SUPERARE L'EVANESCENZA DEL PARLATO

Un vademecum per il trattamento digitale di dati linguistici

a cura di
Giuliano Bernini - Ada Valentini
Jacopo Saturno - Lorenzo Spreafico

BERGAMO UNIVERSITY PRESS
sestante edizioni

UNIVERSITÀ DEGLI STUDI DI BERGAMO | Dipartimento di Lingue, Letterature e Culture Straniere

Sestante Edizioni - Bergamo
www.sestanteedizioni.it

SUPERARE L'EVANESCENZA DEL PARLATO
Un vademecum per il trattamento digitale di dati linguistici
Giuliano Bernini - Ada Valentini - Jacopo Saturno - Lorenzo Spreafico (A cura di)
p. 262 cm. 15,5x22,0
ISBN: 978-88-6642-369-0

Printed in Italy
by Sestanteinc - Bergamo

In copertina: "Evangelista, or letter-writer, and his clients". Immagine tratta da Brown, Robert. 1894. The Countries of the World: being a popular description of the various continents, islands, rivers, seas, and peoples of the globe [with plates]. Londra: Cassell, Petter & Galpin. https://www.flickr.com/photos/britishlibrary/11226480883/
Superare l'evanescenza del parlato: lo sforzo può comportare lo sgomento riflesso nel volto dello scriba di fronte ai modi di parlare di personaggi tanto diversi.

SUPERARE L'EVANESCENZA DEL PARLATO

Un vademecum per il trattamento digitale di dati linguistici

a cura di
Giuliano Bernini - Ada Valentini
Jacopo Saturno - Lorenzo Spreafico

BERGAMO UNIVERSITY PRESS
sestante edizioni

Direttore responsabile
Prof. Giuliano Bernini

Biblioteca di Linguistica e Filologia

6.

Superare l'evanescenza del parlato
Un vademecum per il trattamento digitale di dati linguistici

a cura di
Giuliano Bernini - Ada Valentini
Jacopo Saturno - Lorenzo Spreafico

Questo volume è stato stampato con il contributo del Dipartimento
di Lingue, Letterature e Culture Straniere dell'Università degli Studi di Bergamo.

Contributi rivisti dai curatori.

Indice

Cinzia Avesani
(Consiglio Nazionale delle Ricerche)

Barbara Gili Fivela
(Università del Salento)

Analysing Prosody: Methods, issues, and hints on crosslinguistic comparison and L2 learning[1]

1. *Introduction*

During an act of speaking, the flow of speech is not a simple concatenation of segments, but consonants and vowels are modulated by principled variations of fundamental frequency (F0), duration, intensity. These acoustic modulations are perceived as variations in pitch, length and loudness of speech stretches, but they affect single sound segments to various degrees. They are the acoustic and perceptual reflexes of how the sounds are articulated: segments that have higher fundamental frequency are produced with a higher rate of vocal fold vibration, determined by the configuration of the larynx, the subglottal pressure, and the degree of oral closure; segments that have longer duration are produced with speech gestures that are longer (and have phases which are not truncated, e.g., Byrd & Saltzman 2003); segments that have higher intensity are produced with more articulatory effort and higher subglottal pressure.

These are the parameters that, besides affecting each segment, give rise to a set of phonological phenomena such as stress, rhythm and timing, tone and intonation, usually referred to with the cover term of Prosody.

A broad definition of prosody refers to those non-segmental speech events that participate in the organization of lexicon and syntax and play a decisive role in the semantic and pragmatic interpretation of a given utterance. Non-segmental is here preferred to *suprasegmentals*, a term originally coined by Lehiste (1970) and used – often in the past, but still sometime used nowadays – interchangeably with prosody. With it, she

[1] This work has been designed, discussed and conducted in close collaboration between the two authors. Main responsibility in writing the paper is divided as follows: Avesani: §1, 2, 2.1, 2.3, 3, 4, 5; Gili Fivela: §2.2, 6, 6.1, 7, 7.1; Avesani and Gili Fivela: 6.2, 8.

intended to indicate a set of linguistic phenomena that span over domains larger than a segment, like syllables, phrases and utterances. But the term also evokes that prosodic events stand "above" the segments. Using "suprasegmentals" with such denotation instead of prosody can be misleading, as it overlooks one fundamental aspect of speech: there are no utterances of natural languages in which segments are unaffected by prosody, and because prosody is an intrinsic and unavoidable part of any language, investigating speech events without reference to it misses important aspects of how speech is organized. A conspicuous number of experimental studies has shown that prosody affects all aspects of the speech signal. For example, not only elements found in prosodically prominent positions (i.e. stressed and accented) are longer, higher, louder, and more fully articulated than elements in prosodically weak positions (i.e. unstressed and unaccented), but also elements occurring at the edges of prosodic units are affected in their inner articulation: compared to consonants occurring in unit-internal position, consonants occurring at initial edges are strengthened (Fougeron & Keating 1997; Cho & Keating 2001; Keating *et al*. 2004) and segments and syllables occurring at final edges are regularly lengthened (e.g. Beckman & Edwards 1990; Edwards *et al*. 1991).

A better definition of prosody is due to Beckman (1996) who refers to prosody as the "organizational structure of speech". As a musical score is organized in notes, measures, musical sentences and so on, prosody organizes speech in prosodic constituents, dividing the flow of speech in "chunks of information" that help listeners to parse discourse in meaningful units for further linguistic information (syntactic, semantic, conversational: i.e. turns). In line with other proposals regarding prosody (e.g., rule-based and syntactically related as in Nespor & Vogel 1986; Selkirk 1984), the definition suggests that prosodic units stand in a hierarchical relation, on par with the hierarchy of syntactic constituents that determine the order of morphemes inside words and of words inside sentences. Independently of the specific definition adopted, the subdivision of speech in prosodic units and the organization of such units in a structure is the first task of prosody, referred to as *phrasing*. A second main function of prosody is to mark prominence relations within each prosodic constituent with a language-specific variable combination of acoustic parameters: the abovementioned duration, intensity, F0. This is the second task of prosody: *highlighting* in a principled way some elements within each prosodic unit making them to stand out as prominent

at lexical and postlexical level, and marking the prominence relations that hold among them.

One of the most investigated prosodic phenomenon is intonation, which has been defined in the literature at least in a narrow and in a broad sense. One definition equates intonation with the melody of speech, restricting the use of the term to refer to the variations of pitch in the course of the utterance (t'Hart *et al*. 1990), which are not determined by lexical distinctions as in tone languages (a.o., Gussenhoven 2007) and are used by the speakers "to mark pragmatic force of the information in an utterance" (Pierrehumbert 1999). More generally, intonation can be defined as the "linguistically structured and pragmatically meaningful" modulation of pitch (Arvaniti 2012: 265). In the broader use, which refers to the linguistic structure and pragmatic meaning, the term intonation includes also variations in loudness, length and segmental quality besides pitch, basically equating intonation with prosody. However, even definitions focussing on pitch modulation and its link to linguistic information do not assume that pitch is the only correlate of intonation. Rather, pitch is considered as the main correlate, directly linked with phonological representations in the grammar, while variation in loudness, length and segmental quality co-occur with its modulation.

In spoken language, intonation serves a variety of linguistic and paralinguistic functions, ranging from speech act information (assertions, questions, commands etc.), information structure or information packaging (topic, focus, background), information status (given vs. new information), knowledge state (or epistemic position of the speaker with respect to the information exchange), illocutionary force, affective state, emotions. Since the course of fundamental frequency is the main exponent of intonation through which the speakers convey such an array of communicative functions, it is clear that determining the structure of pitch modulations and unravelling the nature of intonational meaning is a challenging task. The main difficulties reside, first, in defining the primitives that make up the pitch contour, since linguistically related F0 changes are not as easily identifiable as in tone languages; and, second, in determining the meaning associated with those primitives, as in the intonational domain meaning is represented by pragmatic and information structure contrasts, which are notoriously more difficult to determine than stable lexical contrasts in the segmental domain. Therefore, both intonational form and its connection with segmental material and intonational meaning are hard to pinpoint.

One solution adopted in the literature is to treat intonation contours as gestalts, an approach that starting from Bolinger (1951) has been embraced by many scholars until recently (e.g., Cooper & Sorensen 1981; Hirst & Di Cristo 1998; Xu 2005; Grabe *et al*. 2003) and, at least to a certain extent, applied also to Italian in the *Language into Act* model proposed by Cresti and collaborators (e.g. Cresti & Moneglia 2018). In those studies, intonational contours are deemed to be holistic entities that directly reflect certain structural or functional aspect of speech, such as the depth of a syntactic boundary (Cooper & Sorensen 1981), or a speech act (e.g. Cresti 2005; Moneglia 2006; Cresti & Moneglia 2018).

This approach faces important problems, though. The contours in Fig. 1 help illustrating the first point. The figure represents the F0 contours of two utterances differing in length: *È Marina?* 'Is she Marina?' on the left and *È Marina la ragazza che ti piace?* 'Is it Marina the girl that you like?' on the right. Globally, both contours share what at a first view could be considered the "same" rise-fall-rise F0 pattern. Although they show some similarities, they cannot be said to be identical: after the initial rise-

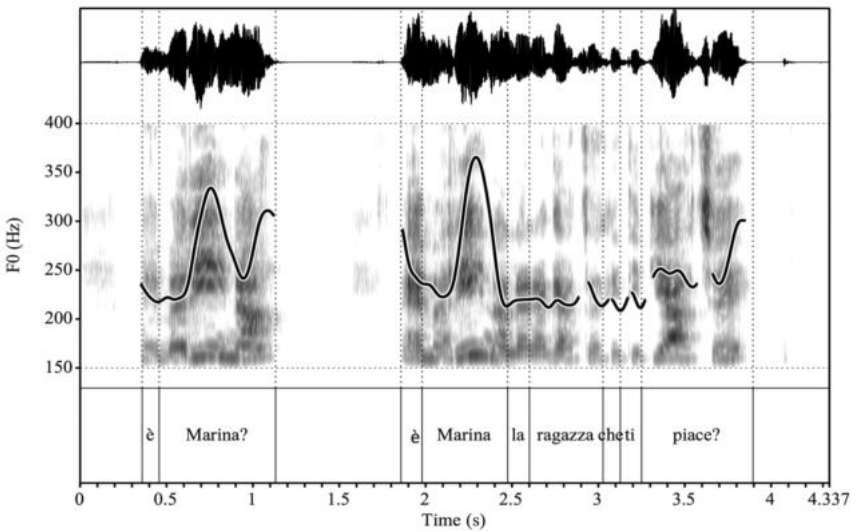

Fig. 1. Intonational contour of the sentences È Marina? *'Is she Marina?' and* È Marina la ragazza che ti piace? *'Is it Marina the girl that you like?'*

fall movement common to both contours, a low plateau follows in the contour of the longer utterance on the right side of the figure. Notwithstanding the difference in their form, both contours are used by the speakers for producing *wh*-questions and are perceived as conveying the same meaning. The same difference in rise-fall-rise contours found in utterances of different length sharing the same communicative meaning are also found in other languages (for example in English, Greek and Polish). As in Italian, in none of those cases it is possible to consider one contour as a "stretched" or "squeezed" version of the other as it would be expected if melodies were undivided wholes (Arvaniti & Ladd 2009; Arvaniti 2012).

This example shows that the shape of intonational contours that share the same pragmatic meaning (asking a question) can vary considerably as a function of the segmental material with which they are coproduced. However, we can make sense of this variation if we factor out the components of the rise-fall-rise pattern and take into consideration the overall prosodic structure of the utterance. That is, if we do not consider the difference of the two contours as the by-product of random variation, but if we take into consideration the main and higher level functions of prosody: highlighting and phrasing. In both utterances the highest prominence is produced on the stressed syllable of word *Marina* 'Marina', which carries in both cases the same rise-fall pattern, and the final rise of the pattern is synchronized with the end of the contour in both cases. However, while in the short utterance the final rise appears at the end of the word that carries the main prominence, and makes the word Marina the docking site of a combined rise-fall-rise pattern, in the longer utterance it appears as a separate pitch event due to the number of segments that separate the last accented syllable from the end of the utterance. Thus, it appears that parts of the melody coordinate independently with parts of the segmental string (Arvaniti 2007; Ladd 2008: chapter 2).

The example illustrates what has been shown in the literature with plenty of evidence, i.e. that when tunes are realized in utterances with different length and metrical structure their form differs substantially. The idea that pitch contours are not non-analyzable gestalts but have an internal structure has been acknowledged also by a configurational approach such as the IPO (Institute for Perception Research, Eindhoven) model of intonation (t'Hart *et al*. 1990). In the IPO model the largest

descriptive unit of intonation is the pitch contour, but contours are decomposed in configurations (Prefix, Root, Suffix) that consist of pitch movements. The pitch movements are specified in terms of features (their direction, timing with regard to syllable boundaries, rate of change, size) and are distinguished based on their function to lend prominence or not to the syllable on which they occur (prominence-lending movements that co-occur with stressed syllables vs. non-prominence lending movements). Importantly, t'Hart and colleagues observed that the same sequence of pitch movements can be distributed differently over an utterance (t'Hart et al. 1990: 98): they could appear either all together on a single syllable or as separated by intervening syllables without affecting the perceptual identity of the contour. By using stylisation techniques, the authors showed that only certain aspects of the contour are important for the listeners while the global shape of the contour is not. Overall, their work showed what will be repeatedly demonstrated in the following years, that parts of the melody appear to coordinate independently with parts of the segmental string.

Beside facing a problem in accounting for intonational form, a holistic approach faces the major difficulty of maintaining a one-to-one relationship between form and meaning. Since the pragmatic functions performed by intonation are manifold, postulating a direct form-function relationship necessarily leads to identifying a specific form for each different meaning across different utterances, even where the diversity between one melody and the other is not justifiable on acoustic and perceptual ground. Typically, different pragmatic functions empirically defined are associated with putatively different types of contours, but their belonging to contrasting categories or to variants of the same category is not always proved on experimental (acoustically, perceptually) ground (e.g. Cresti & Moneglia 2018).

On the contrary, in the past decades many authors have noticed that the mapping between form and function is not a one-to-one, but rather a many-to-many relation: the same melody can be used to convey different meanings, and the same meaning can be expressed by different melodies (a.o., Pike 1945; Lehiste 1970; Ladd 2008; Grice *et al*. 2005; for Italian, Gili Fivela 2008; Gili Fivela *et al*. 2015). Problems like these make it clear that viewing melodies as composed by smaller, phonologically relevant elements is more likely to be successful in accounting for intonational meaning, especially if it is assumed that the mapping between the

superficial form of pitch modulation and linguistic functions/meaning is not direct, but it is mediated by the phonological structure.

2. *Autosegmental-Metrical theory of intonation*

A major breakthrough in our understanding of intonation was achieved with the advent of the *Autosegmental-Metrical theory of intonation* (henceforth, AM). The theory has its origin in Pierrehumbert's dissertation (1980) who incorporates in it the insights of two previous influential theses, Liberman (1975) and Bruce (1977), and it has been further developed into the current model particularly by Beckman and Pierrehumbert (Beckman & Pierrehumbert 1986; Pierrehumbert & Beckman 1988). The basic tenet of the theory is that intonation is part of the grammar and has a phonological structure: it is possible to characterize contours in terms of a string of categorically distinct elements and to provide a mapping from phonological elements to continuous phonetic parameters (Ladd 2008: 43). The term "Autosegmental-Metrical" was coined by Ladd (1996) as it reflects the intellectual heritage and the principles of Intonational Phonology (Bruce 1977; Pierrehumbert 1980; Gussenhoven 1984; Liberman & Pierrehumbert 1984; Beckman & Pierrehumbert 1986; Pierrehumbert & Beckman 1988) and those of Metrical and Prosodic Phonology, with reference to the domains proposed within Prosodic Phonology and the prominence relations holding within them (Liberman 1975; Liberman & Prince 1977; Selkirk 1984, 2004; Nespor & Vogel 1986, 2007).

