

Tonal structure and pitch targets in Italian focus constituents*

Mariapaola D'Imperio

Université de Provence. CNRS UMR 6057

Laboratoire Parole et Langage

29, Av. R. Schuman. 13621 Aix-en-Provence. France

dimperio@lpl.univ-aix.fr

Abstract

This work investigates the tonal structure of the focal accent in narrow focus statements of Neapolitan Italian. The formal properties of this accent lend themselves to two competing analyses. Specifically, this accent could equally be described as a HL accentual fall or as a LH rise. The two analyses were evaluated on the basis of a set of utterances containing focus constituents with varying number of words. Long narrow focus constituents present in fact a medial F0 minimum that appears to be an actual tonal target. Such a target might be part of either a HL or of a LH accent. Tonal as well as timing evidence appear to lend support to the LH hypothesis. An important consequence is that the final fall of statement focus constituents must be analyzed as a tonal event that is separate from the nuclear pitch accent and is analogous to the question final fall.

Key words: Focus, pitch accent, tonal target, alignment, scaling, slope, Italian.

1. Introduction

As in Standard and Palermo Italian (Avesani 1995, Grice 1995a), Neapolitan Italian presents a contrast between the nuclear accent of broad focus declaratives and the one employed in narrow focus declaratives (D'Imperio 1995, D'Imperio 2001). While the first appears to be «downstepped» and acoustically less prominent than a regular prenuclear H* accent (and it is generally described as a H+L*), this is not the case for the narrow focus nuclear accent, which is acoustically and perceptually quite salient. In an earlier study, the narrow focus pitch accent of Neapolitan Italian statements was analyzed as a H*+L (D'Imperio and House 1997), one of the arguments being that the fall is generally completed within the stressed syllable and the F0 peak is quite early (closer to the stressed vowel onset than to its offset). An additional argument for such an analysis was that, for a statement to be perceived as such, the accentual fall needs to span the vocalic portion of the accented syllable.

* This paper was first presented at the 14th International Congress of Phonetic Sciences in San Francisco, USA. The author would like to thank Eva Estebas, Martine Grice and Pilar Prieto for helpful comments and suggestions on an earlier version of this paper.

However, the H*+L analysis appears problematic once various other facts are considered. First, in Neapolitan Italian the narrow focal accent of statements is very similar in shape to the narrow focal accent of polar questions, which has been previously described as a LH sequence (D'Imperio 1995, 1997, 2001). Interestingly, a rise-fall configuration, with a very conspicuous peak, characterizes both pitch accent types. Nevertheless, the timing characteristics of the peak (and rise) differ between the two accents, and appear to be perceptually relevant to the purpose of signaling modality, as shown in D'Imperio and House (1997).

Also, in a previous study, the properties of the focal accent in utterances containing multi-word focus constituents were investigated (D'Imperio 2001). It was noticed that, similarly to questions, statements present an initial rise and a final fall that demarcates the edges of the focus constituent. Specifically, when the constituent in focus consists of a single word, the fall of the rise-fall pattern occurs immediately after the pitch accent rise (Figure 1, upper section). On the other hand, when the focus constituent is made of more than a word, (and two accents are generally produced, one at the beginning and the other at the end of the constituent) statements and questions behave differently. In questions, the falling part of the contour separates from the rise, so that it occurs at the right edge of the narrow focus constituent, creating an intervening high F0 plateau. Statements, on the other hand, present a focus constituent-medial fall at some time after the initial

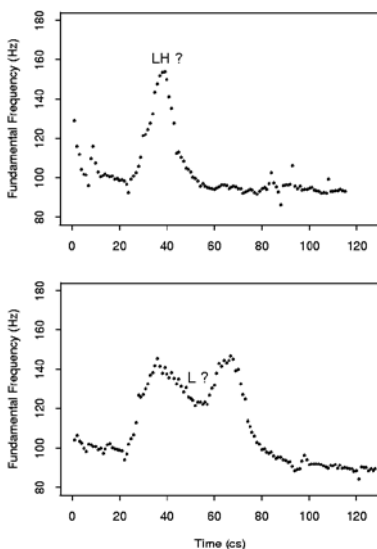


Figure 1. F0 contour of a single-word (upper) and a multi-word (lower) narrow focus statement.

