Topic and Focus marking in an Italian corpus: some results of algorithmic measurement and structural interpretation

Edoardo LOMBARDI VALLAURI[1], Fabio TAMBURINI[2]

[1]Department of Linguistics, University of Roma Tre, Italy; [2]Department of Linguistics and Oriental Studies, University of Bologna, Italy

lombardi@uniroma3.it, fabio.tamburini@unibo.it

Abstract

The results of an experiment on a corpus of spoken Italian suggest a partly new hypothesis on how the main prominence may be interpreted by speakers in the marking of Information Structure (IS). A "topologic" concept of Prominence can be conceived of, as endowed with the function of demarcation between units, beside and before their culmination and characterization. Much of the process by which speakers interpret the IS of utterances may rest upon this, the specific intonational contours of IS units being probably motivated by other functions. In addition, many real utterances seem not always to signal the distinction between Topic-Focus and Broad Focus clearly, remaining rather underspecified in this respect, with no serious effects on communicative dynamism in the subsequent discourse. Such results, obtained by measuring Prominence as a complex entity (not only intonational in nature) strikingly follow the law of least effort. The used algorithm receives confirmation by the fact that automatic measurements and human evaluations of IS patterns show a very high percent of coincidence.

Keywords: information structure; prominence; corpus; spoken italian.

1. Introduction

Acoustic patterns are used to express Information Structure (IS) in linguistic utterances. Adopting the definitions proposed by Cresti (2000) and Lombardi Vallauri (2009), we assume that the Focus is

> "the part of an utterance which carries illocutionary force and realizes the informational purpose of the utterance itself. The Topic, on the contrary, is the part of an utterance that has no illocutionary force, whose function is to allow the comprehension of the Focus with respect to the discourse".

In the present study, Topic and Focus have been located in utterances from two corpora of spoken Italian, by perceptively evaluating acoustic patterns, applying negation tests, and judging which part(s) of utterances convey illocutionary force and New information (Chafe, 1987; 1992). Only three typologies of IS where examined, namely Broad Focus (extending to the whole utterance), Topic-Focus, and Focus-Appendix (i.e. constructions with a Narrow Focus located to the left of the utterance).

Some studies on the matter directly investigate the relations between IS and phonetic phenomena, while others analyse them through an intermediate, phonological level. (e.g. (Ladd, 1996; Pierrehumbert, 1987) and all studies adopting the ToBI labelling scheme (Beckman, *et al.* 2005)). In this second perspective phonological categories are derived from acoustic parameters, mainly considering intonation, i.e. F0 profiles.

Most studies on Italian belong to the Autosegmental Metrical (AM) paradigm, quite often based on read rather than spontaneous speech, and usually examine (typical) tonal profiles, mainly pitch accents, of assertive utterances looking for a specific kind of pitch accent able to mark focalised segments.

Contrastiveness is marked intonationally in Florentine (Avesani & Vayra, 2004), while in Roman (Frascarelli, 2004) and Neapolitan (D'Imperio, 2002b) different pitch accents depend on Focus breadth. It is still unclear whether such differences are due to diatopic variation or to idiosyncrasies of the ToBI transcription scheme. On the one hand ToBI notation seems unable to account for melodic differences clearly perceived by the speakers: Broad Focus of assertive utterances is represented through the same pitch accent although hearers are able to identify the geographic origin of other speakers on the sole basis of intonation (Marotta, 2008). On the other hand, scholars agree on the identification of edge tones and pitch accents, but not about the classification of pitch accents different in nature (Pitrelli, *et al.*, 1994; Syrdal & McGorg, 2000). Disagreement concerns tonal alignment (D'Imperio, 2002a; Gili Fivela, 2002) and tonal target identification, in particular inside plateaux (where a single maximum or minimum cannot be easily discerned) (D'Imperio, 2002a). Information about scaling (i.e. the frequency range within pitch accents) and slope is underestimated, although potentially distinctive (Gili Fivela, 2002).