2.1 Basic elements

According to AM, intonation is represented in terms of a string of static H(igh) and L(ow) tones. H and L tones are the primitives of the abstract phonological representation and are phonetically realized as targets in the F0 contour, typically peaks and dips in the contour. The contrast between H and L tones is paradigmatic, i.e. *ceteris paribus*, a H tone is higher than a L tone in the same context, but the phonetic height of each target is defined in relative terms, with reference to the speaker's range: a L tone is realized as a low target on the hypothetical bottom line of the speaker's range and a H tone as a high target on the topline (Pierrehumbert 1980: 69 and following). Given that the speaker's range shows a natural

downtrend across the utterance and the top and bottom lines tend to converge toward the end of it, a H tone that occurs late in the F0 contour could be as high in F0 as a L tone at the beginning of the contour.

Crucially, tones are represented on an autonomous tier or plane separated from the linguistic material with which they are necessarily co-produced: in line with Autosegmental Phonology, they are auto-segments, connected with units in the skeleton through specific association principles (Leben 1973; Goldsmith 1979)[2]. It is important to notice that the string of tones that represents a melody (or tune) is not intended to describe the whole F0 contour, but to represent only those parts of the melody that are linguistically significant: it is intended not as a mere transcription that describes all the peaks, troughs and turning points in a contour, but rather as an underspecified phonological representation. A direct consequence is that the same tune can be associated with texts of different segmental length and composition, giving rise to intonational contours that can be holistically different in acoustic form, but are the phonetic realization of the same abstract melodic entity.

Tones associate with the string of segments (or *text*) indirectly, through the mediation of the metrical structure of the utterance. With *metrical structure* we refer to a theoretical proposal which considers a given string of language to be organized into a series of hierarchically arranged prosodic constituents (in line with Prosodic Phonology), and that the linguistic units included in those constituents are specified in terms of relative prominence relations (in line with Metrical Phonology; Selkirk 1984, 2004; Nespor & Vogel 1986, 2007; Pierrehumbert & Beckman 1988; Liberman 1975; Libermann & Prince 1977).

In the literature, different theoretical proposals have been made as for the number of prosodic constituents that compose the hierarchy of prosodic domains. Nespor & Vogel (1986) for example propose seven constituents, which are, from the smallest to the larger: *Syllable*, *Foot*, *Prosodic Word*, *Clitic Group*, *Phonological Phrase*, *Intonational Phrase*, and *Utterance*. Others (e.g. Selkirk 1978, 1986), do not posit the existence of a *Clitic Group* but propose a *Minor phrase* and a *Major phrase* between the level

[2] In an autosegmental representation, different characteristics of a sound message - for example tones and phonemes - are represented on different tiers that all converge on a common plane, called a skeleton; this consists of a sequence of temporal units designed to fix the linear order of consonants and vowels (Leben 1973; Goldsmith 1979).

of the *Prosodic Word* and the *Intonation Phrase*. This is the highest level of the hierarchy, i.e., Selkirk does not always posit the *Utterance* as prosodic domain (but see Selkirk 1978), while she does consider the *Mora* as the lowest unit of the hierarchy. Beckman & Pierrehumbert (1988) consider three levels of constituents above the prosodic word: the *Accentual Phrase*, the *Intermediate Phrase* (ip), roughly corresponding to the phonological phrase of Nespor & Vogel (1986) and to the *Major Phrase* of Selkirk (1978, 1986), and the *Intonational Phrase* (IP). As it appears, there is agreement on the higher levels of the hierarchy but not as on the mid-levels (for a discussion of the different prosodic hierarchies and their correspondences: Shattuck-Huffnagel & Turk 1996; Frota 2017).

The constituents proposed in different models are motivated by the theoretical and empirical analyses of specific languages: for example, the proposal of an *Accentual Phrase* as a prosodic domain in Beckman & Pierrehumbert (1986) stems from their prosodic analysis of Japanese and has been shown as pertinent in the analysis of other languages as well, such as French (Verluyten 1982; Jun & Fougeron 1995). So far, most works adopting the AM framework to analyse the intonation of many languages (for an overview, Jun 2005; Frota & Prieto 2015) show that at least two levels of constituents are pertinent for intonation: a minor phrase, be it the *Intermediate*, *Phonological* or *Minor phrase*, and a major one, the *Intonational Phrase*.

The authors of the above proposals agree that all the constituents of a certain hierarchical level are exhaustively included in the constituents of the upper hierarchical level, a constrain known as the Strict Layer Hypothesis; in other words, that the prosodic structure is not recursive, differently from the syntactic structure[3], and that each constituent is endowed with a *head*, a metrical strong element.

Tones can associate with the constituents' heads or with constituents' edges. In the first case, in a language such as Italian, the Tone-Bearing Unit (TBU, the docking site of the tone) is the stressed syllable, which by virtue of this association gets its prominence enhanced. Tones associated with constituent heads are called *pitch accents* (PA) and are marked with a star, e.g. H* or L*. The last pitch accent in the Intermediate Phrase is the *nuclear*

[3] The debate on the recursive vs. non-recursive nature of prosodic constituents is very lively in works developing the original Prosodic Phonology proposal or referring to the prosodic hierarchy in general. For instance, Ladd (2008) opened to a limited recursivity in prosodic domains allowing for compound constituents in a given level of the hierarchy.

accent (Beckman 1996). Tones can also associate with the right boundary of the Intermediate and Intonational phrase: the former is marked by a hyphen and the second by a percent sign, e.g., respectively H- and H%, and are called *phrase accent* and *boundary tone.* Collectively, they are referred to as *edge tones* and their role is to demarcate a phrasal boundary. All languages investigated so far have H and L tones that associate with right boundaries. For several languages it has been postulated also a left edge association for the Intonational Phrase, mostly of a H tone, that in such a case is indicated as %H (e.g. for English: Beckman *et al*. 2005; for Italian: Avesani 1995; Grice *et al*. 2005; Gili Fivela *et al*. 2015). Edge tones associate with the edges of constituents and are phonetically realized on the segments flanking their edges, such as the final vowels or sonorant consonants for H% or L% and phrase initial ones for %H – see Figure 2.

Pitch accents (henceforth, PAs) can be monotonal, i.e. composed by one tone only, as in H* or L*, or bitonal, as L*+H or H*+L, whose phonetic realization gives rise to glissandos, namely rises and falls. The star notation reflects the fact that the starred tone is stronger and is directly associated with the TBU. The weaker tones are called *leading* if they precede the starred tone, *trailing* if they follow it. For a discussion on the nature of the starred tone and for the internal structure of pitch accents the reader is referred to

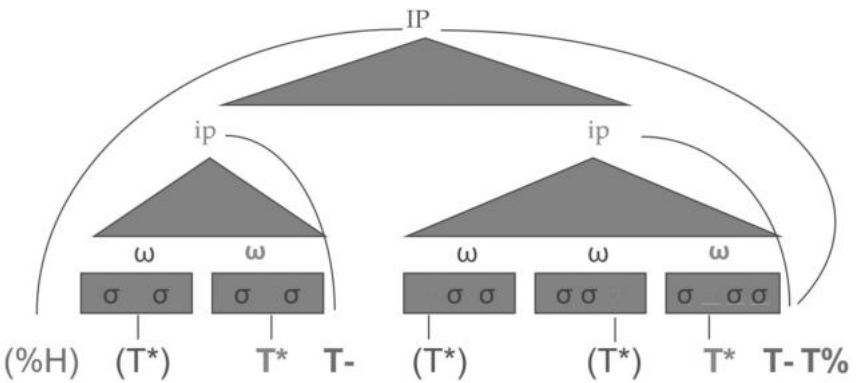

Fig. 2. Example of prosodic tree, including an Intonational Phrase (IP), Intermediate Phrases (ip), prosodic words (ω) and syllables (σ); pitch accents (T, representing both monotonal and bitonal pitch accents) are shown both in prenuclear (bracketed) and nuclear position and are followed by edge tones, either phrase accents (T-) or boundary tones (T%)*

Arvaniti *et al*. (2000) and Grice (1995). Edge tones are monotonal, but in the literature multitonal combination of edge tones have been occasionally proposed in the analysis of specific languages, for example a tritonal LHL% boundary tone has been used for the analysis of Catalan (Prieto 2014).

Since the beginning of the AM approach, a crucial role in the definition of tonal categories and in the coding of the intonative oppositions within the intonational system of a specific language has been recognized to tonal *alignment* and *scaling*. The former corresponds to the synchronization of the F0 peaks and lows to the TBU, and the latter represents the tone height of the H or L tone associated with a structural position (see also § 2.1).

Based on the formal properties of alignment and scaling of PAs and edge tones and on the informational and pragmatic functions they convey, a limited set of contrastive tonal events can be identified as the building blocks of the intonational system of a specific language. Tunes then arise from the phonetic implementation of a linear sequence of pitch accents and edge tones whose targets are assumed to be linearly interpolated. Typically, the melody of an intermediate phrase (ip) is composed by one or more optional prenuclear PAs, one obligatory nuclear PA, and one edge tone. If two ips combine to make up an intonational phrase (IP), then its melodic structure is represented as follows:

(1) $[[(\text{prenuclear PA}) - \text{nuclear PA}]_{ip}\ [(\text{prenuclear PA}) - \text{nuclear PA}]_{ip}]_{IP}$

where at the right boundary of each ip is associated a phrase accent and at the right boundary of the IP is associated a boundary tone. Consequently, the end of the melody is marked by a combination of two edge tones – see Figure 2.

For Mainstream American English (MAE), a consensus analysis has identified two boundary tones and two phrase accents (L%, H%, L-, H-), two monotonal PAs (L*, H*) two bitonal rising PAs, L+H* and L*+H, and one falling PA H*+!H (Beckman *et al*. 2005).

As the inventory of contrastive tonal events is rigorously linguo-specific, not all languages or language varieties share the same tonal inventory. For example, in MAE only one falling PA is deemed to be part of the intonational inventory, while in all the varieties of Italian analyzed so far a H*+!H is not attested, but all of them share a falling H+L* that occurs as the nuclear pitch accent in broad focus declarative sentences. For some of them also a H*+L falling accent is attested (Gili Fivela *et al*. 2015).

Summarizing, in the Autosegmental Metrical framework the two basic tasks of prosody, phrasing and highlighting, are fulfilled respectively by the *placement* of edge tones at the boundaries of prosodic constituents and by the placement of pitch accents within each constituent. The pragmatic meaning of the tune is determined by the linear position and by the selection of the *type* of PAs and edge tones (Pierrehumbert & Hirschberg 1990).

Going back to the example in §1, the similarities and the differences of the two contours represented in Figure 1 can be disentangled by considering how PAs and edge tones are distributed in the contours. In the short utterance shown in Figure 3, the stressed syllable of *Marina* carries a L+H* PA, where H* is associated with the TBU /ri/ and L is the trailing tone. This PA is the only nuclear PA occurring in the contour, as the short utterance is phrased in one intonational phrase only. The PA is then followed by a L- phrase accent and a H% boundary tone realized on the following unstressed syllable[4].

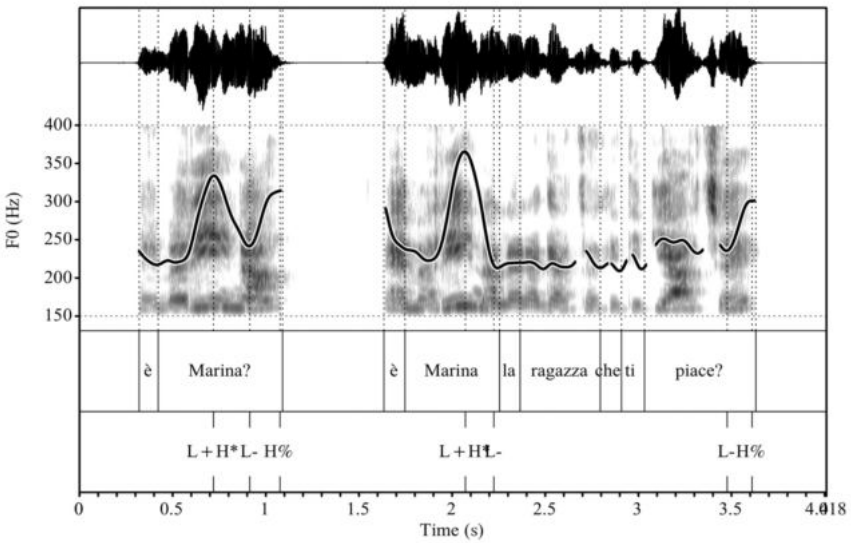

Fig. 3. Intonational contour and tonal transcription of the sentences È Marina? *'is she Marina?' and* È Marina la ragazza che ti piace? *' is it Marina the girl that you like?'*

[4] By default in the first formulations of the AM theory an IP is composed by one ip. Therefore, even a monorematic utterance is right marked by two edge tones.

The long utterance *È Marina la ragazza che ti piace?* 'Is it Marina the girl that you like?' is phrased in two intermediate phrases:

(2) $[[\text{È Marina}]_{ip}\ [\text{la ragazza che ti piace}]_{ip}]_{IP}$

Here *Marina* shares the same PA (L+H*) and the same metrical position (nuclear) of the short utterance and it is followed by the same L- phrase accent that delimits the right boundary of the first intermediate phrase in which it is wrapped. A L- phrase accent delimits the right boundary of the second intermediate phrase and a H% boundary tone marks the right boundary of the intonational phrase[5].

2.2 Phonological representation and phonetic form

As already mentioned, the phonological representation is necessarily language-specific. Notably, such phonological representation is mapped into a phonetic representation through phonetic realization rules, which are again language-specific, and which shape the F0 track in terms of its alignment and scaling characteristics.

The literature on alignment and scaling properties of tones is rich and it offers various points of view on the mapping between phonology and phonetics. For instance, alignment was originally interpreted as somehow defined by phonological association (the association of a tone to a TBU implied its alignment with it in Pierrehumbert's 1980 proposal), while later works did not assume such a strict coordination. The starred tone can also be aligned earlier or later than the tone-bearing unit, pointing to the impossibility to rely on alignment characteristics to identify the starred, associated tone (Arvaniti *et al*. 2000) or a coordination with specific landmarks may be assumed (cfr. Ladd *et al*.'s 1999 Segmental Anchoring Hypothesis). Further, such phonetic features, which were originally discussed with reference to acoustics, have later been investigated as far as their perception is concerned (D'Imperio & House 1997; D'Imperio 2000), and some works also focused on kinematic data (e.g., D'Imperio 2002). However, more recently, some authors argued that alignment and scaling of tonal targets may not be the only properties speakers take care of, as inter-subject differences may be observed in the accuracy speakers

[5] A postfocal compressed L+H* PA occurs on the TBU of the last word *piace* ('you like').

show in aligning targets or preserving pattern shape (Niebuhr 2007; D'Imperio *et al*. 2010).

Further, many works in the literature have basically shown that intra-linguistic alignment differences, e.g., between PAs which are part of a specific linguistic system, convey semantic and pragmatic differences that are (almost) categorically perceived (though results are not consistent in this respect; see Gili Fivela 2012 for an overview). Scaling has long been considered less relevant than alignment in linguistically differentiating linguistic functions, though it is nowadays accepted that scaling too conveys semantic and pragmatic differences and may be categorically perceived (e.g., yes-no questions as opposed to *wh*-questions in Majorcan Catalan: Vanrell 2006, 2007).

One of the key aspects of the AM proposal regards the lack of a "transparent" and direct mapping of phonetic properties onto a phonological representation. Such representation is abstract. However, its units are labelled taking somehow the phonetic properties into account (e.g. a label will include a H+L tone if, in the clearest realization, it is falling). Thus, it is well known that identifying high and low turning points in the phonetic form is not enough to identify phonological targets and to label them. Crucially, a phonetic event may be considered as phonological if its presence/absence implies changes in the linguistic function played by the pattern. Thus, the questions to be answered in analyzing a phonetic continuum are, for instance: Is there a linguistic function played by such F0 event? What is the impact of changing its alignment and scaling? Does the meaning of the utterance change?

