattere o lavare la tovaglia dalle briciole?	

Table A2. Sentences (questions and answers) used for the reading task. Questions elicited different information structures for the answers, thus for each item (firs column on the left) target words (displayed in bold) occurred in BF (broad focus), NF (narrow contrastive focus) and PF (post-focal position).

A.3 Test Material Rating study

List of items of the trial utterances presented to participants. All critical words are underlined, target words are bolded. Utterances corresponding to each item were presented in three focal conditions: target word in broad focus, target word in narrow contrastive focus and target word in post-focal position.

Item	Trial utterance
1	Bisogna sbattere la **tovaglia** dalle briciole
2	Bisogna prendere gli **stivali** per il fango
3	Bisogna girare la **maniglia** sulla porta
4	Bisogna portare il **maglione** nella borsa
5	Bisogna cuocere le **ciliegie** nella pentola
6	Bisogna avere la **cantina** con le botti
7	Bisogna usare la **pianola** con frequenza
8	Bisogna coprire la **ferita** con la pomata
9	Bisogna comprare un **regalo** per gli amici
10	Bisogna pesare la **farina** con la bilancia

Table A3. All trial items for the rating task.

A.4 Test Material EEG study

Complete list of stimuli (trials and fillers) used in the EEG experiment. Items and conditions are here ordered with all questions in presented first. Items and conditions were randomised for the experiment.

Item	Condition prosody	Condition congruence	Stimuli	Q / S	Trial/Filler
2	NF	C+	Per aprire la porta, bisogna girare la MANIGLIA?	Q	T
2	PF	C+	Per aprire la porta, bisogna GIRARE la maniglia?	Q	T
2	NF	C-	Per aprire la porta, bisogna girare la FRITTATA?	Q	T
2	PF	C-	Per aprire la porta, bisogna GIRARE la frittata?	Q	T
6	NF	C+	Per costruire il palazzo, bisogna seguire il PROGETTO?	Q	T
6	PF	C+	Per costruire il palazzo, bisogna SEGUIRE il progetto?	Q	T
6	NF	C-	Per costruire il palazzo, bisogna seguire la DIETA?	Q	T
6	PF	C-	Per costruire il palazzo, bisogna SEGUIRE la dieta?	Q	T
7	NF	C+	Per preparare l'insalata, bisogna lavare la VERDURA?	Q	T
7	PF	C+	Per preparare l'insalata, bisogna LAVARE la verdura?	Q	T
7	NF	C-	Per preparare l'insalata, bisogna lavare la LAMETTA?	Q	T
7	PF	C-	Per preparare l'insalata, bisogna LAVARE la lametta?	Q	T
8	NF	C+	Per preparare la zuppa, bisogna tagliare le CAROTE?	Q	T
8	PF	C+	Per preparare la zuppa, bisogna TAGLIARE le carote?	Q	T
8	NF	C-	Per preparare la zuppa, bisogna tagliare lo SPARTITO?	Q	T
8	PF	C-	Per preparare la zuppa, bisogna TAGLIARE lo spartito?	Q	T
10	NF	C+	Per costruire un muro, bisogna usare il CEMENTO?	Q	T
10	PF	C+	Per costruire un muro, bisogna USARE il cemento?	Q	T
10	NF	C-	Per costruire un muro, bisogna usare il	Q	T

			CACAO?		
10	PF	C-	Per costruire un muro, bisogna USARE il cacao?	Q	T
12	NF	C+	Per preparare il panino, bisogna tagliare il SALAME?	Q	T
12	PF	C+	Per preparare il panino, bisogna TAGLIARE il salame?	Q	T
12	NF	C-	Per preparare il panino, bisogna tagliare il METALLO?	Q	T
12	PF	C-	Per preparare il panino, bisogna TAGLIARE il metallo?	Q	T
13	NF	C+	Per irrigare campi, bisogna riempire il CANALE?	Q	T
13	PF	C+	Per irrigare i campi, bisogna RIEMPIRE il canale?	Q	T
13	NF	C-	Per irrigare i campi, bisogna riempire il PANINO?	Q	T
13	PF	C-	Per irrigare i campi, bisogna RIEMPIRE il panino?	Q	T
14	NF	C+	Per fare la torta, bisogna tagliare il LIMONE?	Q	T
14	PF	C+	Per fare la torta, bisogna TAGLIARE il limone?	Q	T
14	NF	C-	Per fare la torta, bisogna tagliare il TAPPETO?	Q	T
14	PF	C-	Per fare la torta, bisogna TAGLIARE il tappeto?	Q	T
15	NF	C+	Alle prossime elezioni, bisogna votare il GOVERNO?	Q	T
15	PF	C+	Alle prossime elezioni, bisogna VOTARE il governo?	Q	T
15	NF	C-	Alle prossime elezioni, bisogna votare la CANZONE?	Q	T
15	PF	C-	Alle prossime elezioni, bisogna VOTARE la canzone?	Q	T
17	NF	C+	Per far crescere la pianta, bisogna usare il CONCIME?	Q	T
17	PF	C+	Per far crescere la pianta, bisogna USARE il concime?	Q	T