As suggested in some classical studies (such as Ladd, 1996) and substantiated in more recent investigations (Breen *et al.*, 2010; Lee & Yu, 2010), a focused item might involve a complex combination of different acoustic cues, namely duration, pitch and intensity, and cannot be analysed only through its intonational profile. For these reasons, we will try to investigate the correlation between focused items and phonetic features by considering the concept of prosodic prominence as a complex and rich set of acoustic features combined in a sophisticated way.

2. Prominence definition and automatic detection

Following a common view, we can define prosodic prominence as *a perceptual phenomenon, continuous in*

its nature, emphasizing segmental units with respect to their surrounding context, and supported by a complex interaction of prosodic and phonetic/acoustic parameters.

Due to its methodological rigour, we will primarily refer to (Kohler, 2005) for a description of the interactions between the different prosodic features that determine the perception of prominence. In his view, there are two main 'actors' playing a relevant role in supporting sentence prominence (or sentence accent). The first, *pitch accent*, concerns specific movements in F0 profile. The second, *force accent*, is independent from intonation and is connected with intensity, segmental durations and possibly other parameters. Both phenomena seem to play relevant roles in supporting prominence perception at utterance level (see also Ladd, 1996), reinforcing each other without establishing specific antagonistic or hierarchical roles.

One of the major challenges in predicting syllable prominence is the disentangling of various sources of influence such as fundamental frequency excursions, duration, intensity related parameters and the listeners' linguistic expectancies. At the acoustic level, various studies (e.g. Heldner, 2003; Sluijter & van Heuven, 1996; Streefkerk, 1996) suggest, also cross-linguistically, the dependence of force accents from unit duration and spectral emphasis (spectral tilt or spectral balance), while pitch accents would be supported by specific F0 configurations and by the global intensity inside a particular segmental unit. One of the authors has carried out experiments confirming such relations for some languages (Tamburini, 2005, 2006).

Assuming this view, we can introduce a prominence function which should be able to assign a continuous prominence level to each syllabic nucleus using only acoustic information:

$$Prom^i = W_{FA} \cdot \left[SpEmph^i_{SPLH-SPL} \cdot dur^i \right] + W_{PA} \cdot \left[en^i_{ov} \cdot \left(A^i_{event}(at_M, at_m) \cdot D^i_{event}(at_M, at_m) \right) \right]$$

where $SpEmph_{SPLH-SPL}$ is the spectral emphasis, dur is the nucleus duration, en_{ov} is the overall energy in the nucleus and A_{event} and D_{event} are the parameters derived from the TILT model (Taylor, 2000) as a function of the maxima alignment type – at_M – and the minima alignment type – at_m. All parameters are referred to the generic syllable nucleus i. See Tamburini (2006) for further details on parameter computation.

The body of the function *Prom* contains nine parameters. Five of them can be considered as supporting the prominence phenomenon from a cross-linguistic point of view ($SpEmph_{SPLH-SPL}$, dur, en_{ov}, A_{event} and D_{event}), while the other four, represented in the vector $\mathbf{W} = (W_{FA}, W_{PA}, at_M, at_m)$, can be seen as language specific. In our model, W_{FA} and W_{PA} weigh the contribution of the two different accent types, while at_M and at_m model the different pitch accent alignments specific for each language.

All the parameters involved in the *Prom*-function computation are normalised inside the utterance (using mean and variance), thus the contributions of different speakers and numeric ranges should be factored out. In all the experiments we used $\mathbf{W} = (1.0, 1.0, 2, 2)$.

3.   Experiments

The two experiments presented here were aimed at searching invariancies in position and level of the Main Prominence, identified through the automatic algorithm presented in the previous section, compared to the IS assigned to the utterances by an expert annotator.