Depending on the linguistic system, the inventory of intonational units and their alignment and scaling characteristics changes. Thus, differences are found in systems of different languages and even in the case of varieties of the same language. For instance, a set of nine pitch accents and six edge tones is necessary to analyze 13 Italian varieties (Gili Fivela *et al*. 2015), and each variety shows a specific selection and a specific combination of those units. For instance, the L*+H pitch accent is found in Neapolitan and Turin Italian, but it is not found in Pisa Italian; a HL% edge tone is found in Pisa Italian after a H+L* pitch accent, while it is found after L+H* in Neapolitan (for updates on Italian in this line of research, see Gili Fivela & Iraci 2017; Gili Fivela & Nicora 2018; see the latter, together with Gili Fivela *et al.* submitted, for investigations concerning possible cross-varietal similarieties due to contact situations).

2.3 ToBI

The principles of AM theory are reflected in the ToBI (Tone and Break Index) transcription system (Beckman & Ayers 1997), a common transcription system whose immediate benefit is the possibility to compare the prosody of disparate languages and language varieties. Since 1991, date of the first workshop organised to define a set of common principles for transcribing Mainstream American English, ToBI-like analyses have been proposed for a number of different linguistic systems providing the intonational analysis of 35 languages and almost 30 language varieties (we refer to the following collective volumes: Jun 2005, 2014; Frota & Prieto 2015). In fact, the original ToBI system has been adapted for the description of languages which vary geographically (European, Native American, Asian, Australian aboriginal languages) and typologically, in the type and in the degree of lexical specification of prosody (intonational languages such as English, Italian, French, Spanish, Portuguese; lexical pitch accent languages such as Swedish, some Dutch and German dialects, Chickasaw, Japanese; and tone languages such as Cantonese and Mandarin).

What is important to highlight is that ToBI is not comparable to an International Phonetic Alphabet for intonation, with the choice of adopting a broader or narrower transcription, but it is a phonological representation of intonational contrasts[6]. The aim is not a more or less faithful depiction of the F0 contour; rather, its aim is to define a limited set of categories to represent the intonational contrasts in a sound system. Therefore, the transcription should be driven by system internal considerations, by considering a phonetic detail as part of the representations only if there is evidence it is contrastive (Arvaniti 2016: 8). For example, the decision to transcribe a pitch rise on a stressed syllable as a H* or a L+H* PA must be guided by considering the contrastiveness of such event within the whole system under analysis. The transcriber could then decide to include the rise as part of the phonological specification in one language, with the consequent adoption of the label L+H* because L+H* contrasts with H*,

[6] For a thorough discussion of the phonological assumptions behind current approaches to prosodic transcription, for the choice of discrete units and their granularity and the consequences of considering ToBI as a broad phonetic transcription we refer to the special collection on Advancing Prosodic Transcription that appeared in *Laboratory Phonology* (D'Imperio *et al*. 2016).

but to exclude it from the phonological representation of another language, where such a contrast is not attested (e.g. Arvaniti *et al*. 2000) Importantly, a decision about contrastiveness must be guided by form in combination with meaning: differences in the form of a tonal event should be considered contrastive only after taking into account focus, information structure, and the pragmatic function of utterances in discourse (cf. Pierrehumbert 1980: 59-63).

As a good practice suggested by Arvaniti (2016: 8), the analysis should involve several iterations that lead to both bottom-up and top-down decisions: a first set of data determines the original analysis, which is then used to annotate more data.

3. *On metrical structure*

As mentioned in section 2, Intonational Phonology also refers to the existence of prosodic domains and, consistently, the AM framework refers to the principles of Intonational Phonology as well as Metrical and Prosodic Phonology. As a matter of fact, different proposals regarding prosodic domains emerge from independent research traditions and they differ according to theories of the syntax-phonology mapping and theories of the structural relations between constituents of prosodic structure (Frota & Vigario 2018).

In one framework, the structure of phonological representation at the word level and above is a hierarchy of phonological constituents that results from the interaction of a limited set of (morpho)syntactic information with phonological principles related, among others, to constituent size and weight. The prosodic word (PW), the phonological phrase (PhP), and the intonational phrase (IP) are the domains of application of segmental rules and bear a relation to a specific syntactic constituent type: respectively a word-like (lexical) morphosyntactic unit, a phrase-like syntactic unit, and a clause-like syntactic unit[7] (Nespor & Vogel 1986; Selkirk 2011; Truckenbrodt 1995; for a thorough , on syntax-

[7] Two main branches of such framework correspond to relation-based and end-based approaches, which differ as for the syntactic information used in the computation of prosody: the former makes reference to notions like head-complement, modifier-head relations, and syntactic branching, while the latter refers to syntactic heads and maximal projections.

prosody interface we refer to Frota & Vigario 2018, and for an analysis of syntax-prosody interface in Italian we refer to Bocci 2013).

Parallel to theories positing that prosodic structure is rule-based and related to syntax, a different approach posits that prosodic structure is intonation and prominence defined, by relying on intonational, durational and segmental phenomena that characterize the constituents above the word level (Beckman & Pierrehumbert 1986; Pierrehumbert & Beckman 1988; Beckman 1996).

There is ample empirical evidence that a constellation of cues mark prosodic domains, functional to supporting the prosodic structure and the constituents it comprises. Constituents are marked by lengthened duration of the segments right-flanking the boundary (*final lengthening*), with degrees of lengthening that correlate with the prosodic boundary level (a.o. Beckman & Edwards 1994; Wightman *et al*. 1992; Byrd & Saltzman 2003); and by lengthening of earlier segments within the preboundary word (Price *et al*. 1991; Wightman *et al*. 1992; Turk & Shattuck-Hufnagel 2007). Segments in initial position of a prosodic domain have their articulatory properties enhanced as a function of the constituent level in the prosodic hierarchy (ip or IP) in which they appear (*prosodic strenghthening* of e.g. linguo-palatal contact or nasal flow e.g., Keating *et al*. 2004); pre-boundary as well as post-boundary segments can be glottalized (Pierrehumbert & Talkin 1992; Byrd & Saltzman 2003). Tonal marking of prosodic domains include the presence of edge tones, the scaling of subsequent H peaks within a domain and the total or partial resetting of pitch range after a boundary (Pierrehumbert 1980; Beckman & Pierrehumbert 1986; Truckenbrodt 2002; D'Imperio & Michelas 2014).

Interestingly, the presence of a pause is only one - and notably not even crucial - cue of a prosodic boundary. In a study on brain responses to prosodic boundaries, Steinhauer & Friederici (1999) found that locally ambiguous sentences that contain the same words but differ in the presence vs. absence of a prosodic boundary elicit a neural response at the position of the boundary that is marked by pause, final lengthening and edge tone. When the prosodic boundary is perceived and used by the listener to drive the syntactic parsing of the sentence, event-related brain potentials (ERPs) show a waveform positive shift in the temporal interval that corresponds to the boundary. This new ERP component was termed Closure Positive Shift (CPS) since it took the form of a positive shift at the closure of an intonational phrase. In later studies not only was it confirmed that the CPS is a neural response to the

prosodic boundary as a whole, but that by removing the pause from the boundary while maintaining the other cues (final lengthening and edge tone) the ERP component was still observed (Bögels *et al*. 2011).

The results of the abovementioned neural studies are in agreement with models of online and offline sentence processing which argue for the constituents of prosodic structure acting as processing units in human sentence comprehension (Carroll & Slowiaczek 1987), that prosodic information contributes to the final structuring of an initial syntactically determined parse (Pynte & Prieur 1996), and that prosodic and non-prosodic factors may enter a cue-trading relation in the process by which syntactic and semantic analyses are constructed (e.g. Beach 1991; Stirling & Wales 1996).

The importance of prosodic constituency for the comprehension of syntactic structure is also shown in a study on Italian (Bocci & Avesani 2015), which builds on the result of a previous production experiment (Bocci & Avesani 2011). In languages such as English or Italian, the default distribution of phonological prominences assigns the head to the rightmost element in a prosodic domain (e.g. Nespor & Vogel 1986). In broad focus sentences or in out-of-the-blue sentences produced without context the highest level of prominence is then assigned to the last head of the last intermediate phrase in the (last) intonational phrase. Therefore, the last word in the higher-level prosodic constituent gets the main prominence of an utterance and attracts the nuclear pitch accent. The same distribution of prominences occurs in sentences where the last lexical constituent is the (narrow) *focus*. In such pragmatic condition, the last word is at the same time the one which is most important informationally and the one that attracts the strongest metrical prominence. Phonological and pragmatic conditions concur in marking the last item as the most prominent. If the pragmatic conditions vary and the focus occurs sentence-initially, the focal element attracts the main prominence and the rest of the sentence has the informational status of *background* information which is prosodically subordinated. In English it is said that postfocal material must be de-stressed and de-accented (Selkirk 2008, a.o) with no phrase-level metrical prominence. Usually, the F0 contour is low and flat after the initial PA that is associated with the focus element and no other PAs follow. On the contrary, in some southern varieties of Italian and in Portuguese, post-focal constituents can be pitch accented, with the proviso that the associated PAs are not fully-fledged: only some types of PA can occur post-focally and their pitch span is highly compressed (e.g. Frota 2000; Grice *et al*. 2005).

Bocci & Avesani (2011) show that in Tuscan Italian post-focal material can be accented, even if the post-focal portion of the F0 contour is low and flat and shows no sign of pitch rises or falls. Their conclusion is based on sentences like (4), where a sentence-initial focussed subject "Germanico" is followed by the verb "vorrebbe invitare" and a right-dislocated object "Pierangela" represented in Fig. 4. In such sentences the focused subject is always associated with a rising PA and the following contour is low and flat.

Their production results show that the tone-bearing unit [ˈta] of the post-focal infinitive verb is longer, has higher spectral emphasis and more extreme formant trajectories than the same verb in a broad focus sentence (3). Moreover, that the last vowel and last syllable of the verb [re] are longer than in the equivalent sentence in broad focus (3). All the acoustic cues indexing phrasing and prominence indicate that even in absence of a "visible" PA, the verb acts as the metrical head of the independent ip "vorrebbe invitare", which is inserted between the ip that includes the focused subject and the ip that includes the right-dislocated object (5):

(3) [Germanico vorrebbe *invitare* Pierangela]$_{BF}$
'Germanico would like to invite Pierangela'

(4) [Germanico]$_{F}$ la vorrebbe *invitare* [Pierangela]$_{RD}$
'Germanico her-would like to invite Pierangela'

(5) [[Germanico]$_{ip}$ [la vorrebbe *invitare*]$_{ip}$]$_{IP}$ [[Pierangela]$_{ip}$]$_{IP}$

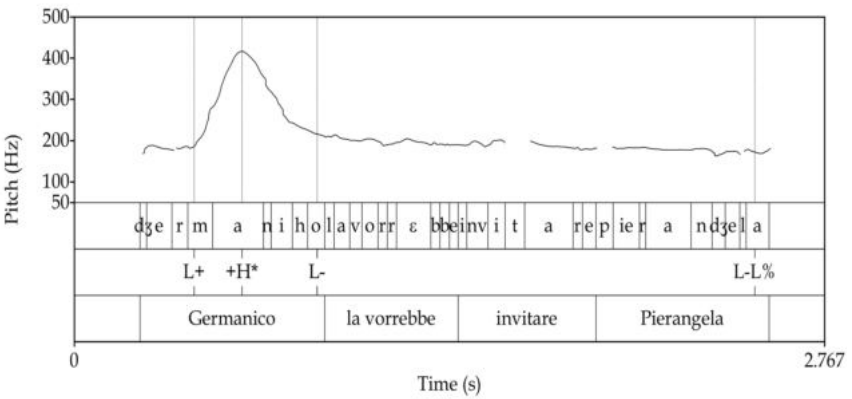

Fig. 4. Intonational contour of the sentence Germanico la vorrebbe invitare Pierangela *'Germanico her-would like to invite Pierangela':* Germanico *is the focus,* la vorrebbe invitare Pierangela *is the background where* Pierangela *is the right-dislocated object*

Building on such results, a comprehension experiment with manipulated stimuli was run in order to determine whether the conclusion of the production experiment had a psychological validity. The rationale of the comprehension experiment is based on two morphosyntactic properties of Italian. First, a clitic cannot double a focus element. Second, Right Dislocated (RDed) objects always involve a resumptive clitic, whereas subjects do not. Starting from sentences like (5) they reasoned that if the sentence is manipulated by deleting the object clitic "la" from the segmental string, in the resulting sentence the phonetic properties of the infinitive's ip-head and of the IP-boundary at its right edge still cue the final proper name "Pierangela" as right dislocated. However, because there is no object clitic, Pierangela cannot be interpreted as a RDed object and the first proper name "Germanico" in focus could be interpreted either as a focused subject or as a fronted focused object. Given the morphosyntactic and prosodic properties of the sentence, they expected the sentence to be interpreted in comprehension as OVS, with "Germanico" being interpreted as a fronted focused object and "Pierangela" being interpreted as a RDed subject. In a second run, they further manipulated the previous sentence with the excised clitic by deleting the phonetic correlates of the ip-head on the infinitive and of the IP-boundary at its right edge. Because no prosodic cue marks "Pierangela" as right dislocated any longer, a SVO order should be restored. Manipulations regarded only segments and specifically the cues of the TBUs [ta] and [re]. The results confirmed that when the infinitive is characterized by durations that correlate with the ip-head and the IP-boundary, the preferred interpretation is OVS. When head and boundary do not occur, the preferred interpretation is SVO. Overall, the results clearly indicate the fundamental role played by metrical structure (constituents and their heads) in sentence comprehension: only small duration differences in relevant positions lead to a specific metrical representation and this, in turn, leads to a specific syntactic representation.

4. *Linguistic functions of prosody and intonation*

As already mentioned in §1, prosody and intonation play a wide range of functions in communication. Some of them are clearly linguistic, such as signalling changes in sentence modality, phrasing, accentuation, and focus (Kohler 2006).

Yes/no questions are often signalled by means of prosodic and intonational changes in comparison to statements, though in some languages morphosyntactic markers are also found (Ladd 1996). Italian, on the other hand, offers a very clear example of the linguistic use of prosody and intonation, as it may indeed express the change from a statement to a yes/no question with no morpho-syntactic means, but rather intonation resources. Further, varieties of Italian may use different patterns to signal yes/no questions, as in Fig. 5 (Grice *et al*. 2005; Savino 2012; Gili Fivela *et al*. 2015).

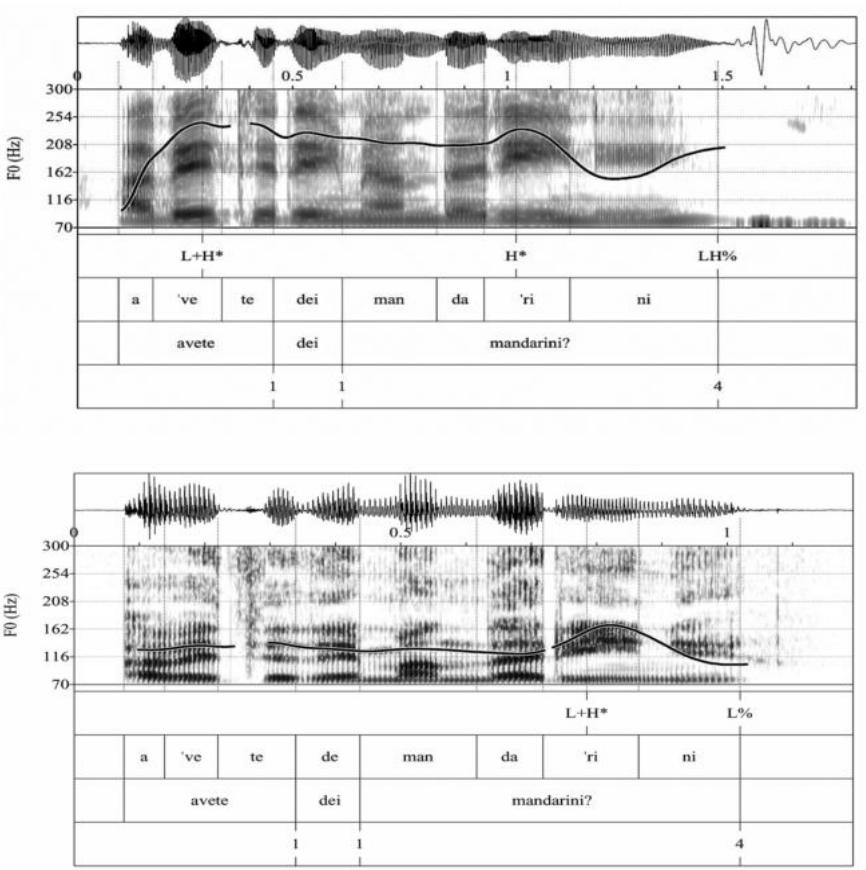

Fig. 5. Information-seeking question Avete dei mandarini? *'do you have mandarins?', speakers from Florence (top) and Salerno (bottom). From Gili Fivela* et al. *(2015)*

Another very important linguistic function, which makes the role of intonation and other prosodic features crucial for the expression of pragmatic meaning, is to highlight the *packaging* of information conveyed by an utterance (Chafe 1976; Krifka 2007). Speakers can choose to structure the information of an utterance in *topic-comment* or *focus-background* or to assign a referent the status of *given* or *new* information. Those choices are reflected in how the utterance is prosodically phrased and how post-lexical prominences are distributed in the utterance through the placement of pitch accents.