17	NF	C-	Per far crescere la pianta, bisogna usare il VESTITO?	Q	T
17	PF	C-	Per far crescere la pianta, bisogna USARE il vestito?	Q	T
18	NF	C+	Per usare il fornello a gas, bisogna pagare il METANO?	Q	T
18	PF	C+	Per usare il fornello a gas, bisogna PAGARE il metano?	Q	T
18	NF	C-	Per usare il fornello a gas, bisogna pagare il CESTINO?	Q	T
18	PF	C-	Per usare il fornello a gas, bisogna PAGARE il cestino?	Q	T
19	NF	C+	Per noleggiare la macchina, bisogna fornire la PATENTE?	Q	T
19	PF	C+	Per noleggiare la macchina, bisogna FORNIRE la patente?	Q	T
19	NF	C-	Per noleggiare la macchina, bisogna fornire la MERENDA?	Q	T
19	PF	C-	Per noleggiare la macchina, bisogna FORNIRE la merenda?	Q	T
20	NF	C+	Per fare la dieta, bisogna pesare le PORZIONI?	Q	T
20	PF	C+	Per fare la dieta, bisogna PESARE le porzioni?	Q	T
20	NF	C-	Per fare la dieta bisogna pesare il MATTONE?	Q	T
20	PF	C-	Per fare la dieta bisogna PESARE il mattone?	Q	T
21	NF	C+	Per scrivere l'indirizzo, bisogna sapere la PROVINCIA?	Q	T
21	PF	C+	Per scrivere l'indirizzo, bisogna SAPERE la provincia?	Q	T
21	NF	C-	Per scrivere l'indirizzo, bisogna sapere la PREGHIERA?	Q	T
21	PF	C-	Per scrivere l'indirizzo, bisogna SAPERE la preghiera?	Q	T
22	NF	C+	Quando si affitta la stanza, bisogna pagare la CAUZIONE?	Q	T
22	PF	C+	Quando si affitta la stanza, bisogna	Q	T

			PAGARE la cauzione?		
22	NF	C-	Quando si affitta la stanza, bisogna pagare la BANANA?	Q	T
22	PF	C-	Quando si affitta la stanza, bisogna PAGARE la banana?	Q	T
23	NF	C+	Per conoscere l'astronomia, bisogna studiare i PIANETI?	Q	T
23	PF	C+	Per conoscere l'astronomia, bisogna STUDIARE i pianeti?	Q	T
23	NF	C-	Per conoscere l'astronomia, bisogna studiare il CERVELLO?	Q	T
23	PF	C-	Per conoscere l'astronomia, bisogna STUDIARE il cervello?	Q	T
25	NF	C+	Per produrre i gioielli, bisogna comprare i DIAMANTI?	Q	T
25	PF	C+	Per produrre i gioielli, bisogna COMPRARE i diamanti?	Q	T
25	NF	C-	Per produrre i gioielli, bisogna comprare i PALLONI?	Q	T
25	PF	C-	Per produrre i gioielli, bisogna COMPRARE i palloni?	Q	T
28	NF	C+	Per andare a pescare, bisogna comprare la LICENZA?	Q	T
28	PF	C+	Per andare a pescare, bisogna COMPRARE la licenza?	Q	T
28	NF	C-	Per andare a pescare, bisogna comprare la CASTAGNA?	Q	T
28	PF	C-	Per andare a pescare, bisogna COMPRARE la castagna?	Q	T
31	NF	C+	Per colorare il foglio, bisogna usare le MATITE?	Q	T
31	PF	C+	Per colorare il foglio, bisogna USARE le matite?	Q	T
31	NF	C-	Per colorare il foglio, bisogna usare le MONETE?	Q	T
31	PF	C-	Per colorare il foglio, bisogna USARE le monete?	Q	T
32	NF	C+	Per andare alla partita bisogna comprare il BIGLIETTO?	Q	T