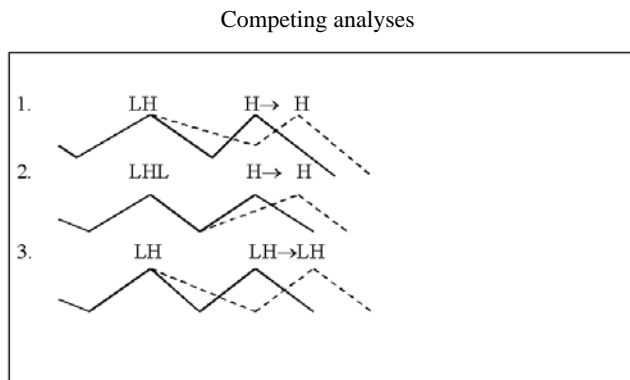


Figure 2. Schematic representation of the 3 predictions of the study.

rise (Figure 1, lower section). As a consequence, narrow focus constituents of statements are characterized by a medial valley, whose nature is explored in this paper. An additional fall is however still realized in the vicinity of the focus constituent right edge, which is much like the final fall of narrow focus question constituents.

The medial F₀ valley can indeed be accounted for by two alternative hypotheses. First, the pitch accent sequence within the focus constituent could be described as a sequence of two H* accents and the medial valley could be a consequence of non-linear or «sagging» interpolation between them. This is the case shown by 1 within the schematic representation of Figure 2. Interpolation between two H targets has been claimed to be of this type in American English (Pierrehumbert 1980), and Italian could behave similarly. Alternatively, the medial valley could be the manifestation of an actual low (L) tonal target. The second hypothesis is shown by both 2 and 3 in Figure 2. Both cases predict that the F₀ valley is an actual target, hence that its melodic and temporal value are constant under certain conditions.

The last hypothesis appears to be supported by the data presented in this study, though offering an additional layer of complexity. In fact, though evidence for a medial L target might be found, it would still be unclear whether such a target would structurally belong to the first or the second pitch accent within the focus constituent. In other words, the medial L could plausibly be either the trailing tone of the first accent or the leading tone of a second pitch accent. Both cases are shown by 2 and 3 in Figure 2. Specifically, case 2 predicts that the target is the trailing tone of a complex LHL pitch accent, which would be supported by the fact that neither the slope from the first H to the medial L would change, nor would the melodic value of the L. On the other hand, case 3 predicts that the medial L is the leading tone of a LH pitch accent, with an invariant slope from the medial L to the second H. After presenting evidence speaking to this issue, consequences for the tonal system of Neapolitan Italian will be discussed.

2. Methods

The corpus consisted of sentences varying in focus constituent size and number of syllables to the right of the nuclear stressed syllable (Table 1). Focus constituents could be made of one, two or three content words. Readings with the required scope of focus were obtained by setting up a felicitous context by means of a previously uttered question. The utterances thus obtained were read 5 times each by 3 speakers of Neapolitan Italian, two males (RM and LD) and one female (MD), with statement intonation, for a total of 135 utterances.

Table 1. Corpus utterances. Scope of focus is indicated by brackets. The stressed syllable within the target word is indicated by capital letters.

CORPUS UTTERANCES	
<u>Focus Constituent Size 1</u>	Vedrai [MA/MAMma/MAMmola] domani. 'You will see mom tomorrow.'
<u>Focus Constituent Size 2</u>	Vedrai [la mano di MA/MAMma/MAMmola] domani. 'You will see mom's hand tomorrow.'
<u>Focus Constituent Size 3</u>	Vedrai [la bella mano di MA/MAMma/MAMmola] domani. 'You will see the beautiful mom's hand tomorrow.'

Various locations in the F0 contours were labeled (Figure 3). Here, results for the following locations are presented:

1. F0 peak (H1).
2. F0 peak (H2) (in two and three-word focus constituents).
3. F0 minimum preceding the first F0 peak (L1).
4. Focus constituent-medial F0 minimum (L2).