Phrasing information is crucial to highlight the information structure of the utterance as well as to solve syntactic ambiguities. Separate intermediate phrases, for instance, may separate the part of utterance that expresses what the speakers are talking about (the sentence *topic*) from the part that is related to what the speaker predicates about the topic (the sentence *comment*).

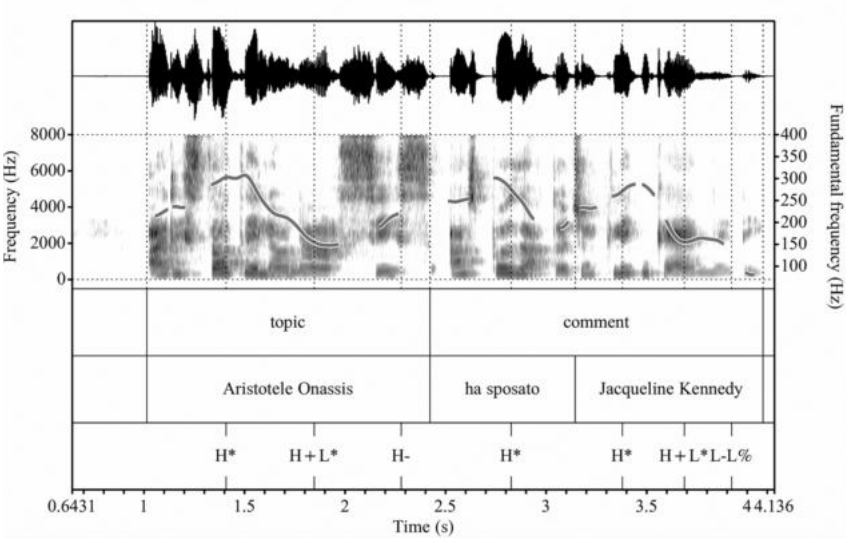

Fig. 6. Intonational contour of the sentence Aristotele Onassis ha sposato Jacqueline Kennedy *'Aristotele Onassis married Jacqueline Kennedy'. The boundary between the two ips that include the topic* (Aristotele*) and the comment* (ha sposato Jacqueline Kennedy*) are marked by a pause, final lengthening in the first ip, resetting of the pitch range and peaks downtrend in the second ip*

As an example, in (6) the *topic* "Aristotele Onassis" and the *comment* "ha sposato Jacqueline Kennedy" are coextensive with two intermediate phrases separated by a boundary which is signalled via a cluster of prosodic cues: the presence of a H- edge tone, a short pause, lengthening of the unstressed syllable before the boundary, resetting of the pitch range after the boundary and H targets downtrend within the *comment* (Fig. 6):

(6) [Aristotele Onassis]$_{\text{Topic}}$ [ha sposato Jacqueline Kennedy]$_{\text{Comment}}$
'Aristotele Onassis married Jacqueline Kennedy'

Phrasing signals also the disambiguation of a constituent's syntactic attachment, particularly of prepositional phrases, relative clauses, and adverbial phrases (a.o. Schafer 1997; Kjelgaard & Speer 1999; Avesani 1999; Hirschberg & Avesani 2000). For instance, in (7a) the absence of a prosodic boundary between "parlato" and "chiaramente" favours the low attachment of the Adverbial Phrase (AdvP) to the Verbal Phrase (VP), while in (7b) the presence of a prosodic boundary favours the high attachment of the Adverbial Phrase to the sentence root. In (7b), the high attachment of AdvP is favoured also by the prominence relation between the nuclear pitch accents of the ips that wrap respectively the VP and the AdvP: the lower height of the nuclear PA on the adverb indexes a prominence subordination of the AdvP with respect to the VP.

(7a) [[Lui] $_{\text{ip}}$ [le aveva parlato chiaramente]$_{\text{ip}}$]$_{\text{IP}}$
'He to-her talked clearly'
(7b) [[Lui] $_{\text{ip}}$ [le aveva parlato] $_{\text{ip}}$ [chiaramente] $_{\text{ip}}$] $_{\text{IP}}$
'It was clear that he talked to her'

The placement of pitch accents within an utterance serves the function to indicate which words or phrases are most salient to the purpose of the discourse, a function that directly relates to the notion of *focus* of information and to the *information status* of referents in the discourse (*given* and *new* information). Focus is a semantic-pragmatic notion (Krifka 2007). A pragmatic use of focus is to highlight the part of an utterance which the speaker presents as being important or assumes to be highly informative for the listener. If the focus is restricted on a constituent as opposed to the whole sentence, the sentence is partitioned in *focus* (the informative part) and *background* (the uninformative part). Focus is usually

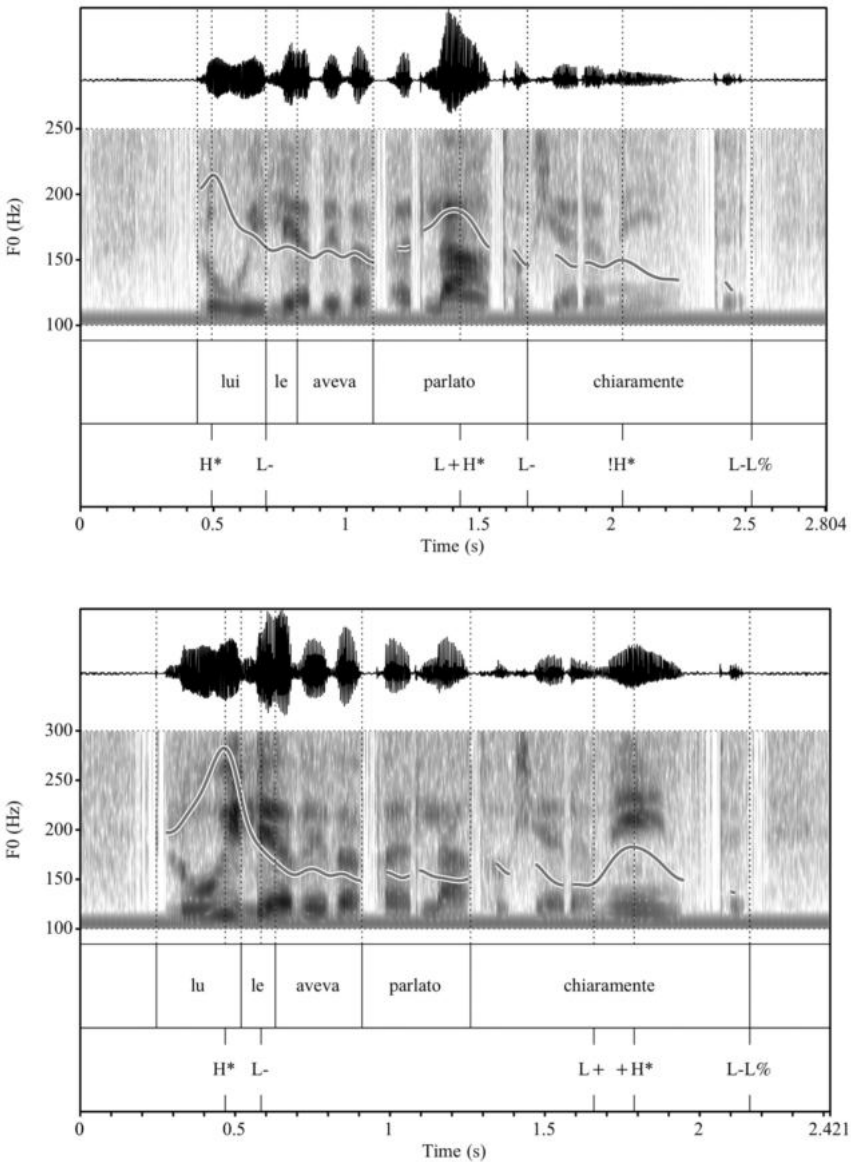

Fig. 7. Intonational contour of the sentence Lui le aveva parlato chiaramente *'he talked to her clearly' with a high attachment (top) and a low attachment (bottom) of the adverb* chiaramente. *In the high attachment the adverb is separated from the rest of the utterance by a prosodic boundary*

determined by taking into consideration the immediately preceding context. That is, according to a classic definition, the focus corresponds to the *wh*-part of a constituent question (Paul 1880, quoted in Krifka 2007) and is defined in terms of Question-Answer Congruence (a.o., Büring 2016). In the exchange in (8), the focused part "Michelangelo" is the answer to the question "a chi hanno presentato Marinella le tue sorelle? (to whom your systers presented Marinella?)".

(8) Q: A chi hanno presentato Marinella le tue sorelle
'to whom your systers presented Marinella?'
A: [Le mie sorelle hanno presentato Marinella]$_{background}$
[a Michelangelo]$_{Focus}$
'my systers presented Marinella to Michelangelo'

According to a more recent formulation, focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions (Rooth 1992). Focus evokes a set of alternative propositions that differ only for the focused element: in the case of (8) the focus on Michelangelo evokes a set of alternative propositions (indicated in braces) such as {le mie sorelle hanno presentato Marinella *a Giacomo* ("my sisters presented Marinella *to Giacomo"*); le mie sorelle hanno presentato Marinella *a Luca* ("my sisters presented Marinella *to Luca"*)*;*….} and from this set the proposition "le mie sorelle hanno presentato Marinella *a Michelangelo* ("my sisters presented Marinella *to Michelangelo"*)" has been selected. The background is defined as the invariant part of the alternative propositions.

The answer in (8A) is a case of *narrow focus,* and in languages like Italian it is marked by a (nuclear) pitch accent on the focus constituent[8]. A special type of focus is *contrastive focus* (CF) that can also be used as a correction of what has been previously said. In Figure 8, (8A) has been produced by a Sienese speaker with a focus of contrastive-corrective import on "a Michelangelo". In the F0 contour, a nuclear L+H* pitch

[8] When the focus is not restricted to a single constituent it is *broad*. In these structures the relation between focus and accent is no longer straightforward and a pitch accent on one word, called the focus exponent, marks the larger focus domain (the phenomenon is called *focus projection*). In Italian the focus exponent in broad focus structures is the last word of the sentence.

accent is associated with the TBU of the focused constituent ("CFocus") and two prenuclear pitch accents are associated with two noun phrases in the background ("background (given)"). If the focus phrase is moved sentence initially as in (9), the element in focus is marked by same L+H* nuclear pitch accent but the prosodic properties of the background are radically different, as no fully-fledged pitch accent occurs post-focally (Fig. 8).

(9) [a Michelangelo]$_{\text{Focus}}$ [le mie sorelle hanno presentato Marinella]$_{\text{background}}$
'to Michelangelo my systers introduced Marinella'

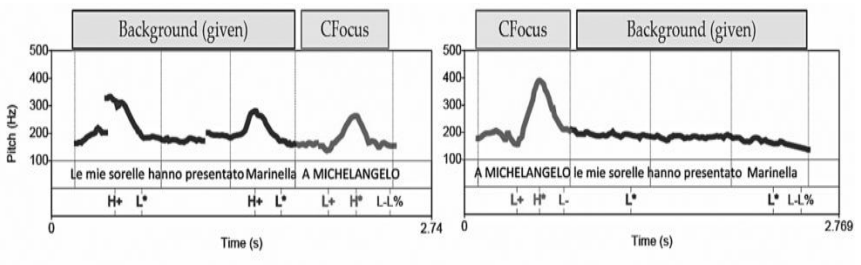

Fig. 8. Intonational contour of the sentence Le mie sorelle hanno presentato Marinella a Michelangelo *'my sisters introduced Marinella to Michelangelo' (left) and* A Michelangelo le mie sorelle hanno presentato Marinella *'to Michelangelo my sisters introduced Marinella' (right). Michelangelo is a contrastive-corrective focus in sentence-final position (left) and in sentence initial position (right). (Courtesy of Giuliano Bocci)*

Besides a pragmatic, there is also a semantic use of *focus*, that leads to change the truth-conditional value of a proposition. This is the case of focus-sensitive operators such as the particles "only", "even" or the negative quantifier "not", in which the linguistic element modified by the logical operator is marked as focused by the association with a pitch accent (for examples in Italian and English see Hirschberg & Avesani 2000).

Further, in Italian different types of focus are distinguished by different types of pitch accents: *information focus* (IF) and *contrastive-corrective focus* (CF) are marked by different types of pitch accents in many varieties

of Italian (Avesani 2003; Avesani & Vayra 2004; Bocci & Avesani 2008; Gili Fivela *et al*. 2015). That is, differently from English in which the same type of pitch accent is used for the two imports of *focus* in the same sentence position, in Italian no ambiguity arises in identifying IF and CF in sentence final position. In Florentine and Sienese Italian, for instance, IF associates with a falling accent, phonologically specified as H+L*, while CF associates with a rising accent, phonologically specified as L+H* (Fig. 9).

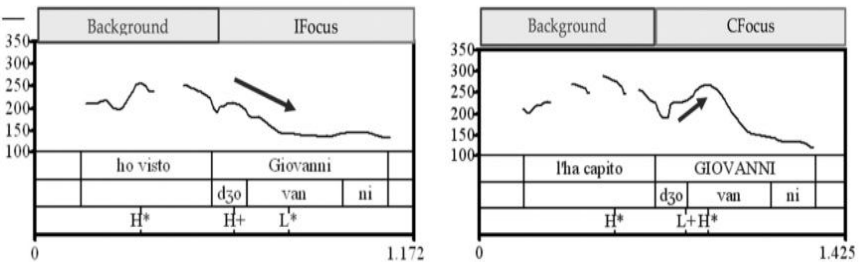

Fig. 9. Information focus (left) in the sentence Ho visto Giovanni 'I have seen John' (as a reply to the question Chi hai visto? *'Who did you see?') and contrastive-corrective focus in the sentence* L'ha capito Giovanni *'it-understood John' (as a reply to the question* Sembra che lo abbia capito Leo *'it looks like it-understood Leo').*

Intonation also encodes discourse-related properties such as the *information status* of referential expressions. With information status of a referent we indicate the specific relation between a linguistic entity and the corresponding non-linguistic entity that holds in the mind of the speaker/hearer or in the discourse model at the moment of communication, which can dynamically change as the discourse evolves (e.g. Chafe 1976). A linguistic expression can be *given* if its representation is already present in the mind of the speaker/hearer because: a) it is part of the encyclopedic knowledge or b) it is part of the shared knowledge of the speaker and the hearer; c) it is visible in the external context, d) it is explicitly present in the immediate linguistic context (i.e. already mentioned), or e) it stands in a hyperonymy relation with its antecedent (Baumann & Riester 2013). The status of *given* is uniquely determined by the knowledge and attention

state of the interlocutor at a specific moment of a conversation (Prince 1981; Lambrecht 1994; Grosz & Sidner 1986). *New* is a referent recently entered in a discourse and not recoverable from the preceding context.