32	PF	C+	Per andare alla partita bisogna COMPRARE il biglietto?	Q	T
32	NF	C-	Per andare alla partita bisogna comprare il QUADERNO?	Q	T
32	PF	C-	Per andare alla partita bisogna COMPRARE il quaderno?	Q	T
36	NF	C+	Per digerire meglio, bisogna filtrare la TISANA?	Q	T
36	PF	C+	Per digerire meglio, bisogna FILTRARE la tisana?	Q	T
36	NF	C-	Per digerire meglio, bisogna filtrare le RICHIESTE?	Q	T
36	PF	C-	Per digerire meglio, bisogna FILTRARE le richieste?	Q	T
38	NF	C+	Per guardare la tv, bisogna spostare il DIVANO?	Q	T
38	PF	C+	Per guardare la tv, bisogna SPOSTARE il divano?	Q	T
38	NF	C-	Per guardare la tv, bisogna spostare il MERCATO?	Q	T
38	PF	C-	Per guardare la tv, bisogna SPOSTARE il mercato?	Q	T
40	NF	C+	Per decorare la stanza, bisogna appendere il DIPINTO?	Q	T
40	PF	C+	Per decorare la stanza, bisogna APPENDERE il dipinto?	Q	T
40	NF	C-	Per decorare la stanza, bisogna appendere il CATINO?	Q	T
40	PF	C-	Per decorare la stanza, bisogna APPENDERE il catino?	Q	T
43	NF	C+	Per restaurare il mobile, bisogna scrostare la VERNICE?	Q	T
43	PF	C+	Per restaurare il mobile, bisogna SCROSTARE la vernice?	Q	T
43	NF	C-	Per restaurare il mobile, bisogna scrostare la PADELLA?	Q	T
43	PF	C-	Per restaurare il mobile, bisogna SCROSTARE la padella?	Q	T
45	NF	C+	Per fare i cannoli, bisogna scolare la	Q	T

			RICOTTA?		
45	PF	C+	Per fare i cannoli, bisogna SCOLARE la ricotta?	Q	T
45	NF	C-	Per fare i cannoli, bisogna scolare la LATTUGA?	Q	T
45	PF	C-	Per fare i cannoli, bisogna SCOLARE la lattuga?	Q	T
47	NF	C+	In questo tipo di gioco, bisogna trovare il TESORO?	Q	T
47	PF	C+	In questo tipo di gioco, bisogna TROVARE il tesoro?	Q	T
47	NF	C-	In questo tipo di gioco, bisogna trovare il RISOTTO?	Q	T
47	PF	C-	In questo tipo di gioco, bisogna TROVARE il risotto?	Q	T
49	NF	C+	Per fare il soffritto, bisogna tritare la CIPOLLA?	Q	T
49	PF	C+	Per fare il soffritto, bisogna TRITARE la cipolla?	Q	T
49	NF	C-	Per fare il soffritto, bisogna tritare il CARTONE?	Q	T
49	PF	C-	Per fare il soffritto, bisogna TRITARE il cartone?	Q	T
51	NF	C+	Per riavviare il computer, bisogna toccare la TASTIERA?	Q	T
51	PF	C+	Per riavviare il computer, bisogna TOCCARE la tastiera?	Q	T
51	NF	C-	Per riavviare il computer, bisogna toccare il BUDINO?	Q	T
51	PF	C-	Per riavviare il computer, bisogna TOCCARE il budino?	Q	T
53	NF	C+	Per servire la minestra, bisogna riempire le SCODELLE?	Q	T
53	PF	C+	Per servire la minestra, bisogna RIEMPIRE le scodelle?	Q	T
53	NF	C-	Per servire la minestra, bisogna riempire le CASELLE?	Q	T
53	PF	C-	Per servire la minestra, bisogna RIEMPIRE le caselle?	Q	T