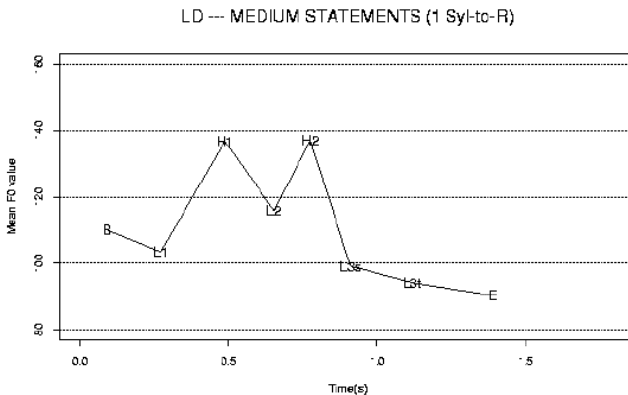


Figure 3. Target values for two-word focus constituent utterances for speaker LD.

The utterances were subsequently digitized at 16 kHz on a SUN Sparc Station and were analyzed with Entropic's ESPS Waves+ at the Phonetics Laboratory of the Department of Linguistics, Ohio State University. F0 peaks were marked at points in which the contour started to rapidly fall, while F0 minima, being more difficult to locate consistently, were calculated through an automatic algorithm (this procedure is illustrated in D'Imperio [2000]) and were hand-checked for errors. Generally, the location of a minimum corresponded to points where the F0 contour started to rapidly rise. Timing and F0 values relative to each of the labeled locations were subsequently obtained.

Slope values for the region falling from H1 to L2 and the one rising between L2 and H2 were also measured. We hypothesized that, if L2 is an actual L target and not the result of sagging interpolation, the slope value obtained for the H1-to-L2 region would be smaller in three-word than two-word constituents. The slope value for the L2-to-H2 region was instead obtained in order to distinguish between the two competing hypotheses arising once evidence is found for the medial L to be an actual target. Specifically, we predicted that if the slope of the L2-to-H2 rise does not significantly vary between two and three-word constituents, this would be evidence for a strict relationship between the L and the subsequent H, such as the one that is typical for tones within a bitonal accent. Such a result would support the idea of a LH nuclear accent, against a HL analysis. In order to further validate this hypothesis, slope values for two and three-word constituents were compared to one-word constituents L1-to-H1 slopes. In fact, we postulated that the nuclear accent of single and multi-word constituents would have the same structure.

Slopes were calculated by fitting linear regression curves to the H1-to-L2, L2-to-H2 and L1-to-H1 F0 values. Because the F0 time series is often noisy, we could not locally estimate the shape of the curve, hence we fitted a model over the entire curve. We first tried to fit a logistic curve, but its asymptotes departed substantially from the F0 minima and maxima. The choice of the linear regression model resulted, instead, in a better fit, and correlation coefficients accounting for the data will be presented. To further test the hypothesis by which L2 is the leading tone of a LH pitch accent, we also measured the temporal distance (latency) of H2 relative to L2. We hypothesized that if the latency between the posited L and the subsequent H were constant across different constituent sizes, this would constitute further evidence for a bitonal LH accent. In fact, it appears that while a starred tone is generally aligned with the associated tone bearing unit (which is the stressed syllable in Italian), the «weak» (leading or trailing) tone precedes or follows it at a fixed temporal distance (Pierrehumbert and Beckman 1988).

3. Results

Figure 4 shows representative examples of regression line fitting to the H1-to-L2 region for two and three-word focus constituents. The results accounted for a high portion of the variability, with values of R^2 spanning from 0.87 to 0.93. Therefore, we believe that the F0 stretch intervening between H1 and L2 can safely be modeled as a straight line connecting the two locations.

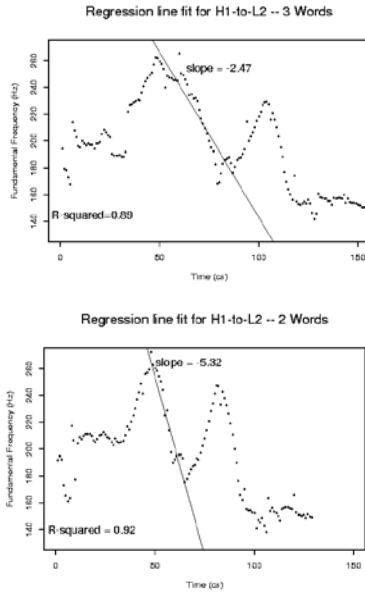


Figure 4. Linear regression curve for F0 values between H1 and L2 for a two-word utterance (upper) and a three-word utterance (lower) for speaker MD.