Many studies in the recent and past literature have shown that there is a correlation between how prosody marks a linguistic entity and its information status: a *new* linguistic entity is usually pitch accented and a *given* entity is usually *deaccented*, that is, it lacks a pitch accent that would otherwise be used to mark the same elements if it was occurring in all-new utterances. Listeners are sensitive to the prosodic marking of an entity's information status: accenting of new information and deaccenting of *given* information affects off-line sentence comprehension (Birch & Clifton 1995) as well as on-line processing (e.g. Dahan *et al*. 2002). Moreover, neurolinguistic studies using event-related potentials and investigating the impact of different types of accentuation on the comprehension of referents have shown that appropriate prosodic cues affect the construction of a mental model (e.g. Schumacher & Baumann 2011).

However, not all languages follow the same pattern of accentuation/deaccentuation of a referent according to its status of *new* or *given* information. First of all, the association between deaccenting and information status is not to be conceived as an exceptionless one-to-one relationship, but at most as a strong association (Brown 1983; Terken & Hirschberg 1994; Bard & Aylett 1999), as many intonation patterns that are claimed to convey a certain meaning only represent the most frequent pattern that speakers choose to use in that context (Braun & Chen 2012). Second, *given* can be seen as a scalar notion in which, based on a scale of assumed familiarity, at least three categories are defined: *new*, *given* and *accessible* information (Chafe 1976; Lambrecht 1994; Prince 1981, 1992). Along this line, it was shown that in German *given* and *textually accessible* information are preferably deaccented (respectively: 78% and 63%) while *inferentially accessible* information is preferably accented (64%) with a H* or a downstepped !H* pitch accent (Röhr & Bauman 2010). Even with those provisos, though, it is widely accepted that Germanic languages avoid marking as prosodically prominent referential expressions that strictly convey *given* information.

Contrary to Germanic languages, Romance languages fail to deaccent referents which are informationally *given*. For Italian this was firstly observed by Cruttenden (1993) and Ladd (1996) and later experimentally proved by Avesani (1997), Avesani & Vayra (2005), Swerts et al. (2002).

Pitch accenting *given* referents is reported to occur in different speaking styles such as spontaneous speech (Avesani 1997) or task-directed dialogues (Avesani & Vayra 2005). In the latter study, only 6.5% of coreferential expressions were reported as lacking a pitch accent. Further, in comparing accentuation strategies of typologically different languages such as Dutch and Italian, Swerts *et al*. (2002) showed that Italian speakers always accent *given* items while Dutch speakers always deaccent them; moreover, Italian speakers cannot perceive any difference in prominence between *given* and *new* items (while Dutch listeners can) and they are unable to reconstruct the dialogue history on the basis of the accentuation of an item, while Dutch listeners are able to guess whether a referent was already mentioned in the preceding dialogue.

Avesani & Vayra (2005), however, observed that some cases of coreferential nouns, albeit few, were produced with a pitch accent. All cases related to *given* referents which occurred in longer syntactic constituents, specifically in post-focal position of sentences with fronted foci. Bocci & Avesani (2011) and Bocci (2013) disentangled the question arguing that *given* constituents which occur post-focally are not deaccented, but are assigned phrasal stress, overriding their information status of *given* and part of the background. They are marked by all prosodic cues that identify them as post-lexically stressed, and by a pitch accent that in Tuscan Italian is a L*. By taking into account only the melodic contour though, it could be said that post-focal *given* elements are "deaccented", as superficially no fully-fledged pitch accent (high, rising or falling) is observed. But a more thorough prosodic analysis and distributional considerations argue for the contrary: as can be appreciated in Fig. 8, changing the focus-background partition of the sentence, the same *given* elements that in post-focal position appear as deprived by a fully-fledged pitch accent in pre-focal position clearly bear a H+L*.

5. *The meaning of tunes*

In a seminal paper Pierrehumbert & Hirschberg (1990) address the contribution of the choice of *tune*, or intonational contour, to discourse interpretation. While in the literature the characterization of the meaning of a given tune has been interpreted in terms of speaker attitudes (politeness, surprise, deference etc.), speech acts (statements, requests, contradictions), propositional attitude (belief, uncertainty, etc.), presupposition and focus,

Pierrehumbert & Hirschberg (1990) claimed that neither speech acts nor propositional attitude provided sufficient characterization of available tunes in English. Rather, they claimed that tunes specify a particular relationship between the propositional content of the utterance and the mutual beliefs of discourse participants: speakers choose a specific tune to convey a particular relationship between an utterance, the current beliefs of the hearer(s) and the anticipated contributions of subsequent utterances. They also proposed that these relationships are compositional, composed of pitch accents, phrase accents and boundary tones that make up a tune. Therefore, the main components of intonation offer separate and distinct contributions to discourse interpretation which are related to mutual belief spaces in conversation, capturing the intuition that tunes sharing certain tonal features also share some aspects of meaning (Gili Fivela 2008, Prieto 2015).

Differences in accent type convey differences in meaning when interpreted in conjunction with differences in the discourse context and variation in other acoustic properties of the utterance. For example, in English H* accents are typically found in standard declarative utterances and are commonly used to convey that the accented item should be treated as *new* information in the discourse, and is part of what is being asserted in an utterance. L* accents are broadly characterized as conveying that the accented item should be treated as salient, but not part of what is being asserted. In English, L+H* accents can be used to produce a pronounced "contrastive" effect and H+!H* accents are associated with some implied sense of familiarity with the mentioned item.

As for phrasal tones, phrase accents indicate the presence of an interpretive as well as a phonological boundary (Pierrehumbert & Hirschberg 1990: 302): H- indicates that the current phrase is to be taken as forming part of a larger composite interpretive unit with the following phrase, while a L- emphasizes the separation of the current phrase from a subsequent phrase. The type of boundary tone conveys whether the current intonational phrase is forward-looking or not, that is whether this is to be interpreted with respect to some succeeding phrase or whether the direction of interpretation is unspecified.

Recent proposals have built on Pierrehumbert & Hirschberg's compositional approach to explore the meaning of English (Truckenbrodt 2012) and French pitch contours (Portes & Beyssade 2012). They argue for a systematic relationship between tonal features and their semantic primitives, but they also assume that these meanings are to some extent

context-dependent. Many of these proposals consider that intonation encodes basic meanings from which context-dependent conversational implicatures can be derived.

As an example, Armstrong & Prieto (2015) explored how intonation and context conspire to lead a listener to a given meaning. Their experimental evidence points to the dynamic interaction between context and contour, and also to the fact that individual intonation contours can differ in the type and number of meanings they convey.

6. *Acquisition and crosslinguistic comparisons*

The impact of first language (L1) prosody on second or foreign language (henceforth, L2) is widely discussed in the literature. Think of the grounding work by Mennen (2004) who, analyzing Dutch L1 consecutive bilinguals who learned Greek as L2 in their early adulthood and used it regularly, showed that pitch alignment characteristics of Dutch L1 affect the alignment features of prenuclear rises produced in Greek L2.

However, in our aim to discuss how linguists analyze prosody and intonation, we describe some studies that address issues concerning Italian and L2 learning as well as crosslinguistic comparison. Besides the specific phenomena and languages considered, the first study more generally regards prosodic structure and segmental phenomena in the acquisition of an L2, the others relate to the implementation of prominences in L2.

6.1 Prosody, constituency and vowel insertion in French L2

The prosodic structure has an impact on various aspects of speech production, among which the realization of segments that precede or follow prosodic boundaries. The role of prosodic structure in this respect has been observed in relation to both L1 and L2. As for the latter case, interesting observations stem from an acoustic and articulatory investigation related to the production of consonant clusters in French L2 by Italian speakers (D'Apolito & Gili Fivela 2013, 2018).

Sibilant clusters are common in French (where they can also undergo place assimilation; Niebuhr *et al*. 2008), while they are marked in Italian, where they are not even found across word boundaries, being the word

ending usually a vowel (few exceptions are represented by prepositions, loanwords and contexts in which word final vowel truncation occurs; Muliačić 1973; Farnetani & Busà 2004). Thus, sibilant clusters are phonotactically marked (Eckman 2008) for Italian speakers who, as a general repair strategy to produce such unusual sequences, may insert a vowel between the two consonants. Also depending on prosodic conditions, such vowel insertion may actually correspond to either an epenthetic or an intrusive vowel (Hall 2003, 2006, 2011), that is a vowel with or without an articulatory target. Besides theoretical implications, distinguishing between the two types may show the influence of prosodic factors on segment production and may be useful in order to shed light on their phonetic transcription.

D'Apolito & Gili Fivela (2013, 2018), investigated French heterosyllabic sibilant clusters (alveolar-postalveolar and postalveolar-alveolar sibilant sequences, such as /sʃ/, /ʃs/, /sʒ/, /ʒs/, /zʃ/ e /zʒ/), by creating a speech corpus of acoustic and articulatory (electromagnetic articulography, EMA) data in which consonant sequences were inserted in carrier sentences in which they were realized at a word boundary (/a_/#/_i/). Such word boundary could correspond to either a phonological phrase boundary (e.g., *Il dit tasse chinoise rapidement* 'Dice tazza cinese rapidamente') or an intonation phrase boundary (e.g. *D'abord il a dit tasse. Chinoise l'a dit après* 'Prima ha detto tazza. Cinese l'ha detto dopo'). Three advanced Italian learners of French-L2 (Lecce, Italy) and two native French speakers (Nantes, Paris) produced seven repetitions of the French corpus, both at a fast and at a normal speech rate. The authors performed an auditory evaluation (aimed at verifying the presence of the expected prosodic boundary, the presence of an inserted vowel and the consonant realization) and both an acoustic and an articulatory analysis. The analysis related to the V1C1#C2V2 sequences, including the presence of possible schwas (V0) and/or pauses (P), and it was carried on by performing acoustic measurements of duration (single segments, as well as utterance duration), speech rate (number of syllables/utterance duration), and formant values (F1, F2) for /a/, /i/ and the possible schwa. Acoustic results show that speakers differentiated between the two speech rates, and, almost with no exception, inserted a vowel at normal rate and in the case of a weak boundary; Italians kept inserting a vowel-like segment in the case of a stronger boundary, while French speakers showed a more variable behavior (as for articulatory analysis and results, we refer

to D'Apolito & Gili Fivela 2013, 2018). As far as the fast rate is concerned, Italians inserted vowels, though less than in the normal rate condition, while French inserted very few vowels in the case of strong boundaries and no vowel at all in the case of weak boundaries. Thus, the presence of a prosodic boundary is shown to clearly interact with vowel insertion in both French L1 and French L2 speech. Interestingly, formant values showed that French speakers realized schwa-like vowels, while Italians produced a more closed and anterior vowel, whose quality seemed to be affected by the prosodic context and, in any case, resembled more the following [i], rather than a schwa – see Fig. 10.

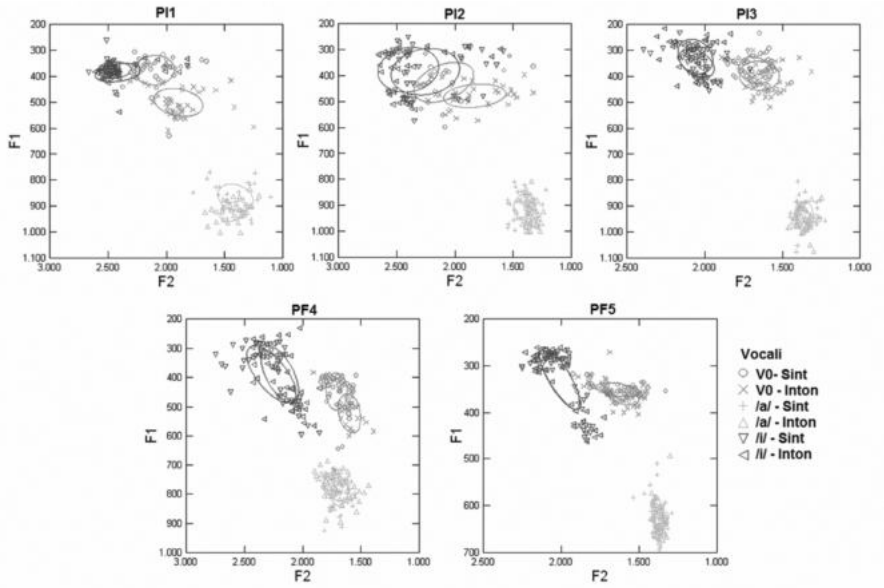

Fig. 10. Formant values of /a/, /i/ and inserted V0 vowels, produced in different prosodic conditions by Italian learners (upper panels) and French natives (lower panels; adapted from D'Apolito & Gili Fivela 2018).

Overall acoustic (and articulatory) results point to the need of taking care of prosodic conditions and differentiating the phonetic transcription, using [ɨ] rather than /ə/ for productions by Italians, who probably realize an intrusive, rather than an epenthetic vowel (with no articulatory target).

6.2 Pitch accent realization, distribution and information status in German L2

As already mentioned, pitch accents play a crucial role in conveying linguistic functions and are indeed language specific as for both their form-function mapping and their implementation, i.e. their phonetic characteristics including the temporal relation between the tonal target(s) and the segmental chain (see §2.1 and §2.2).

Not surprisingly, pitch accent realization in L2 may be as difficult as the production of other phonological events. In investigating the L2 Greek speech by Dutch L1 speakers, for instance, Mennen (2004) showed that non-native productions of L+H* prenuclear rises in L2 Greek declaratives are characterized by an earlier F0 peak alignment in comparison to native Greek productions. Along similar lines, Atterer & Ladd (2004) showed that Northern and Southern speakers of German differed in aligning the L+H* prenuclear rise in English L2 productions, in that Southern speakers produced a later alignment of both L and H targets in comparison not only to English speakers but also to Northern German speakers. Very similarly, results on Italians producing prenuclear pitch accents in German L2 (Stella 2013) showed that production accuracy varies depending on the learner's competence in L2. By means of both acoustic and articulatory data, the author shows that low competence speakers show the same pitch accent observed in their L1 (corresponding to a L+H* transcription). On the other hand, high competence learners show a more stable anchoring of tonal targets to segments and a later alignment of the expected low target, which goes in the direction of the German pitch accent (described as L*+H; Braun 2006).

In learning a foreign language, the strategies of PA assignment specific of an L2 can pose difficulties as well, especially if the native language and the target language differ typologically and what needs to be attained is mastering the prosodic properties at the interface with information structure. That is the case of Italian and German, which differ as for the accenting or deaccenting of referents that are informationally *given*. We have seen in §4 that German tends to deaccent *given* referents while Italian does not. Moreover, German differs from Italian on another respect: in broad-focus verb-final sentences the verb can be accented or deaccented according to the status of argument or adjunct of the element that precedes it (Truckenbrodt 2007), while the last lexical item of the same sentences in Italian are always pitch accented. From a typological point of view,

German can be said to be marked with respect to Italian, because both languages obey the same phonological rules of pitch accent placement (the last metrical head of the final intonational phrase gets the highest prominence of the sentence), but in German deaccenting is driven both by informational principles ('deaccent *given'*) and by syntactic constraints that do not apply in Italian. Based on those structural differences, an asymmetry in the acquisition process can be predicted: structures of an L2 that are marked will be more difficult to be learned than unmarked ones (Eckman, 2008).

A study on the acquisition of prosody of L2-Italian and L2-German by native speakers of German and Italian explores this topic (Avesani *et al*. 2015). By exploiting the same card game methodology used by Swerts *et al*. (2002), a total of five pairs of German-German and Italian-Italian speakers produced in a semi-spontaneous way NPs in which the Adjective (a colour) or the Noun (a fruit name) is *given*, *new* or *contrastive*. The speakers first played the game in their L2 and then played the game in their L1. The experimental set up allows to combine a contrastive analysis of the native languages (L1-German vs. L1-Italian), a contrastive analysis of the speakers' interlanguages (L2-Italian vs. L2-German) and an analysis of speakers' interlanguages with their native languages (L1-Italian vs. L2-German; L1-German vs L2-Italian). Results show that in Italian the final word of the noun phrase is always accented independently from its information status. When it represents *given* information, it is pitch accented in 100% of the cases. When the *given* word is NP-initial, it can be optionally accented, as it occurs in prenuclear position. On the contrary, in German the last word in the NP when it is *given* is deaccented (i.e. it is not associated with a pitch accent) in 87% of the cases and in a lesser percent also when it occurs NP-initially.