The first experiment is a pilot study on a limited corpus of spoken Roman Italian. The second experiment was aimed to verify the results for the same kind of Italian on a different corpus, and to extend the analysis to two further diatopic varieties, namely Florentine and Neapolitan Italian. The annotator identified the mandatory unit of Focus and possible units of Topic and Appendix, if present. He also determined Focus breadth and possible contrastiveness. We will consider here utterances of 3 classes on the basis of IS: (a) TOPIC | FOCUS; (b) BROAD FOCUS; (c) FOCUS | APPENDIX, NARROW FOCUS, CONTRASTIVE FOCUS. The utterances containing retracting, hesitations and speech disfluencies have been discarded.

**(a) TOPIC	FOCUS**						
Var.-	*Main Prominence on the…*						*No Main Prom*
Corp.	*LsT*	*LsF*	*LsA*	*IsT*	*IsF*	*IsA*	
R–B	18	1	-	0	1	-	3
R–C	12	3	-	1	0	-	3
F–C	24	1	-	0	1	-	7
N–C	8	0	-	2	1	-	2
(b) BROAD FOCUS							
Var.-	*Main Prominence on the…*						*No Main Prom*
Corp.	*LsT*	*LsF*	*LsA*	*IsT*	*IsF*	*IsA*	
R–B	-	4	-	-	0	-	4
R–C	-	4	-	-	6	-	8
F–C	-	3	-	-	3	-	2
N–C	-	4	-	-	7	-	6
**(c) FOCUS	APPENDIX, Narrow F, Contrastive F**							
Var.-	*Main Prominence on the…*						*No Main Prom*
Corp.	*LsT*	*LsF*	*LsA*	*IsT*	*IsF*	*IsA*	
R–B	-	14	0	-	2	0	0
R–C	-	22	1	-	2	0	2
F–C	-	14	1	-	1	0	2
N–C	-	25	0	-	6	0	0

Table 1: Number of utterances divided by Variety-Corpus pairs (R=Rome, F=Florence, N=Naples; B=Bonvino, C=CLIPS) and configurations (e.g. LsT=Last syl. of Topic, IsF=Internal syl. of Focus). Some combination pairs are not possible; in those cases we have inserted a '-' in the corresponding cells
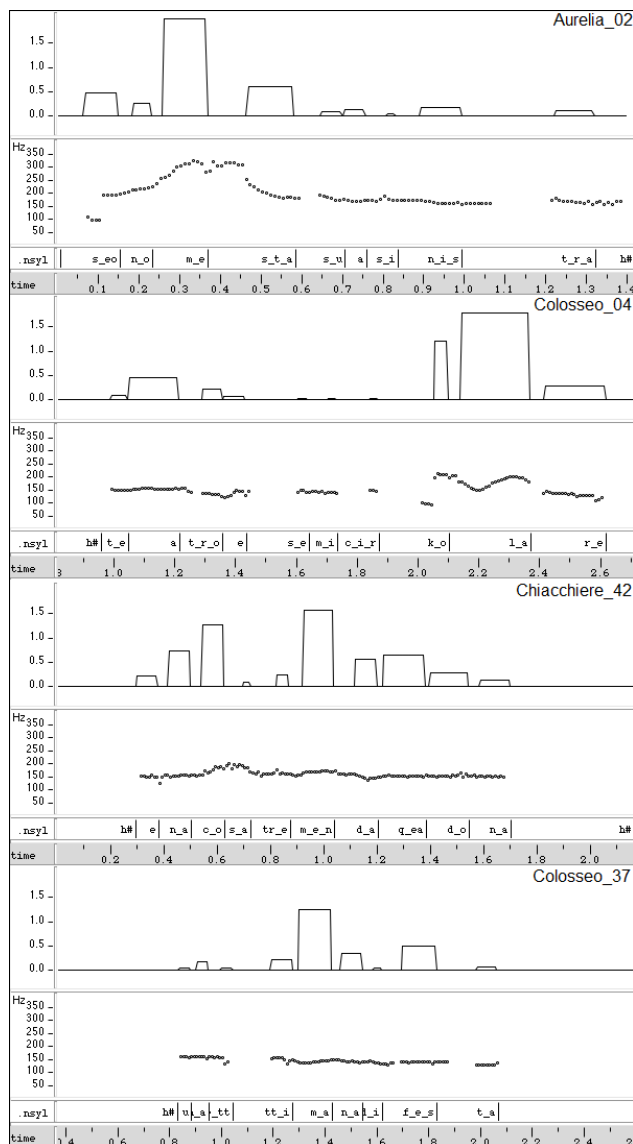
Figure 1: The prominence function profiles – *Prom* – and pitch profiles for some utterances considered in this study.
Aurelia_02: "*Secondo me* T | *stava sulla sinistra* F".
Colosseo_04: "*Il teatro è semicircolare* F".
Chiacchiere_42: "*E' una cosa tremenda* F | *quella donna* A". Colosseo_37: "*Una settimana* F | *di festa* A"