60	NF	C+	Per chiamare alla messa, bisogna suonare la CAMPANA?	Q	T
60	PF	C+	Per chiamare alla messa, bisogna SUONARE la campana?	Q	T
60	NF	C-	Per chiamare alla messa, bisogna suonare la TARANTA?	Q	T
60	PF	C-	Per chiamare alla messa bisogna SUONARE la taranta?	Q	T
61	FI	C+	Per fare la cena, la mamma ha comprato le melanzane?	Q	F
62	FI	C-	Per viaggiare in treno, compri sempre il processo?	Q	F
63	FI	C+	Per togliere il dente del giudizio, vai dal dentista?	Q	F
64	FI	C-	Quando l'aereo si ferma, bisogna scendere dalle nuvole?	Q	F
65	FI	C+	Per allargare i tuoi orizzonti, hai girato per il mondo?	Q	F
66	FI	C-	Per guadagnare soldi, bisogna avere un sasso?	Q	F
67	FI	C+	Per diventare un pilota, dovrai fare l'addestramento?	Q	F
68	FI	C-	Durante il processo, ascoltarono le difese delle chiavi?	Q	F
69	FI	C+	Appena é arrivato, ha baciato in bocca la ragazza?	Q	F
70	FI	C-	Per imbiancare la parete, hai dovuto usare il rullo?	Q	F
71	FI	C+	Per lo spavento, il passante chiamò subito la polizia?	Q	F
72	FI	C-	Per scrivere una mail, devi accedere all'università?	Q	F
73	FI	C+	Per risolvere il problema, devi conoscere la teoria?	Q	F
74	FI	C-	Sei andato al lago per camminare sulle braci?	Q	F
75	FI	C+	Per rimanere in forma, l'atleta si allena ogni giorno?	Q	F
76	FI	C-	Se ti sei ammalato, devi prendere le navi?	Q	F

77	FI	C+	Per essere sicura, hai chiesto la ricevuta?	Q	F
78	FI	C-	Quando andrai in piscina, in testa metterai la corona?	Q	F
79	FI	C+	Per essere preparato, hai fatto tutti i compiti?	Q	F
80	FI	C+	Per usare il dizionario, devi imparare l'alfabeto?	Q	F
81	FI	C-	Per non sprecare energia, aveva spento la sigaretta?	Q	F
82	FI	C+	Per godersi la propria libertà, ha lasciato il fidanzato?	Q	F
83	FI	C-	Il macellaio lavora tagliando la carrozzeria?	Q	F
84	FI	C-	Per essere educato, hai portato un vassoio di strade?	Q	F
85	FI	C+	Per non disturbare nessuno, avevi indossato le cuffie?	Q	F
86	FI	C-	Devo aggiustare l'armadio a cui mancano le spiagge?	Q	F
87	FI	C+	Per avvertire del suo arrivo, suonò il campanello?	Q	F
88	FI	C+	Per dar da bere alle piante, prendi l'innaffiatoio?	Q	F
89	FI	C-	Per i tuoi piedi, vuoi provare nuovi elicotteri?	Q	F
90	FI	C-	Per sedermi comodo, posso prendere quella felicità?	Q	F
91	FI	C+	Sei andato dal veterinario per il mio canarino?	Q	F
92	FI	C-	Per riciclare la carta, bisogna fare la penitenza?	Q	F
93	FI	C-	Per essere elegante devi indossare un accappatoio?	Q	F
94	FI	C-	In una VERA amicizia, si dovrebbero aiutare gli ospedali?	Q	F
95	FI	C+	Un SOLO testimone, racconterà l'intera verità?	Q	F
96	FI	C+	IERI, ti sei dimenticato il portafoglio?	Q	F
97	FI	C-	Le PIANTE, hanno bisogno di tanta	Q	F

			mozzarella?		
98	FI	C-	Per GUARDARE quel cartellone, devi mettere la musica?	Q	F
99	FI	C+	Per SOCCORRERE il ferito, bisogna chiamare il medico?	Q	F
100	FI	C+	Per il FREDDO, ti è venuta la pelle d'oca?	Q	F
101	FI	C+	Per fare FELICI i tuoi genitori, hai visitato il museo?	Q	F
102	FI	C-	IERI, siete andati a mangiare nella nuova discarica?	Q	F
103	FI	C+	Per il LORO anniversario, hai comprato i fiori?	Q	F
104	FI	C+	Prima di AGGIUNGERE il pomodoro, devi cuocere la pasta?	Q	F
105	FI	C-	Per fare la SPESA, si deve prendere l'aereo?	Q	F
106	FI	C+	Per prendere il CIBO, il gatto è salito sul tavolo?	Q	F
107	FI	C-	Per giocare a CALCIO, è necessaria la protesta?	Q	F
108	FI	C+	Ogni sabato SERA, tuo fratello balla la salsa?	Q	F
109	FI	C-	Per essere PROMOSSI, bisogna presentarsi all'audizione?	Q	F
110	FI	C-	Quando si viaggia su un AEREO, bisogna allacciare il grembiule?	Q	F
111	FI	C+	Per accontentare il PADRE, è diventato avvocato?	Q	F
112	FI	C-	Per imparare la BIOLOGIA, devi leggere tanti indirizzi?	Q	F
113	FI	C+	Per farci vedere le FOTO, avevi portato l'album?	Q	F
114	FI	C-	Per sapere il FINALE, devi leggere il numero?	Q	F
115	FI	C-	Lo scrittore impiegò TRE anni, per scrivere il nome?	Q	F
116	FI	C+	Per il bene di TUTTI, troveremo un accordo?	Q	F
117	FI	C-	Chi lavora in OFFICINA, deve riparare le	Q	F