The analysis of interlanguages confirms that Italian speakers (who are advanced learners of German) transfer in their interlanguage the distribution of the accentual prominences of their L1 and do accent a German Noun or an Adjective if it is informationally *given*. Differently, when German learners of the same proficiency level speak L2-Italian they show to have acquired the prosodic accentuation of the target language and properly accent *given* information. The authors interpret the results in terms of the different cognitive weight faced by the Italian and the German learners in producing the correct accentuation in the L2. The Germans have only to select one of the strategies of accentuation already

present and active in their L1: the "structural" accentuation, according to which the last word in a broad-focus sentence that is not a verb gets a pitch accent. Conversely, to properly produce the prosody of L2-German, Italians have to master a specific type of pragmatic (de)accentuation that is not present in their L1, as well as its interplay with the phonological structure. For the Germans, the acquisition process is reduced to a suspension of the pragmatic constraints that govern the distribution of the prosodic prominences in their mother tongue; consequently, the default phonological rules do take over, and all NP final words are accented independently of their information status. On the contrary, Italians have a more difficult task: they must realize that prominences' distribution is not only phonologically-based, and that the highest prominence is not necessarily allocated rightmost in a phrase. Then, they have to master a new type of pitch accent association, which is largely ruled by the information status of the lexical items in the NP.

7. *Prosody and gestures*

Since the beginning of the Seventies, a tight relation has been observed between spoken utterances and movements of the hands, head, face and torso. According to Kendon (1972, 1980, 2004), gestures accompanying speech are organized into a hierarchy of constituents which resemble the hierarchy proposed for prosody. Since his work, arguing in favour of a coordination between gestures and Tone Groups (with reference to the proposals of the British School of Intonation, Crystal 1969), various works have been showing that gestures and speech are synchronous. Specifically, the most prominent segment of gestures tends to co-occur with the most prominent segment of speech (e.g., Birdwhistell 1952, 1970; Kendon 1972, 1980; Loehr 2012, among many others), that is, gesturing is timed to prominent syllables. Further, this timing can be influenced by prosodic boundaries.

Co-speech gestures play a crucial role in helping people to comprehend speech, especially in the case of unclear or ambiguous stretches. Further, different types of co-speech gestures have been identified, showing specific relations with the speech message and, therefore, different roles in speech coding and decoding (McNeill 1992; Kelly & Church 1997; Morsella & Krauss 2004; Goldin-Meadow & Beilock 2010; Goldin-Meadow 2013).

The main co-speech gestures, in line with McNeill's (1992) book, are *referential* and *non-referential* gestures. The former visually refer to the content of speech, while the latter offer information on the form rather than the content of the utterance. Among referential gestures, some well-known types are *deictic*, which are typically pointing gestures to specific locations, performed by means of, e.g., fingers, head, nose; *iconic* and *metaphoric*, visually illustrating concrete and abstract aspects respectively of the speech content. On the other hand, non-referential gestures, also known as *beats* (or batons; Efron 1941/1972), are two-phase movements (in/out, up/down) which do not present any discernible meaning, though they co-occur with important parts of the message. For instance, both pointing gestures and beat gestures have been shown to be tightly synchronized to prosodic prominence in spontaneous speech (Loehr 2012; Esteve-Gibert & Prieto 2013; Wagner et al. 2014). In general, *hits*, which are defined as "gestures with sudden sharp end points" (Shattuck-Hufnagel & Ren 2018: 206, but see also Shattuck-Hufnagel *et al.* 2007: 39), occur "towards the end or just after a spoken accented syllable" (Hufnagel & Ren 2018: 214). Hits have been found to be temporally synchronized to prosodic events such as pitch accents and boundary tones in Italian too (Esposito *et* al. 2007).

As far as pointing gestures are concerned, Esteve-Gibert & Prieto (2013) analyzed fifteen Catalan speakers while they pointed at a screen and produced target words with stress in different positions (es. "máma" vs "mamá"). They induced the production of target words in contrastive focus contexts, in which they were pitch accented and followed by a phrase boundary. Their analysis of pointing gestures, pitch accents and boundary tones showed that the apex of deictic gestures is coordinated with the intonation peak, and that the entire pointing gestures are bound by prosodic phrasing. As the authors observed, "the timing of their starting movements and prominence peaks (F0 peak and apex) varies if there is a preceding or an upcoming prosodic phrase boundary" (Esteve-Gibert & Prieto 2013: 863).

Besides being aligned with prosodic prominences, and pitch accents in particular, pointing gestures have indeed been found to be coordinated with prosodic boundaries. The literature on the marking of prosodic boundaries is rich, and regards signals of both incoming breaks (e.g., preboundary lengthening, boundary tone implementation) and post-boundary events (e.g. F0 reset, phrase initial strengthening).

In their articulatory investigation, Krivokapic *et al*. (2015), for instance, found that pointing gestures show phrase initial temporal lengthening which increases with boundary strength. They collected audio and synchronized kinematic data by means of electromagnetic articulometery and a motion capture system while asking their participants to point to the appropriate picture of a doll (named either miMA or MIma) while reading a sentence including the target name. By changing the stress position in the target word (first or second syllable of a disyllable, i.e., MIma and miMA), and the structure of the sentence including the target word, specifically with respect to the strength of the phrase initial boundary (word boundary, ip—intermediate phrase boundary, IP—Intonation Phrase boundary), they could collect data on the alignment of the pointing gesture and oral constriction gestures corresponding to the production of the two target words in the different prosodic conditions. Their results, though preliminary, clearly show that "1) manual and oral gestures are longer phrase-initially than phrase medially and 2) manual and oral gestures lengthen under phrase-level prominence" (2015: 1) (Fig. 11).

Summing up, the prosodic structure (heads and edges) plays a strong role on gesture timing. Gesture movements are bound by both prosodic heads (pitch accents), that is by prosodic prominence - which is related to, e.g., the management of information structure, focus, and discourse marking -, and by prosodic edges (prosodic boundaries), that is by phrasing - which is related to the managements of, e.g., syntactic grouping and turn-taking.

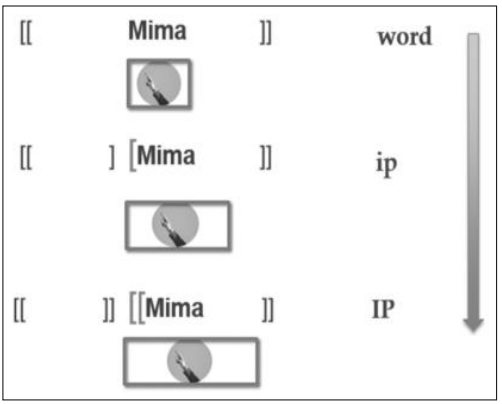

Fig. 11. Schematic representation of Krivokapic et al.*'s (2015) results on the coordination of the pointing gesture with the beginning of different prosodic boundaries.*

However, the relation between prosody and gestures goes beyond the involvement of hand movement, and prosody-related visual information may "spread" over different units. In their work on the interplay of contextual and prosodic information in the coding of politeness, Gili Fivela & Bazzanella (2014) showed the role of visual information in creating two local contexts. Specifically, the local context available to participants in the conversation, who can see each other, and a second, related context, that involves also someone who is not physically present and cannot see the interlocutors. The example the authors describe concerns a conversation that someone may have on the phone, with interlocutors who do not share visual information, while communicating also with someone who is physically present and therefore shares visual information on the context. For instance, in one of their recordings, involving a speaker who expresses on the phone her appreciation with regards to a piece of furniture while denying it with body (hand) gestures and visual expressions, it may be easily seen that both highbrow rising and hand movements participate in conveying the denial of the oral message, and that the eyebrow rising stops before the hand movement does.

The specific contribution of visual and audio information in the coding and decoding of prosody is still a debated issue (Massaro 1989; House 2002; Krahmer & Swerts 2005; Borràs-Comes & Prieto 2011, Crespo Sendra *et al*. 2013, Ambrazaitis & House under review), but the intertwined contribution of both channels has been shown even in relation with the communication of clearly linguistic information. Gili Fivela (2015), for instance, reported on the role of visual information in conveying sentence modality in Italian (variety of Lecce). Specifically, the paper focusses on the way speakers differentiate statements, *wh*-questions and exclamations by means of both prosodic and visual information. In analysing utterances supposing a positive attitude, the author finds that *wh*-questions differ from the other modalities considered as for head movement, while statements differ from other modalities as for eyebrow and lid movements.

7.1 Intonation and visual expressions in Catalan, Dutch and Italian

As already mentioned, there is a tight connection between prosodic events and visual expressions, and, not surprisingly, crosslinguistic differences have been reported with reference to multimodality as well. Crespo Sendra *et al*. (2013), for instance, investigated information seeking

and incredulity yes/no questions in Catalan and Dutch in order to check if differences in the phonology of intonation have an impact on visual information. The authors observe that in Catalan information seeking questions are expressed by means of a L*+H H% pattern, and incredulity questions show the same phonological patter which is realized, though, by reaching a higher final tonal target. On the contrary, in Dutch the two question types are expressed by means of different sequences of phonological events. In particular, information seeking questions are expressed by means of a L* HH%, while incredulity questions show a L+H* LH% pattern. As far as visual information is concerned, the authors observed that incredulity questions show some degree of eyebrow furrowing and eyelid closure in both languages. Results of Crespo Sendra *et al*. (2013) investigation on the perception of audio-visual information by Catalan and Dutch listeners show that the former give more importance to facial cues than the latter, and suggest that this may be due to the more subtle (or ambiguous) information conveyed by the audio channel in Catalan.

A question then arises as for Italian, where the same phonological pattern, H+L* L%, may be used to convey both statements and *wh*-questions, with no disambiguation at the phonological level, similarly to what was observed for Catalan in relation to yes-no information-seeking questions and incredulity questions. Gili Fivela (2015) addresses this issue by means of both production and perception data, regarding neutral statements, *wh*-questions suggesting surprise, and exclamations (for the sake of clarity, here only data on statements and questions are reported; as for exclamations, see Gili Fivela 2015). In the production experiment, five subjects were audio and video recorded while producing dialogues including the target sentences. The AM analysis of intonational patterns confirmed that both statements and *wh*-questions were realized by means of a H+L* L% pattern; however, the analysis of visual expression and head movements (*Facial Action Coding Scheme*, Ekman 1982; Ekman *et al.* 2002) pointed out that *wh*-questions differed as for both head movement and visual expressions. Specifically, statements differed from other utterances as for eyebrow and lid movements, and questions differed from others as for head movements. These results seem to point to a possible integration of audio and visual information, with the latter compensating for the lack of phonological differences.

However, perception results offered a different picture. Subjects had to decide on the modality expressed by audio-video stimuli that were either regular, congruous stimuli corresponding to a specific modality, or

incongruous stimuli in which audio and video did not match. Results showed that subjects were strongly influenced in their judgements when the video corresponded to a question and the audio to a statement. In such situation, statements were recognized only in 55% of cases, while in 40% a question was identified. In the reversed case, in which the audio of a question was imposed over the video of a statement, nothing similar happened, as listeners perceived questions in 93% of cases. Thus, visual information corresponding to a question influences the perception of the audio of a statement, and not the other way around. This is taken to point to a more complex picture than that depicting a balancing of audio and visual information in terms of a negative correlation, that is a balance in which the role of one channel depends on what happens in the other. The balancing of information within the same channel seems also to be important, in that marked facial expressions (in this case, corresponding to surprised *wh*-questions) seem to affect the interpretation of utterances which are not associated to marked information on the same video channel (here, neutral statements)

Thus, a more complex integration could possibly take place and should be considered even in a crosslinguistic perspective. Such integration could possibly play a role in accounting for the variability observed in audio realizations, and could be taken into account in the transcription of intonation categories and contours, in both L1 and L2 studies.

8. *Conclusions*

The paper offered an overview of how linguists analyze prosody and intonation. By discussing the main features of prosody and intonation in terms of both form and function, the added value of a framework that hypothesizes a phonological structure for intonation was shown. The Autosegmental-Metrical framework allows investigators to identify linguistically relevant units, which may be both phonologically labelled and phonetically measured. The analysis may then regard both phrasing and prominences. Further, it may also be performed with reference to a transcription system, the ToBI system, which needs to be developed for each specific (variety of) language, being phonological in nature, but it is based on the very same principles, in that it requires a phonological coding and it also allows for a phonetic analysis of relevant units. Such framework may then be ideal in analyzing language/variety specific phonological inventories.

Specific attention was devoted to the main functions that are crucially played by prosody and intonation, showing that, besides conveying prominence in general, they express sentence modality, focus, phrasing and information structure and, unsurprisingly, they may solve global ambiguities.

Further, issues were addressed concerning acquisition and crosslinguistic comparison, again with the aim of discussing how linguists analyze prosody and intonation. The attention was oriented towards Italian and some works investigating L2 acquisition with respect to prosodic structure and segmental phenomena (specifically, prosody, constituency and vowel insertion in French L2), and prominence patterns and L2 acquisition (specifically, the study of pitch accent realization, distribution and information status in German L2).

Finally, multimodality was also addressed, showing the tight interplay of prosody, co-speech gestures and visual expressions, and pointing out that the analysis of multimodal information requires to refer to prosodic units and may be relevant for linguistic purposes. Some works underlying the interplay of crosslinguistic differences and multimodal information modulation were also discussed in the end (specifically, the role of intonation and visual expressions in Catalan, Dutch and Italian).

Overall these studies show that prosody and intonation have been analyzed by adopting various perspectives and methodologies, but they all refer to a linguistic structure and to linguistic functions, as well as to an abstract representation of intonational events. These aspects allow not to be mislead by phonetic details and variability, offering the reference for analyzing a continuously varying signal, that is the verbal chain. Only adopting a shared framework, including both a phonological and a phonetic level and accounting for both phrasing and prominences, allows to succesfully face investigations on either L1 or L2. Models assuming a direct correspondence between form and function would not be equally successful in this respect.

References

Ambrazaitis, Gilbert & House, David. Under review. Is phonetic prominence underlyingly multimodal? *Laboratory Phonology, Special Collection on Phonological categories and prosodic modulation: importance, representation and implementation*.

Armstrong, Meghan & Prieto, Pilar. 2015. The contribution of context and contour to perceived belief in polar questions. *Journal of Pragmatics* 81. 77–92.

Arvaniti, Amalia & Ladd, Robert D. & Mennen, Ineke. 2000. What is a starred tone: Evidence from Greek. In Broe, Michael & Pierrehumbert, Janet (eds.), *Papers in Laboratory Phonology V*, 119–131. Cambridge: Cambridge University Press.

Arvaniti, Amalia & Ladd, Robert D. 2009. Greek wh-questions and the phonology of intonation. *Phonology* 26. 43–74.

Arvaniti, Amalia. 2007. On the Relationship between Phonology and Phonetics (Or Why Phonetics is not Phonology). In Trouvain, Jürgen & Barry, William John (eds.), *Proceedings of the XVI International Congress of Phonetic Sciences*, 19–24. Saarbrücken: Universität des Saarlandes.

Arvaniti, Amalia. 2012. Segment-to-Tone Association. In: Cohn, Abigail & Fougeron, Cécile & Huffman, Marie (eds.), *The Oxford Handbook of Laboratory Phonology*, 256–275. Oxford: Oxford University Press.

Arvaniti, Amalia. 2016. Analytical Decisions in Intonation Research and the Role of Representations: Lessons from Romani. *Laboratory Phonology* 7(1). 6. http://doi.org/10.5334/labphon.14

Atterer, Michaela & Ladd, Robert D. 2004. On the phonetics and phonology of "segmental anchoring of F0: Evidence from German". *Journal of Phonetics* 32. 177–197.