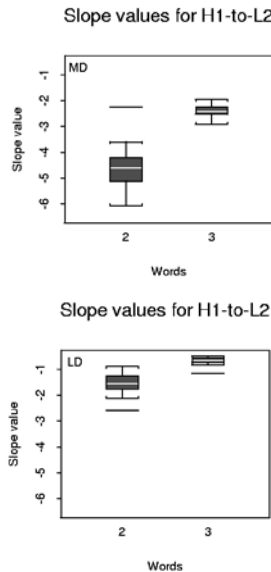


Figure 5. Boxplots for slope values for the H1-to-L2 region in two-word utterances for speakers MD and LD. The white bar within the box represents the median.

Slope values for the H1-to-L2 region for speakers MD and LD are shown in Figure 5. As we can see, slopes were shallower in three-word than in two-word focus constituents (means for two-word constituents: LD = -1.6, MD = -4.6; means for three-word constituents: LD = -0.68, MD = -2.39). The data were tested through a one-way ANOVA for each speaker, with Constituent Size as a factor. Statistic results revealed a significant difference between two and three-word focus constituent slope values within each speaker [LD: $F(1,28) = 46.69$; $p < 0.01$; MD: $F(1,28) = 86.16$; $p < 0.01$; RM: $F(1,28) = 68.2$; $p < 0.01$]. The direction of the difference was the same for all speakers.

The linear regression curves fitted to the L2-to-H2 rising region did also explain most of the variability in terms of a straight line model, with R^2 values ranging from 0.87 to 0.94. Figure 6 clearly shows that slope values for this region are very similar between two and three-word constituents, with mean values for speaker MD equal to 3.53 and 3.28 respectively, and 1.87 and 1.81 for speaker LD. As expected, the ANOVAs conducted on the slope values did not show any significant difference due to Constituent Size for any of the speakers [MD: $F(1,28) = 0.497$; $p = 0.48$; LD: $F(1,28) = 0.301$; $p = 0.587$; RM: $F(1,28) = 0.005$; $p = 0.94$]. When the slope values for L1-to-H1 in one-word constituents were included in the model, results still showed no significant difference for two speakers [MD: $F(1,43) = 3.51$; $p = 0.07$; RM: $F(1,43) = 4.97$; $p = 0.03$], and were only marginally significant for LD [$F(1,43) = 6.81$; $p = 0.01$].

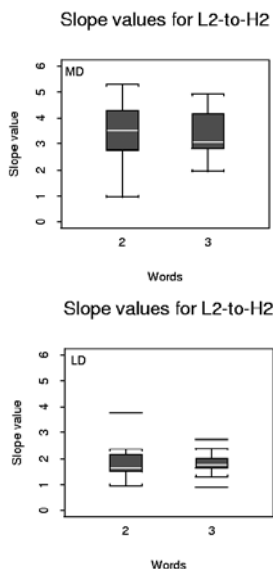


Figure 6. Boxplots for slope values for the H2-to-L2 region in three-word utterances for speakers MD and LD. The white bar within the box represents the median.

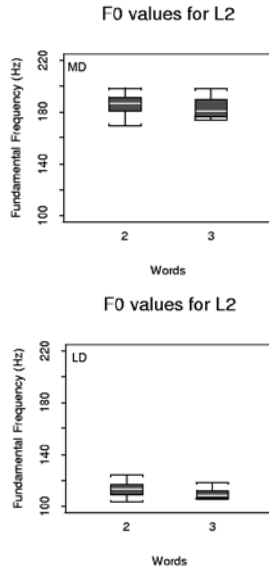


Figure 7. Boxplot for F0 values for L2 in two and three-word utterances for speakers MD and LD. The white bar within the box represents the median.