3.1 Experiment 1

The data have been extracted from the "Bonvino" corpus (2005). It consists of 47 utterances selected from 3 out of 12 conversations by speakers from Rome, homogeneous in social level, age, level of education and geographical origin. A reference transcription has been manually added to the extracted waveform to mark the syllabic nuclei needed for prominence identification.

3.2 Experiment 2

The data have been selected from the spoken dialogue sub-corpus of *CLIPS*, stratified through diatopic and diaphasic dimensions (Albano Leoni, 2003). The choice fell on the labeled texts from Rome, to replicate the first experiment using a different data set, Florence and Naples,

so far particularly studied in the AM phonology approach. 184 utterances have been selected: 64 for Rome, 59 for Florence and 61 for Naples.

The results of both experiments, depicted in Table 1, show relevant regularities considering the position of the Main Prominence in relation to the kind of IS. First of all, considering each specific IS, there are no relevant differences between the Italian varieties: the distribution of the Main Prominences seems to follow similar patterns in the different Variety-Corpus pairs. Moreover, the position of the Main Prominence tend to be placed at the border between the two IS components for the TOPIC | FOCUS and the FOCUS | APPENDIX IS, while, in case of BROAD FOCUS utterances, the overall picture seems to be less clear, even if a slight tendency of the Main Prominence to be at the end of the utterance can be found. Figure 1 outlines these regularities for three example utterances from the Bonvino corpus.

It is worth to note that a relevant number of the Main Prominences considered here (e.g. 14 samples out of the 47 from the "Bonvino" corpus) are supported mainly, or uniquely, by force-accents, as shown by the utterance Colosseo_37 in Figure 1, meaning that no intonational phenomena contributed to support them.

These regularities showed to be highly relevant also when testing them by the Fisher exact test.

4. Demarcation rather than culmination

Table 1 shows that the majority of *Topic-Focus* utterances have the Main Prominence at the Right end of the Topic, while a minority seems not to distinguish between the two units, with comparable Prominences. *Left, Narrow Focus* is always marked by a Main Prominence located at the Right of the Focus itself. About half of *Broad Focus* utterances have the Main Prominence at the Right. The other half show several equivalent Prominences.

In other words, where the Main Prominence is regularly associated is the Right end of constituents located at the Left of the utterance. This suggests that its primary function may be *demarcation*, rather than culmination. There would be *a specific function of the Main Prominence bare presence and position*, whose first effect may be to draw a boundary between two information units, rather than "describing" one of them. For the recognition of *which kind* of units they are, it is sufficient that the contour of the one located to the right signals if this is a Focus or an Appendix.

This may explain that Topics are marked more strongly than both Broad Focuses and Right Focuses after a Topic, though the communicative import of Focuses is greater: because Topics are followed by another major Information Unit, so that the boundary between the two needs to be signaled. Narrow Focuses (at the Left) are also strongly marked, in that they are followed by another information unit within the utterance.