Avesani, Cinzia & Bocci, Giuliano & Vayra, Mario & Zappoli, Alessandra (2015). Prosody and information status in Italian and German L2 intonation. In Chini, Maria (a cura di). *Il parlato in [italiano] L2: aspetti pragmatici e prosodici /[Italian] L2 Spoken Discourse: Pragmatic and Prosodic Aspects*, 93-116. Milano, Franco Angeli (Materiali Linguistici, 71).

Avesani, Cinzia & Vayra, Mario. 2004. Focus ristretto e focus contrastivo in italiano. In Albano Leoni, Federico & Cutugno, Franco & Pettorino, Massimo & Savy, Renata (a cura di), *Atti del Convegno nazionale Il Parlato Italiano*, F01, 1–20. Napoli: D'Auria Editore.

Avesani, Cinzia & Vayra, Mario. 2005. Accenting deaccenting and information structure in Italian dialogues. *SIGdial6–2005*: 19–24. https://www.isca-speech.org/archive_open/sigdial6/sgd6_019.html.

Avesani, Cinzia. 1995. ToBIt. Un sistema di trascrizione per l'intonazione italiana. In Lazzari, Giovanni (ed.), *Atti delle V Giornate di Studio del Gruppo di Fonetica Sperimentale,* 85–98. Trento: Servizio Editoria ITC.

Avesani, Cinzia. 1999. Quantificatori, negazione e costituenza sintattica. Costruzioni potenzialmente ambigue e il ruolo della prosodia. In Benincà, Paola & Mioni, Alberto & Vanelli, Laura (eds.), *Fonologia e morfologia dell'italiano e dei dialetti d'Italia*, 153–200. Roma: Bulzoni.

Avesani, Cinzia. 2003. La prosodia del focus contrastivo. Un accento particolare? In Marotta, Giovanna & Nocchi, Nadia (a cura di), *La coarticolazione. Atti delle XII Giornate di studio del Gruppo di Fonetica Sperimentale*, 157–167. Pisa: ETS.

Bard, Ellen G. & Aylett, Matthew. 1999. The dissociation of deaccenting, givenness, and syntactic role in spontaneous speech. In Ohala, John J. & Hasegawa, Yoko & Ohala, Manjari & Granville, Daniel & Bailey, Ashlee C. (eds.), *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco: University of California.

Baumann, Stefan & Riester, Arndt. 2013. Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In Elordieta, Gorka & Prieto, Pilar (eds.), *Prosody & Meaning*, 119–162. Berlin – New York: Mouton de Gruyter.

Beach, Ceryl. 1991. The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language* 30. 644–663.

Beckman, Mary & Ayers, Gail. 1997. *Guidelines for ToBI labelling*. Columbus: The Ohio State University Research Foundation.

Beckman, Mary & Edwards, Jane. 1990. Lengthening and shortening and the nature of prosodic constituency. In Kingston, John & Beckman, Mary (eds.), *Laboratory Phonology I*, 152–178. Cambridge: Cambridge University Press.

Beckman, Mary & Edwards, Jane. 1994. Articulatory evidence for differentiating stress categories. In Keating, Patricia (ed.), *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form,* 7–33. Cambridge: Cambridge University Press.

Beckman, Mary & Hirschberg, Julia & Shattuck–Huffnagel, Stefanie. 2005. The original ToBI system and the evolution of the ToBI framework. In: Jun, Sun-Ah (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, 10–54. Oxford: Oxford University Press.

Beckman, Mary & Pierrehumbert, Janet. 1986. Intonational structure in English and Japanese. *Phonology Yearbook* 3. 255–310.

Beckman, Mary. 1996. The Parsing of Prosody. *Language and Cognitive Processes* 11(1/2). 17–68.

Birch, Stacy & Clifton, Charles. 1995. Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech* 38(4). 365–391.

Birdwhistell, Ray L. 1952. *Introduction to Kinesics: An annotated System for the Analysis of Body Motion and Gesture*. Louisville: University of Louisville.

Birdwhistell, Ray L. 1970. *Kinesics and Context: Essays on Body–Motion Communication*. Philadelphia: University of Pennsylvania Press.

Bocci, Giuliano & Avesani, Cinzia, 2011. Phrasal Prominences Do Not Need Pitch Movements: Postfocal Phrasal Heads in Italian. In Cosi, Piero & De Mori, Renato & Di Fabbrizio, Giuseppe & Pieraccini, Roberto (eds.), *INTERSPEECH*-2011: 1357–1360.

Bocci, Giuliano & Avesani, Cinzia. 2015. Can the metrical structure of Italian motivate focus fronting? In Shlonsky, Ur (ed.), *Beyond Functional Sequence. The cartography of syntactic structures,* 23–41. Oxford: Oxford University Press.

Bocci, Giuliano. 2013. *The Syntax–Prosody Interface*, Amsterdam: John Benjamins.

Bögels, Sara & Schriefers, Herbert & Vonk, Wietske & Chwilla, Dorothee. 2011. Prosodic Breaks in Sentence Processing Investigated by Event–Related Potentials. *Language and Linguistics Compass* 5(7). 424–440.

Bolinger, Dwight. 1951. Intonation: Levels versus configurations. *Word* 7. 199–210.

Borràs–Comes, Juan & Prieto, Pilar. 2011. Seeing tunes. The role of visual gestures in tune interpretation. *Laboratory Phonology* 2(2). 355–380.

Braun, Bettina & Chen, Aoju. 2012. Now for something completely different: Anticipatory effects of intonation. In Niebuhr, Oliver (ed.), *Understanding Prosody: The Role of Context, Function and Communication*, 289–311. Berlin: De Gruyter.

Braun, Bettina. 2006. Phonetics and phonology of thematic contrast in German. *Language and Speech*, 49(4). 451–493.

Brown, Gillian. 1983. Prosodic structure and the given/new distinction. In Cutler, Anne & Ladd, Robert D. (eds.), *Prosody: Models and Measurement*, 67–77. BerlIn Springer.

Bruce, Gösta. 1977. *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.

Büring, Daniel. 2016. *Intonation and Meaning*. Oxford: Oxford University Press.

Byrd, Dany & Saltzman, Elliot. 2003. The Elastic Phrase: Modelling the Dynamics of Boundary-Adjacent Lengthening. *Journal of Phonetics* 31. 149–180.

Carroll, Patrick & Slowiaczek, Maria. 1987. Modes and Modules: Multiple Pathways to the Language Processor. In Garfield, Jay L. (ed.), *Modularity in Knowledge Representation and Natural–language Understanding*, 221–247. Oxford: The MIT Press.

Chafe, Wallace L. 1976. Givenness, Contrastiveness, Definiteness, Subjects, Topics and Point of View. In Li, Charles (ed.), *Subject and Topic*, 27–55. New York, Academic Press.

Cho, Taehong & Keating, Patricia. 2001. Articulatory and Acoustic Studies on Domain–Initial Strengthening In Korean. *Journal of Phonetics* 29(2). 155–190.

Cooper, William & Sorensen, John. 1981. *Fundamental Frequency in Sentence Production*. Heidelberg: Springer.

Crespo–Sendra, Veronica & Kaland, Constantijn & Swerts, Mark & Prieto, Pilar. 2013. Perceiving Incredulity: The Role of Intonation and Facial Gestures. *Journal of Pragmatics* 47. 1–13.

Cresti, Emanuela & Moneglia, Massimo. 2018. The illocutionary basis of Information Structure. Language into Act Theory (L–AcT). In Adamou, Evangelia & Haude, Katharina & Vanhove, Martine (eds.), *Information structure in lesser described languages: Studies in prosody and syntax*, 359–401. Amsterdam: Benjamins.

Cresti, Emanuela. 2005. Per una nuova classificazione dell'illocuzione a partire da un corpus di parlato (LABLITA). In Burr, Elisabeth (ed.), *Tradizione e innovazione: il parlato. Atti del VI Convegno internazionale SILFI*, 233–246. Pisa: Cesati.

Cruttenden, Alan. 1993. The de–accenting and reaccenting of repeated lexical items. *Working Papers 41, Dept. of Linguistics and Phonetics, Lund, Sweden*.16–19.

Crystal, David. 1969. *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press CUP.

D'Apolito, Sonia & Gili Fivela, Barbara. 2013. Acoustic and articulatory study of French sibilant clusters. In Galatà, Vincenzo (ed.), *Multimodalità e multilingualità: la sfida più avanzata della comunicazione orale*, 135–150. Roma: Bulzoni.

D'Apolito, Sonia & Gili Fivela, Barbara. 2018. Strategie nella produzione del parlato non nativo: l'inserzione vocalica in francese L2. *Studi Italiani di Linguistica Teorica e Applicata* XLVII (2). 269–292.

D'Imperio, Mariapaola & Gili Fivela, Barbara & Niebhur, Oliver. 2010. Alignment perception of high intonational plateaux in Italian and German. In *Proceedings of Speech Prosody 2010. Special Session on Shape, Scaling, and Alignment Effects in the Production and Perception of F0 Event*s. Chicago: ISCA Speech 100186:1–4.

D'Imperio, Mariapaola & House, David. 1997. Perception of questions and statements in Neapolitan Italian. In Botinis, Antonis (ed.), *Intonation: Theory, Models and Applications*, 251–254. Athens: Athanasopoulos.

D'Imperio, Mariapaola & Michelas, Amandine. 2014. Pitch scaling and the internal structuring of the Intonation Phrase in French. *Phonology* 31(1). 95–122.

D'Imperio, Mariapaola. 2000. The role of perception in defining tonal targets and their alignment. Columbus: Ohio State University. (Tesi di dottorato.)

D'Imperio, Mariapaola. 2002. Language–specific and universal constraints on tonal alignment: The nature of targets and anchors. In Bel, Bernard & Marlien, Isabelle (eds.), *Proceedings of the Speech Prosody 2002 Conference*, 101–106. Aix-en-Provence: Laboratoire Parole et Langage.

D'Imperio, Mariapaola & Cangemi, Francesco & Grice, Martine. 2016. *Introducing Advancing Prosodic Transcription. Laboratory Phonology* 7(1). 4. http://doi.org/10.5334/labphon.32.

Dahan, Delphine & Tanenhaus Michael & Chambers, Craig. 2002. Accent and reference resolution in spoken–language comprehension. *Journal of Memory and Language* 47. 292–314.

Eckman, Fred R. 2008. Typological markedness and second language phonology. In Hansen Edwards, Jette G. & Zampini, Mary L. (eds.), *Phonology and Second Language Acquisition*, 95–115. Philadelphia: Benjamins.

Edwards, Jane & Beckman, Mary & Fletcher Jane. 1991. The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America* 89(1). 369–382.

Efron, David. 1972 (1941). *Gesture, race and culture*. The Hague: Mouton.

Ekman, Paul & Friesen, Wallace V. & Hager, Joseph C. 2002. *The Facial Action Coding System* CD–ROM. Salt Lake City: Research Nexus.

Ekman, Paul. 1982. Methods for Measuring Facial Action. In Scherer, Klaus R. & Ekman, Paul (eds.), *Handbook of Methods in Nonverbal Behavior Research*, 45–90. Cambridge: Cambridge University Press.

Esposito, Anna & Esposito, Daniela & Refice, Mario & Savino, Michelina & Shattuck–Hufnagel, Stefanie. 2007. A Preliminary Investigation of the Relationships between Gestures and Prosody in Italian. In Esposito, Anna & Bratanić, Maja & Keller, Eric & Marinaro, Maria (eds.), *Fundamentals of verbal and nonverbal communication and the biometric issue*, 65–75. Amsterdam: IOS Press.

Esteve–Gibert, Nuria & Prieto, Pilar. 2013. Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research* 56(3). 850–865.

Farnetani, Edda & Busà, Maria Grazia. 2004. Italian clusters in continuous speech. In *Proceedings of the 3rd International Conference on Spoken Language Processing,* 359–362. Yokohama: Japan.

Fougeron, Cécile & Keating, Patricia K. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101(6). 3728–3740.

Frota, Sonia & Prieto, Pilar. 2015. Intonation in Romance: Similarities and differences. In Frota, Sonia & Prieto, Pilar (eds.), *Intonation in Romance*, 140–197. Oxford: Oxford University Press.

Frota, Sonia & Vigario, Marina. 2018. Syntax-Phonology Interface. In Aronoff, Mark (ed.), *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.111.

Frota, Sonia. 2000. *Prosody and focus in European Portuguese: Phonological phrasing and intonation.* Abingdon: Routledge.

Frota, Sonia. 2017. Segment-to-Tone Association. In Cohn, Abigail & Fougeron, Cécile & Huffman, Marie (eds.), *The Oxford Handbook of Laboratory Phonology*, 265–275. Oxford: Oxford University Press.

Gili Fivela, Barbara & Avesani, Cinzia & Bocci, Giuliano & D'Imperio, Mariapaoala & Giordano, Rosa & Marotta, Giovanna & Savino, Michelina & Sorianello, Patrizia. 2015. Intonational phonology of the regional varieties of Italian. In Frota, Sonia & Prieto, Pilar (eds.), *Intonation in Romance*, 140–197. Oxford: Oxford University Press.

Gili Fivela, Barbara & Bazzanella, Carla. 2014. The relevance of prosody and context to the interplay between intensity and politeness. An exploratory study on Italian. *Journal of Politeness Research* 10(1). 97–126.

Gili Fivela, Barbara & Iraci, Massimiliano. 2017. Variation of intonation across Italy: The case of Palermo Italian. In Bertini, Chiara & Celata, Chiara & Lenoci, Giovanna & Meluzzi, Chiara & Riccci, Irene (a cura di), *Fattori sociali e biologici nella variazione fonetica – Social and Biological Factors in Speech Variation*, 169–190. Milano: Officinaventuno.

Gili Fivela, Barbara & Nicora, Francesca. 2018. Intonation in Liguria and Tuscany: checking for similarities across a traditional isogloss boundary. In Vietti, Alessandro & Spreafico, Lorenzo & Mereu, Daniela & Galatà, Vincenzo (a cura di), *Il parlato nel contesto naturale – Speech in the natural context*, 131–156. Milano: Officinaventuno.

Gili Fivela, Barbara & Avesani, Cinzia & Nicora, Francesca. submitted. Variation and contact in Italian intonation across the La Spezia-Rimini line. *Language and Speech*, Special Issue on *Language contact and speaker accommodation.*

Gili Fivela, Barbara. 2008. *The Phonetics and Phonology of Intonation: The case of Pisa Italian*. Alessandria: Edizioni dell'Orso.

Gili Fivela, Barbara. 2012. Meanings, shades of meanings and prototypes of intonational categories. In Gorka, Elordieta & Prieto, Pilar (eds.), *Prosody and meaning*, 197–237. Berlin: Mouton de Gruyter.

Gili Fivela, Barbara. 2015. L'integrazione di informazioni multimodali: prosodia ed espressioni del volto nella percezione del parlato. In Pistolesi, Elena & Pugliese, Rosa & Gili Fivela, Barbara (a cura di), *Parole, gesti, interpretazioni. Studi linguistici per Carla Bazzanella*, 107–127. Roma: Aracne.

Goldin–Meadow, Susan & Beilock, Sian L. 2010. Action's influence on thought: The case of gesture. *Perspectives on Psychological Science* 5(6). 664–674.

Goldin–Meadow, Susan. 2013. How our gestures help us learn. In Müller, Cornelia & Cienki, Alan & Fricke, Ellen & Ladewig, Silva H. & McNeill, David & Teßendorf, Sedinha (eds.), *Body–language–communication: An international handbook on multimodality in human interaction*, 792–803. Berlin: Mouton De Gruyter.

Goldsmith, John. 1979. *Autosegmental phonology*. Garland: New York.

Grabe, Esther & Greg Kochanski & John Coleman. 2003. Quantitative modelling of intonational variation. *Proceedings of Speech Analysis and Recognition in Technology, Linguistics and Medicine*: 1–23. http://www.phon.ox.ac.uk/oxigen/Grabe_Kochanski_Coleman_2003.pdf

Grice, Martine & D'Imperio, Mariapaola & Savino, Michelina & Avesani, Cinzia. 2005. Strategies for intonation labelling across varieties of Italian. In Jun, Sun-Ah (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, 362–389. Oxford: Oxford University Press.