			ingiustizie?		
118	FI	C+	Per il picnic di DOMANI, bisogna scegliere il luogo adatto?	Q	F
119	FI	C-	Oggi ti sei lavato i DENTI con lo pneumatico?	Q	F
147	FI	C+	Per fare la cena, devi cuocere la pasta?	Q	F
1	NF	C+	Per arieggiare la stanza, bisogna aprire la FINESTRA	S	T
1	PF	C+	Per arieggiare la stanza, bisogna APRIRE la finestra	S	T
1	NF	C-	Per arieggiare la stanza, bisogna aprire la LATTINA	S	T
1	PF	C-	Per arieggiare la stanza, bisogna APRIRE la lattina	S	T
3	NF	C+	Per andare alla festa, bisogna portare la CAMICIA	S	T
3	PF	C+	Per andare alla festa, bisogna PORTARE la camicia	S	T
3	NF	C-	Per andare alla festa, bisogna portare la SARDINA	S	T
3	PF	C-	Per andare alla festa, bisogna PORTARE la sardina	S	T
4	NF	C+	Per proteggere l'ambiente, bisogna salvare la FORESTA	S	T
4	PF	C+	Per proteggere l'ambiente, bisogna SALVARE la foresta	S	T
4	NF	C-	Per proteggere l'ambiente, bisogna salvare la SERATA	S	T
4	PF	C-	Per proteggere l'ambiente, bisogna SALVARE la serata	S	T
5	NF	C+	Prima di bere il vino, bisogna stappare la BOTTIGLIA	S	T
5	PF	C+	Prima di bere il vino, bisogna STAPPARE la bottiglia	S	T
5	NF	C-	Prima di bere il vino, bisogna stappare le ORECCHIE	S	T
5	PF	C-	Prima di bere il vino, bisogna STAPPARE le orecchie	S	T
9	NF	C+	Quando ci si taglia, bisogna coprire la	S	T

			FERITA		
9	PF	C+	Quando ci si taglia, bisogna COPRIRE la ferita	S	T
9	NF	C-	Quando ci si taglia, bisogna coprire la GRANITA	S	T
9	PF	C-	Quando ci si taglia, bisogna COPRIRE la granita	S	T
11	NF	C+	Quando si va in vacanza, bisogna riempire la VALIGIA	S	T
11	PF	C+	Quando si va in vacanza, bisogna RIEMPIRE la valigia	S	T
11	NF	C-	Quando si va in vacanza, bisogna riempire il TEGAME	S	T
11	PF	C-	Quando si va in vacanza, bisogna RIEMPIRE il tegame	S	T
16	NF	C+	Per non avere freddo, bisogna scaldare il LOCALE	S	T
16	PF	C+	Per non avere freddo, bisogna SCALDARE il locale	S	T
16	NF	C-	Per non avere freddo, bisogna scaldare il SAPONE	S	T
16	PF	C-	Per non avere freddo, bisogna SCALDARE il sapone	S	T
24	NF	C+	Per preparare il caffè, bisogna scaldare la TAZZINA	S	T
24	PF	C+	Per preparare il caffè, bisogna SCALDARE la tazzina	S	T
24	NF	C-	Per preparare il caffè, bisogna scaldare la PISCINA	S	T
24	PF	C-	Per preparare il caffè, bisogna SCALDARE la piscina	S	T
26	NF	C+	Per eliminare la forfora, bisogna lavare i CAPELLI	S	T
26	PF	C+	Per eliminare la forfora, bisogna LAVARE i capelli	S	T
26	NF	C-	Per eliminare la forfora, bisogna lavare gli ARMADI	S	T
26	PF	C-	Per combattere i germi, bisogna LAVARE gli armadi	S	T