The explanation we propose, based only on the *presence and position*, not on the quality of Prominence and intonation contours, is

A topologic hypothesis on main prominence:

"What is marked through the Main Prominence is the boundary between Information Units within the utterance."

Structurally, the only *qualitative* difference strictly needed in order to recognize the IS of an utterance is that between the marking of a Topic and the marking of a Left (Narrow) Focus, because both are followed by another unit. They can be kept apart either by the different intonation contours of the following units (respectively a Right Focus or an Appendix), and (with some redundancy) by the specific intonational contours of the Topic and the Left Focus themselves. The absence of a Main Prominence, or its being located on the last stressed syllable of the utterance, both signal a Broad Focus (not preceded by a Topic), whose boundaries match those of the whole utterance and don't need to be signaled.

Scheme 1 summarizes the minimal steps by which the addressee can "compute" the IS of an utterance.
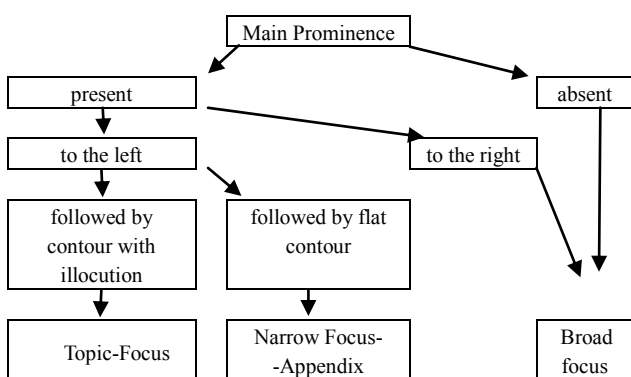
Scheme 1: Steps to the recognition of IS units

If this is true, speakers consistently obey to the *law of least effort*. The only "devices" afforded are (i) one Main Prominence per utterance, and (ii) the difference between a Focus contour and the contour of an Appendix, devoid of illocution. Now, since the different Focus contours are independently needed to express the different linguistic acts, the specific cost required for expressing Information Structure is very low. Culminating each information unit with a specialized Prominence would cost more effort, because distinguishing Topic from Focus would require two different Prominences (one for each) instead of just one at the boundary; and distinguishing Broad Focus from Narrow Focus would require two recognizably different Prominences. As it also happens elsewhere, language prefers to behave economically, marking only the marked element (i.e. Narrow Focus).

5. A continuum

As shown in Table 1, some of the utterances in the corpus that are perceived as Topic-Focus have no Main Prominence. And some of the utterances evaluated as Broad Focuses have an internal Main Prominence, in a position similar to that of Topic-Focus structures.

That utterances acoustically measurable as Broad Focuses can be perceived as Topic-Focus and vice versa, depends on Topic-Focus and Broad Focus being not separate and reciprocally exclusive structures, rather the extremes of a continuum whose center is occupied by utterances with no neat boundary between two units, where the distinction between the two possible ISs remains underspecified. *The speaker is not bound to decide*, at least not prosodically, between Topic-Focus and Broad Focus (possible disambiguation being effected by contextual factors).

In discourse, any content can be focused at different degrees (Daneš, 1974; Firbas, 1989; Sgall *et al*., 1973), or even remain underspecified from this respect. One should always expect for some utterances to have intermediate status between Topic-Focus and Broad Focus, and to contain information, typically "in the middle", with uncertain information status. In sum, Topic vs. Focus seems not to be a black & white story, rather one in a grey scale.

This is the case for the utterances in Figure 2. Topic-Focus and Broad Focus structures do not always need to be clearly distinguished because they are often possible in the same contexts, and compatible with the same development of discourse.