Grice, Martine. 1995. *The Intonation of Interrogation in Palermo Italian: Implications for Intonational Theory*. Tubingen: Niemeyer.

Grosz, Barbara & Sidner, Candice. 1986. Attentions, intentions and the structure of discourse. *Computational Linguistics* 12. 175–204.

Gussenhoven, Carlos. 1984. *On the grammar and semantics of sentence accents*. Dordrech: Foris.

Gussenhoven, Carlos. 2007. The phonology of intonation. In de Lacy, Paul (ed.), *The Cambridge Handbook of Phonology*, 253–280. Cambridge, Cambridge University Press.

Hall, Nancy E. 2003. Gestures and segments: Vowel intrusion as overlap. Amherst: University of Massachusetts. (Tesi di dottorato.)

Hall, Nancy E. 2006. Cross-linguistic patterns of vowel intrusion. *Phonology* 23. 387–429.

Hall, Nancy E. 2011. Vowel epenthesis. In van Oostendorp, Marc & Ewen, Colin J. & Hume, Elisabeth & Rice, Keren (eds.), *The Blackwell Companion to Phonology*, 1576–1596. Malden: Wiley–Blackwell.

Hirschberg, Julia & Avesani, Cinzia. 2000. Prosodic disambiguation in English and Italian. In Botinis, Antonis (ed.), *Intonation. Analysis, modelling and technology*, 87–96. Dordrecht: Kluwer Academic Publishers.

Hirst, Daniel & Di Cristo, Albert. 1998. A survey of intonation systems. In Hirst, Daniel & Di Cristo, Albert (eds.), *Intonation Systems: A Survey of Twenty Languages*, 1–44. Cambridge: Cambridge University Press.

House, David. 2002. Intonation and visual cues in the perception of interrogative mode in Swedish. In Hansen, John & Pellom, Bryan (eds.) *Proceedings of the 7th International Conference on Spoken Language Processing*, 1957–60. http://www.isca–speech.org/archive/icslp02.

Jun, Sun–Ah & Fougeron, Cécile. 1995. The accentual phrase and the prosodic structure of French. In Elenius, Kjell & Branderud, Peter (eds.) *Proceedings of the XIIIth International Congress of Phonetic Sciences*, volume 2, 722–725. Stockholm, Stockholm University.

Jun, Sun-Ah (ed.). 2005. *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

Jun, Sun-Ah (ed.). 2014. *Prosodic Typology II: The Phonology of Intonation and Phrasing*, Oxford: Oxford University Press.

Keating, Patricia & Cho, Taehong & Fougeron, Cécile & Hsu Chai-shune. 2004. Domain–initial articulatory strengthening in four languages. In Local, John & Ogden, Richard & Temple, Rosalind (eds.), *Phonetic interpretation. Papers in laboratory phonology VI*, 145–163. Cambridge: Cambridge University Press.

Kelly, Spencer D. & Church, Breckinridge R. 1997. Can children detect information conveyed through other children's nonverbal behaviors? *Cognition and Instruction* 15. 107–134.

Kendon, Adam. 1972. Some relationships between body motion and speech: an analysis of an example. In Siegman, Aron Wolfe & Pope, Benjamin (eds.), *Studies in Dyadic Communication*, 177–210. New York: Pergamon.

Kendon, Adam. 1980. Gesticulation and speech: two aspects of the process of utterance. In Kay, Mary Ritchey (ed.), *The Role of Nonverbal Communication*, 207–227. Berlin: De Gruyter.

Kendon, Adam. 2004. *Gesture: Visible Action as Utterance*. Cambridge. Cambridge University Press.

Kjelgaard, Margaret M. & Speer, Shari R. 1999. Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language* 40. 153–194.

Kohler, Klaus. 2006. Paradigms in experimental prosodic analysis: From measurements to function. In Sudhoff, Stefan & Lenertové, Denisa & Meyer, Roland & Pappert, Sandra & Augurzky, Petra & Mleinek, Ina & Richter, Nicole & Schliesser, Johannes (eds.), *Methods in Empirical Prosody Research*, 123–152. Berlin: De Gruyter.

Krahmer, Emiel & Swerts, Mark. 2005. How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech* 48(1). 29–54.

Krifka, Manfred. 2007. Basic notions of information structure. In Fery, Caroline & Fanselow, Gisbert & Krifka, Manfred (eds.), *The Notions of Information Structure. Working Papers of the SFB 632 Volume 6*, 13–55. Potsdam: Universitätsverlag Potsdam.

Krivokapic, Jelena & Tiede, Marc & Tyrone, Martha. 2015. A kinematic analysis of prosodic structure in speech and manual gestures. The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*, paper number 658. Glasgow: the University of Glasgow. https://www.internationalphoneticassociation.org/icphs–proceedings/ICPhS2015/Papers/ICPHS0658.pdf. 1-5.

Ladd, Robert & Faulkner, Dan & Faulkner, Hanneke & Schepman, Astrid. 1999. Constant segmental anchoring of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America* 106(3). 1543–1554.

Ladd, Robert. 1996/2008. *Intonational Phonology*. Cambridge: Cambridge University Press.

Lambrecht, Knud. 1994. *Information structure and sentence form: Topics, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Leben, William. 1973. *Suprasegmental phonology*. Cambridge: Massachusetts Institute of Technology. (Tesi di dottorato.)

Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge: The MIT Press.

Liberman, Mark, & Pierrehumbert, Janet. 1984. Intonational Invariance under Changes in Pitch Range and Length. In Aronoff, Mark & Oehrle, Richard & Kelley, Frances & Stephens, Bonnie Wilker (eds.), *Language Sound Structure*, 157–234. Cambridge: The MIT Press.

Liberman, Mark. 1975. The intonational system of English. Cambridge: Massachusetts Institute of Technology. (Tesi di dottorato.)

Libermann, Mark & Prince, Alan. 1977. On Stress and Linguistic Rhythm. *Linguistic Inquiry* 8. 249–336.

Loehr, Daniel. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology* 3. 71–89.

Massaro, David. 1989. Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology* 21(3). 398–421.

McNeill, David. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.

Mennen, Ineke. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32. 543–563.

Moneglia, Massimo. 2006. Units of Analysis of Spontaneous Speech and Speech Variation in a Cross-linguistic Perspective. In Kawaguchi, Yuji & Zaima, Susumu & Takagaki, Toshihiro (eds.), *Spoken Language Corpus and Linguistics Informatics*, 153–179. Amsterdam: Benjamins.

Morsella, Ezequiel & Krauss, Robert. 2004. The role of gestures in spatial working memory and speech. *The American Journal of Psychology* 117(3). 411–424.

Muliačić, Žarco. 1973. *Fonologia della lingua italiana*. Bologna: Il Mulino.

Nespor, Marina & Vogel, Irene. 1986/2007. *Prosodic phonology*. Dordrecht: Foris.

Niebuhr, Oliver & Lancia, Leonardo & Meunier, Christine. 2008. On place assimilation in French sibilant sequences. In Sock, Rudolph & Fuchs, Susanne & Laprie, Yves (eds.), *Proceedings of the VII International Seminar on Speech Production*, 221–224. Le Chesnay-Rocquencourt: INRIA.

Niebuhr, Oliver. 2007. The signalling of German rising-falling intonation categories – The interplay of synchronization, shape, and height. *Phonetica* 64(2–3). 174–193.

Pierrehumbert, Janet & Beckman, Mary. 1988. *Japanese tone structure*. Cambridge, The MIT Press.

Pierrehumbert, Janet & Hirschberg, Julia. 1990. The meaning of intonational contours in the interpretation of discourse. In Cohen, Phillip R. & Morgan, Jerry & Pollack, Martha E. (eds.), *Intentions in communication*, 271–310. Cambridge: The MIT Press.

Pierrehumbert, Janet & Talkin, David. 1992. Lenition of /h/ and Glottal Stop. In Docherty, Gerard & Ladd, Robert D. (eds.), *Gesture, Segment, Prosody*, 90–117. Cambridge: Cambridge University Press.

Pierrehumbert, Janet. 1980. *The Phonology and Phonetics of English intonation*. Cambridge: Massachusetts Institute of Technology. (Tesi di dottorato.)

Pierrehumbert, Janet. 1999. Prosody and intonation. In Wilson, Robert & Keil, Frank (eds.), *The MIT Encyclopedia of the Cognitive Sciences,* 679–682. Cambrige: The MIT Press.

Pike, Kenneth. 1945. *The intonation of American English*. Ann Arbor: University of Michigan Press.

Portes, Cristel & Beyssade, Claire. 2012. Is intonational meaning compositional? *Verbum* XXXIV. 3–27.

Price, Patti Jo & Ostendorf, Mary &, Shuttuck–Hufnagel, Stefanie & Fong, Cynthia. 1991. The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America* 90(6). 2956–2970.

Prieto, Pilar. 2014. The intonational phonology of Catalan. In Jun, Sun-Ah (ed.), *Prosodic Typology II: The Phonology of Intonation and Phrasing*, 43–80. Oxford. Oxford University Press.

Prieto, Pilar. 2015. Intonational meaning. *WIREs Cognitive Science* 6. 371–381.

Prince, Ellen. 1981. Toward a taxonomy of given–new information. In Cole, Peter (ed.), *Radical pragmatics*, 223–255. New York: Academic Press.

Prince, Ellen. 1992. The ZPG letter: subject, definiteness and information status. In Thompson, Sandra & Mann, William (eds.), *Discourse description*, 295–325. Amsterdam: Elsevier.

Pynte, Joël & Prieur, Bénédicte. 1996. Prosodic breaks and attachment decisions in sentence processing. *Language and Cognitive Processes* 11. 165–192.

Röhr, Christine & Baumann, Stefan. 2010. Prosodic marking of information status in German. In *Proceedings of the 5th International Conference on Speech Prosody*: 1–4. http://www.isca–speech.org/sp2010/ITRW

Rooth, Maths. 1992. A theory of focus interpretation. *Natural Language Semantics* 1. 75–116.

Savino, Michelina. 2012. The intonation of polar questions in Italian: Where is the rise? *Journal of the International Phonetic Association* 42(1). 23–48.

Schafer, Ami. 1997. Prosodic parsing: The role of prosody in sentence comprehension. University of Massachusetts. (Tesi di dottorato).

Schumacher, Petra & Baumann, Stephan. 2011. (De-)Accentuation and the Processing of Information Status: Evidence from Event–Related Brain Potentials. *Language and Speech* 55(3). 361–381.

Selkirk, Elisabeth. 1978. On prosodic structure and its relation to syntactic structure. In Fretheim, Thorse (ed.), *Nordic Prosody II*, Trondheim, TAPIR. 111-140.

Selkirk, Elisabeth. 1984. *Phonology and Syntax*. Cambridge, The MIT Press.

Selkirk, Elisabeth. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3. 371–405.

Selkirk, Elisabeth. 1995. Sentence Prosoedy: Intonation, Stress and Phrasing. In Goldsmith, John, (ed.), *The Handbook of Phonological Theory*, 550–569. Cambridge: Blackwell.

Selkirk, Elisabeth. 2005. Comments on Intonational Phrasing in English. In Frota, Sonia & Vigario, Marina & Freitas, Maria Joao (eds.), *Prosodies*, 11-58. Berlin: Mouton de Gruyter.

Selkirk, Elisabeth. 2008. Contrastive Focus, Givenness and the Unmarked Status of Discourse-New. *Acta Linguistica Hungarica* 55(3). 331–346.

Selkirk, Elisabeth. 2011. The Syntax–Phonology Interface. In Goldsmith, John & Riggle, Jason & Yu, Alan (eds.), *The Handbook of Phonological Theory*, Second Edition, 435–484. Malden, Blackwell.

Shattuck-Hufnagel, Stefanie & Ren, Ada. 2018. The Prosodic Characteristics of Non-referential Co-speech Gestures in a Sample of Academic-Lecture-Style Speech. *Frontiers in Psychology* 9. 1514. 205-2017.https://www.ncbi.nlm.nih.gov / pmc/articles/PMC6137092/

Shattuck-Hufnagel, Stefanie & Turk, Alice. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25(2). 193–247.

Shattuck-Hufnagel, Stefanie & Yasinnik, Yelena & Veilleux, Nanette & Renwick, Margaret. 2007. Method for Studying the Time Alignment of Gestures and Prosody in American English: 'Hits' and Pitch Accents in Academic-Lecture-Style Speech. In Esposito, Anna & Bratanić, Maja & Keller, Eric & Marinaro, Maria (eds.), *Fundamentals of verbal and nonverbal communication and the biometric issue*, 34–44. Amsterdam: IOS Press.

Steinhauer, Karsten & Friederici, Angela. 2001. Prosodic Boundaries, Comma Rules, and Brain Responses: The Closure Positive Shift in ERPs as a Universal Marker for Prosodic Phrasing in Listeners and Readers. *Journal of Psycholinguistic Research* 30(3): 191–196.

Steinhauer, Karsten & Friederici, Angela. 1999. Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience* 2(2): 191-196

Stella, Antonio. 2013. Coordinazione interarticolatoria nella produzione dell'intonazione del tedesco come lingua straniera. In Galatà, Vincenzo (a cura di), *Multimodalità e multilingualità: la sfida più avanzata della comunicazione orale*, 411–426. Roma: Bulzoni.

Stirling, Lesley & Wales, Roger. 1996. Does prosody support or direct sentence parsing? *Language and Cognitive Processes* 11(1-2). 193–212.

Swerts, Mark & Krahmer, Emiel & Avesani, Cinzia. 2002. Prosodic marking of information status in Dutch and Italian: a comparative analysis. *Journal of Phonetics* 30(4). 629–654.

't Hart, Johan & Collier, René & Cohen, Antonie. 1990. *A Perceptual Study of Intonation: An Experimental–Phonetic Approach to Speech Melody*. Cambridge: Cambridge University Press.

Terken Jack & Hirschberg, Julia. 1994. Deaccentuation of words representing given information: effects of persistence of grammatical function and surface position. *Language and Speech* 37(2): 125–145.

Truckenbrodt, Hubert. 1995. *Phonological phrases: Their relation to syntax, focus and prominence*. Massachusetts Institute of Technology. (Tesi di dottorato.)

Truckenbrodt, Hubert. 2002. Upstep and embedded register level. *Phonology* 19. 77–120.

Truckenbrodt, Hubert. 2007. The syntax phonology interface. In de Lacy, Paul (ed.), *The Cambridge Handbook of Phonology*, 435-456. Cambridge: Cambridge University Press

Truckenbrodt, Hubert. 2012. Semantics of intonation. In Maienborn, Claudia & von Heusinger, Klaus & Portner, Paul (eds*.), Semantics: An International Handbook of Natural Language Meaning*, 2039–2969. Berlin: Mouton de Gruyter.

Turk, Alice & Shattuck-Hufnagel, Stefanie. 2007. Phrase–final lengthening in American English. *Journal of Phonetics* 35. 445–472.

Vanrell, Maria del Mar. 2006. A scaling contrast in Majorcan Catalan interrogatives. In Hoffmann & H. Mixdorff (eds), *Proceedings of Speech Prosody 2006*, Dresden, Germany, May 2-5, 2006, 807–810.

Vanrell, Maria del Mar. 2007. A tonal scaling contrast in Majorcan Catalan interrogatives. *Journal of Portuguese Linguistics* 6(1). 147–178.

Verluyten, Sylavin-Paul. 1982. *Investigation on French Prosodics and Metrics*, Université d'Anvers. (Tesi di dottorato.)

Wagner, Petra & Malisz, Zofia & Kopp, Stefan. 2014. Gesture and speech in interaction. *Speech Communication* 57. 209–232.

Wightman, Colin & Shattuck-Hufnagel, Stefanie & Ostendorf, Mary & Price, Patti J. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91(3). 1707–1717.

Xu, Yi. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communicati*on 46(3-4). 220–251.

Questa pubblicazione è stata realizzata utilizzando carta fabbricata nel pieno rispetto dell'ambiente senza l'utilizzo di sostanze nocive e con l'impiego di prodotti ecocompatibili nella fase di stampa e confezione.

Finito di stampare
nel mese di luglio 2021
sestanteinc - Bergamo