27	NF	C+	Per preparare il tè, bisogna usare la BUSTINA	S	T
27	PF	C+	Per preparare il tè, bisogna USARE la bustina	S	T
27	NF	C-	Per preparare il tè, bisogna usare la POMATA	S	T
27	PF	C-	Per preparare il tè, bisogna USARE la pomata	S	T
29	NF	C+	Per mantenere l'ordine, bisogna pulire la CUCINA	S	T
29	PF	C+	Per mantenere l'ordine, bisogna PULIRE la cucina	S	T
29	NF	C-	Per mantenere l'ordine, bisogna pulire la TROMBETTA	S	T
29	PF	C-	Per mantenere l'ordine, bisogna PULIRE la trombetta	S	T
30	NF	C+	Per scalare la montagna, bisogna portare gli SCARPONI	S	T
30	PF	C+	Per scalare la montagna, bisogna PORTARE gli scarponi.	S	T
30	NF	C-	Per scalare la montagna, bisogna portare gli INVITI.	S	T
30	PF	C-	Per scalare la montagna, bisogna PORTARE gli inviti.	S	T
33	NF	C+	Per cambiare rotta, bisogna girare il TIMONE	S	T
33	PF	C+	Per cambiare rotta, bisogna GIRARE il timone	S	T
33	NF	C-	Per cambiare rotta, bisogna girare il CUSCINO	S	T
33	PF	C-	Per cambiare rotta, bisogna GIRARE il cuscino	S	T
34	NF	C+	Per orientarsi in città, bisogna guardare la CARTINA	S	T
34	PF	C+	Per orientarsi in città, bisogna GUARDARE la cartina	S	T
34	NF	C-	Per orientarsi in città, bisogna guardare la PATATA	S	T
34	PF	C-	Per orientarsi in città, bisogna	S	T

			GUARDARE la patata		
35	NF	C+	Per scaldarsi i piedi, bisogna prendere la COPERTA	S	T
35	PF	C+	Per scaldarsi i piedi, bisogna PRENDERE la coperta	S	T
35	NF	C-	Per scaldarsi, bisogna prendere la RUBRICA	S	T
35	PF	C-	Per scaldarsi, bisogna PRENDERE la rubrica	S	T
37	NF	C+	Per appendere il chiodo, bisogna usare il MATELLO	S	T
37	PF	C+	Per appendere il chiodo, bisogna USARE il martello	S	T
37	NF	C-	Per appendere il chiodo, bisogna usare il SETACCIO	S	T
37	PF	C-	Per appendere il chiodo, bisogna USARE il setaccio	S	T
39	NF	C+	Dopo aver scontato la pena, bisogna lasciare la PRIGIONE	S	T
39	PF	C+	Dopo aver scontato la pena, bisogna LASCIARE la prigione	S	T
39	NF	C-	Dopo aver scontato la pena, bisogna lasciare la SFILATA	S	T
39	PF	C-	Dopo aver scontato la pena, bisogna LACIARE la sfilata	S	T
41	NF	C+	Per costruire la macchina, bisogna montare il MOTORE	S	T
41	PF	C+	Per costruire la macchina, bisogna MONTARE il motore	S	T
41	NF	C-	Per costruire la macchina, bisogna montare il FILMATO	S	T
41	PF	C-	Per costruire la macchina, bisogna MONTARE il filmato	S	T
42	NF	C+	Per non puzzare di sudore, bisogna lavare le ASCELLE	S	T
42	PF	C+	Per non puzzare di sudore, bisogna LAVARE le ascelle	S	T
42	NF	C-	Per non puzzare di sudore, bisogna lavare le ORTENSIE	S	T

42	PF	C-	Per non puzzare di sudore, bisogna LAVARE le ortensie	S	T
44	NF	C+	Per costruire la casa, bisogna comprare il TERRENO	S	T
44	PF	C+	Per costruire la casa, bisogna COMPRARE il terreno	S	T
44	NF	C-	Per costruire la casa, bisogna comprare il GIORNALE	S	T
44	PF	C-	Per costruire la casa, bisogna COMPRARE il giornale	S	T
46	NF	C+	Per giocare a basket, bisogna centrare il CANESTRO	S	T
46	PF	C+	Per giocare a basket, bisogna CENTRARE il canestro	S	T
46	NF	C-	Per giocare a basket, bisogna centrare il PROBLEMA	S	T
46	PF	C-	Per giocare a basket, bisogna CENTRARE il problema	S	T
48	NF	C+	Per votare consapevolmente, bisogna capire il PROGRAMMA	S	T
48	PF	C+	Per votare consapevolmente, bisogna CAPIRE il programma	S	T
48	NF	C-	Per votare consapevolmente, bisogna capire il LATINO	S	T
48	PF	C-	Per votare consapevolmente, bisogna CAPIRE il latino	S	T
50	NF	C+	Per mangiare la pasta, bisogna tenere la FORCHETTA	S	T
50	PF	C+	Per mangiare la pasta, bisogna TENERE la forchetta	S	T
50	NF	C-	Per mangiare la pasta, bisogna tenere la CARTELLA	S	T
50	PF	C-	Per mangiare la pasta, bisogna TENERE la cartella	S	T
52	NF	C+	Per preparare la pastasciutta, bisogna bollire gli SPAGHETTI	S	T
52	PF	C+	Per preparare la pastasciutta, bisogna BOLLIRE gli spaghetti	S	T
52	NF	C-	Per preparare la pastasciutta, bisogna	S	T