If we add all utterances underspecified between Topic-Focus and Broad Focus to the patterns explained above within the topologic working of Prominence (summarized in Scheme 1), the matchings between previous perceptive evaluations and the results of measurement all belong in one of the following patterns:

Evaluated IS	Measured position of MP
Topic-Focus	MP at Right end of Topic
Focus-Appendix	MP at Right end of Focus
Broad Focus	MP at Right end, or no MP
Topic-Focus or Broad Focus	No evident MP

The cases that fit this model are almost 90% of the total in the corpus, as shown in Table 2.

	corresponding to the description	not corresponding to the description
Rome – Bonvino	43 (91.49%)	4 (8.51%)
Rome – Clips	55 (85.94%)	9 (14.06%)
Florence – Clips	53 (89.83%)	6 (10.17%)
Naples – Clips	53 (86.89%)	8 (13.11%)
TOTAL	170 (87.88%)	28 (12.12%)

Table 2: confirmation of the analysis by acoustic realizations of IS

6. Conclusions

1. The mere location of Prominence may suffice to signal the demarcation between IS units, allowing speakers to interpret the IS of utterances in discourse. From this respect, the specific intonational contours of the different Information Units may represent a certain amount of redundancy.

2. Acoustically, many utterances remain underspecified for the distinction between Topic-Focus and Broad Focus, with no serious effects on subsequent discourse.

3. Such results seem confirmed by the law of least effort, while the used algorithm receives validation by the very high percent of matching between perceptual evaluations and automatic measurement.
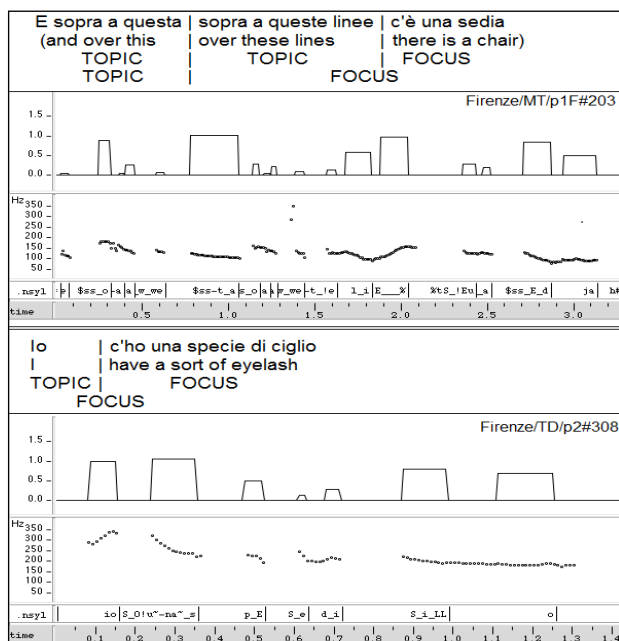

Figure 2: Utterances underspecified between Topic-Focus and Broad Focus

7. References

Albano Leoni, F. (2003). Tre progetti per l'italiano parlato. In *Atti del XXXIV Congr. SLI*, Firenze, pp. 675--683.

Avesani, C., Vayra, M. (2004). Focus ristretto e focus contrastivo in italiano. In F. Albano Leoni, F. Cutugno, M. Pettorino, R. Savy (Eds.), *Il Parlato Italiano. Atti del Convegno Nazionale*, Napoli, pp. 1--20.

Beckman, M.E., Hirshberg, J. and Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S. Jun (Ed.), *Prosodic models and transcription: Towards prosodic typology*, Oxford University Press, pp. 9--54.

Bonvino, E. (2005). *Le sujet postverbal. Une étude sur l'italien parlè*. Paris, Ophrys.

Breen, M., Fedorenko, E., Wagner, M. and Gibson, E. (2010). Acoustic correlates of information structure. In *Language and Cognitive Processes*, 25 (7/8/9): pp. 1044--1098.

Chafe, W. (1987). Cognitive Constraints on Information Flow. In R.S. Tomlin (Ed.), *Coherence and Grounding in Discourse*, Benjamins, pp. 21--51.

Chafe, W. (1992). Information Flow in Speaking and Writing. In P. Downing, S.D. Lima and M. Noonan (Eds.), *The Linguistics of Literacy*, Benjamins, pp. 17--29.

Cresti, E. (2000). *Corpus di italiano parlato*. Firenze, Accademia della Crusca.

Daneš, F. (1974). Functional Sentence Perspective and the Organization of the Text. In F. Daneš (Ed.), *Papers on Functional Sentence Persepctive*. Prague: Academia /The Hague: Mouton, pp. 106--128.

D'Imperio, M. (2002a). Language-specific and universal constraints on tonal alignment: the nature of targets and anchors. In *Proc. Speech Prosody 2002*, pp. 101--106.

D'Imperio, M. (2002b). Italian Intonation: An overview and some questions. In *Probus*, 14(1), pp. 37--69.

Firbas, J. (1989). Degrees of communicative dynamism and degrees of prosodic prominence (weight). In *Brno Studies In English*, 18, pp. 21--66.

Frascarelli, M. (2004). L'interpretazione del Focus e la portata degli operatori sintattici. In F. Albano Leoni, F. Cutugno, M. Pettorino and R. Savy (Eds.), *Il Parlato Italiano. Atti del Convegno Nazionale*, B06, Napoli.

Gili Fivela, B. (2006). Tonal alignment in two Pisa Italian peak accents. In *Proc. of Speech Prosody 2002*, pp. 339--342.

Heldner, M. (2003). On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in swedish. In *Journal of Phonetics*, 31, pp. 39--62.

Kohler, K.J. (2005). Form and Function of Non-Pitch Accents. In *Prosodic Patterns of German Spontaneous Speech*, AIPUK, 35a: pp. 97--123.

Ladd, D.R. (1996). *Intonational Phonology*. Cambridge University Press.

Lee, Y., Xu, Y. (2010). Phonetic Realization of Contrastive Focus in Korean. In *Proc. of Speech Prosody 2010*, Chicago.

Lombardi Vallauri, E. (2009). *La struttura informativa. Forma e funzione negli enunciati linguistici*. Carocci.

Marotta, G. (2008). Phonology or non phonology? That is the question (in intonation). In *Est. de Fonética Exper.*, Univ. Autònoma de Barcelona, XVII, pp. 177--206.

Mertens, P. (1991). Local prominence of acoustic and psychoacoustic functions and perceived stress in french. In *Proc. ICPhS'91*, Aix-en-Provence, pp. 218--221.

Pierrehumbert, J. (1987). *The Phonology and Phonetics of English Intonation*. Indiana Univ. Linguistics Club.

Pitrelli, J.F., Beckman, M.E. and Hirschberg, J. (1994). Evaluation of Prosodic Transcription Labelling Reliability in the ToBI Framework. In *Proc. ICSLP'94*, Yokohama, pp. 123--126.

Sgall, P., Hajicová, E. and Benesová, E. (1973). *Topic, Focus and Generative Semantics*. Kronberg Taunus, Scriptor.

Sluijter, A., van Heuven, V. (1996). Spectral balance as an acoustic correlate of linguistic stress. In *J.Acoustical Society of America*, 100, pp. 2471--2485.

Streefkerk, B. (1996). Prominent accent and pitch movements. *Inst. of Phon. Sciences Proceedings*, University of Amsterdam, 20, pp. 111--119.

Syrdal, A. and McGorg, J. (2000). Inter-transcriber reliability of ToBi prosodic labelling. *In Proc. ICSLP2000*, Bejing, pp. 235--238.

Tamburini, F. (2005). Automatic Prominence Identification and Prosodic Typology. In *Proc. InterSpeech 2005*, Lisbon, pp. 1813--1816.

Tamburini, F. (2006). Reliable Prominence Identification in English Spontaneous Speech. In *Proc. Speech Prosody 2006*, Dresden, PS1-9-19.

Taylor, P.A. (2000). Analysis and Synthesis of Intonation using the Tilt Model. In *J. Acoustical Society of America*, 107, pp. 1697--1714.