			bollire le SIRINGHE		
52	PF	C-	Per preparare la pastasciutta, bisogna BOLLIRE le siringhe	S	T
54	NF	C+	Per prendere il treno, bisogna trovare la STAZIONE	S	T
54	PF	C+	Per prendere il treno, bisogna TROVARE la stazione	S	T
54	NF	C-	Per prendere il treno, bisogna trovare la CASCATA	S	T
54	PF	C-	Per prendere il treno, bisogna TROVARE la cascata	S	T
55	NF	C+	Per non schizzare d'acqua, bisogna montare la TENDINA	S	T
55	PF	C+	Per non schizzare d'acqua, bisogna MONTARE la tendina	S	T
55	NF	C-	Per non schizzare d'acqua, bisogna montare il MANUBRIO	S	T
55	PF	C-	Per non schizzare d'acqua, bisogna MONTARE il manubrio.	S	T
56	NF	C+	Per seminare il campo, bisogna guidare il TRATTORE	S	T
56	PF	C+	Per seminare il campo, bisogna GUIDARE il trattore	S	T
56	NF	C-	Per seminare il campo, bisogna guidare il PARTITO	S	T
56	PF	C-	Per seminare il campo, bisogna GUIDARE il partito	S	T
57	NF	C+	Prima di aprire un negozio, bisogna creare il PARCHEGGIO	S	T
57	PF	C+	Prima di aprire un negozio, bisogna CREARE il parcheggio	S	T
57	NF	C-	Prima di aprire un negozio, bisogna creare il RIFUGIO	S	T
57	PF	C-	Prima di aprire un negozio, bisogna CREARE il rifugio	S	T
58	NF	C+	Per il compleanno del bambino, bisogna comprare il REGALO	S	T
58	PF	C+	Per il compleanno del bambino, bisogna COMPRARE il regalo	S	T

58	NF	C-	Per il compleanno del bambino, bisogna comprare la BENZINA	S	T
58	PF	C-	Per il compleanno del bambino, bisogna COMPRARE la benzina.	S	T
59	NF	C+	Per ridurre l'inquinamento, bisogna fermare le INDUSTRIE	S	T
59	PF	C+	Per ridurre l'inquinamento, bisogna FERMARE le industrie	S	T
59	NF	C-	Per ridurre l'inquinamento, bisogna fermare il PORTONE	S	T
59	PF	C-	Per ridurre l'inquinamento, bisogna FERMARE il portone	S	T
120	FI	C+	Dopo aver mangiato, BISOGNA lavare i piatti	S	F
121	FI	C-	Per prendere appunti, DEVI cercare una bicicletta	S	F
122	FI	C-	Le immondizie BISOGNA buttarle nel water	S	F
123	FI	C+	I legumi sono una BUONA fonte di proteine	S	F
124	FI	C+	Per preparare il PROFITEROL, serve la cioccolata	S	F
125	FI	C-	Per perdere PESO, è necessario correre al senato	S	F
126	FI	C+	Per accendere il FUOCO, ho dovuto usare i fiammiferi	S	F
127	FI	C-	Se vuoi LICENZIARTI, devi parlare con il tuo idraulico	S	F
128	FI	C+	Per rinnovare la CASA, ho dovuto pagare tanto	S	F
129	FI	C-	Per andare a SCIARE, devi salire sulla metropolitana	S	F
130	FI	C+	Per finire in TEMPO, è necessario sollecitare i lavoratori	S	F
131	FI	C-	Per asciugare la MACCHINA prendi il chiavistello	S	F
132	FI	C-	Per votare il SINDACO, è necessaria la matematica	S	F
133	FI	C+	Per cercare LAVORO, ho dovuto	S	F

			chiamare l'agenzia		
134	FI	C+	Per gli SPOSTAMENTI, furono usati gli aerei	S	F
135	FI	C-	Se vuoi andare in CINA, devi prendere la falciatrice	S	F
136	FI	C+	Il ministro PRESENTÒ la legge al parlamento	S	F
137	FI	C+	Per un CONTROLLO, ho contattato il pediatra	S	F
138	FI	C-	Per accendere la CANDELA, è necessario avere un fagiolo	S	F
139	FI	C+	Per la grigliata di DOMANI, hanno già fatto la spesa	S	F
140	FI	C-	L'omicidio si deve PUNIRE con tanti anni di felicità	S	F
141	FI	C+	A difesa del CASTELLO, era stato messo il cane da guardia	S	F
142	FI	C-	Le MERCI al supermercato hanno tutte un cortile	S	F
143	FI	C-	É NECESSARIO, che le ricette vengano seguite alla fontana	S	F
144	FI	C-	Per ENTRARE all'università, dovrò sostenere un frigorifero	S	F
145	FI	C+	Per il MATRIMONIO, ho acquistato la torta con la panna	S	F
146	FI	C+	Per GUARIRE, assumo regolarmente i medicinali	S	F
148	FI	C+	Per dipingere, l'artista ha comprato i pennelli	S	F
149	FI	C-	Per dormire, devi chiudere i macelli	S	F
150	FI	C+	Quando mangi il pesce, assicurati che sia fresco	S	F
151	FI	C-	Per prendere il diploma, bisogna passare il tempo	S	F
152	FI	C+	Durante l'estate, si mangia spesso il gelato	S	F
153	FI	C-	Per stare in salute, devi mangiare tanta colla	S	F
154	FI	C+	Per prepararmi alla gara, alleno tutti i giorni il mio corpo	S	F

155	FI	C-	Per pagare l'affitto, è necessario avere i pidocchi	S	F
156	FI	C+	Per realizzare lo spettacolo, hanno organizzato i provini	S	F
157	FI	C+	Se hai tante cose inutili, devi metterle in cantina	S	F
158	FI	C+	Hai donato parte dei tuoi soldi per una buona causa	S	F
159	FI	C-	Dormire la notte, é necessario per una buona pianola	S	F
160	FI	C+	Per risolvere il caso, gli investigatori chiamano le spie	S	F
161	FI	C-	É doveroso che la legge venga rispettata da pochi	S	F
162	FI	C+	Per il ritardo gli istruttori hanno punito gli allievi	S	F
163	FI	C-	Quando hai fatto il bucato, devi poi stirare i muscoli	S	F
164	FI	C+	Per lavare i capelli, ti serve utilizzare lo shampoo	S	F
165	FI	C+	Perché crescano i fiori, é necessario aspettare molte settimane	S	F
166	FI	C-	Durante la primavera, i prati si riempiono di complimenti	S	F
167	FI	C+	Per maggiore chiarezza, leggerai il resoconto delle indagini	S	F
168	FI	C-	Per fare la glassa dei biscotti, devi sciogliere il catarro	S	F
169	FI	C+	Per non dimenticare niente, ho disegnato lo schema	S	F
170	FI	C-	Per ottenere il formaggio, bisogna lavorare il ferro	S	F
171	FI	C+	Per l'esame, ho dovuto preparare la presentazione	S	F
172	FI	C-	Prima di andare a dormire, devi spegnere la speranza	S	F
173	FI	C+	In aiuto dei profughi, hanno fatto la raccolta fondi	S	F
174	FI	C-	Per arrivare in cima alla montagna, devi	S	F

			imboccare il bambino		
175	FI	C+	Per lavorare in Germania, devi imparare il tedesco	S	F
176	FI	C-	Per scrivere al computer, c'è bisogno della sepoltura	S	F
177	FI	C+	Per tenersi aggiornata, la mamma legge le notizie	S	F
178	FI	C-	Al ristorante, i camerieri vogliono portare le lucertole	S	F
179	FI	C+	Per l'inizio del mese, stiamo allestendo la mostra	S	F
180	FI	C-	Per mangiare l'insalata, devi condirla con l'acido	S	F

Table A4. List of stimuli (T = trials and F = fillers) submitted to listeners. Each trial item (first column on the left) is realised in two prosodic conditions (NF = narrow contrastive focus, PF = post-focal position), two congruence conditions (C+ = congruent, C- = incongruent). Items can be either questions (Q) or statements (S).