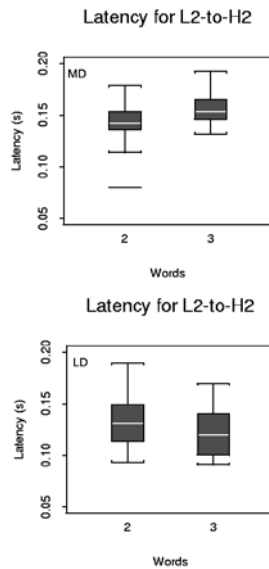


Figure 8. Boxplots for latency values for L2-to-H2 in two and three-word utterances for speakers MD and LD. The white bar within the box represents the median.

To further test the hypothesis that the medial L is a tonal target, we also calculated means for its F0 value in all utterances. As Figure 7 shows, F0 values do not differ much between two and three-word constituents, with an overall mean of 111.9 Hz for speaker LD and of 184.3 for speaker MD. A one-way ANOVA for each speaker, with Constituent Size as a factor, confirmed the absence of reliable differences [MD: $F(1,28) = 0.9$; $p = 0.35$; LD: $F(1,28) = 5.04$; $p = 0.03$; RM: $F(1,28) = 1.89$; $p = 0.18$]. Finally, latency values of H2 relative to L2 showed the expected consistency between two and three-word constituents, as shown in Figure 8. Overall means for latency, independent of Constituent Size, were between 128 ms for speaker LD and 148 ms for speaker MD. A two-way ANOVA conducted on pooled data for all speakers, with Speaker and Constituent Size as factors, showed no significant effect of Constituent Size, nor any reliable interaction, while Speaker was significant [Constituent Size: $F(1,84) = 0.162$; $p = 0.688$; Speaker: $F(1,84) = 10.52$; $p < 0.01$; Interaction: $F(1,84) = 3.02$; $p = 0.05$].

4. Discussion

The present investigation appears to provide evidence for a LH structure of the pitch accent of narrow focus statements. By increasing the number of words within the focus constituent, the slope of the H1-to-L2 contour becomes shallower. This piece of evidence, together with the evidence that F0 values for L2 do not vary with constituent size, points to the existence of a medial L target. Therefore, the hypothesis of «sagging» interpolation between two subsequent H peaks can be discarded. Once the existence of a tonal target was confirmed, evidence that it is the leading tone of a LH accentual rise was shown. Slopes for the nuclear accent rise (L2-to-H2) were found to be constant in both single and multi-word constituents.

This further supports the idea that the same LH accent is the nuclear accent of all the utterances in the corpus. As a further test for the hypothesis of a rising LH accent, it was shown that the temporal distance between L2 and H2 does not change across constituent sizes. This is consistent with a representation of bitonal accents as binary branching trees (Pierrehumbert and Beckman 1988) and is an index of structural unity.

Revising the analysis of the nuclear accent of narrow focus statements from a H+L* fall to a LH was instrumental in deciding starredness status within the LH of polar questions. In fact, from alignment facts alone, it is not clear which of the tones of the bitonal accent is associated to the stressed syllable in question pitch accents (both tones fall within the boundaries of the associated stressed syllable), hence question pitch accents could be equally labeled as L+H* or L*+H. However, once an additional rise is added to the Neapolitan tonal system, contrastive association of the L and the H in the two rises becomes necessary to postulate. By comparing formal properties of the two rises, it appears that it is more appropriate to use the L+H* label for the narrow focus statement rise, while the question accent, whose H target is timed later, should be labeled as L*+H.

As a consequence of the results, a further similarity between the tonal structure of questions and statements was found. Similarly to the complex rise-fall configuration of polar questions, the focus constituent-final fall of statements could be attributed to the presence of a tone similar to the «sentence accent» of Swedish, as proposed in D'Imperio (2001), or to the phrase accent, which has been extensively documented in a number of languages (Grice *et al.* 2000). The phrase accent analysis has been privileged for Neapolitan within a recent review of intonational characteristics of a number of Italian varieties (Grice *et al.*, forthcoming). Independent evidence for the existence of a phrase accent comes also from the analysis of other Italian varieties. Avesani and Vayra (2000), for instance, show evidence for a phrase accent in Tuscan Italian. Specifically, they found that L tone following the H accent peak can be measured at the baseline of the speaker, and is timed at the end of the accented word, hence cannot be part of the pitch accent proper. Also, Gili Fivela (2002) shows that while contrastive accents of Pisa Italian have a final L at a fixed distance from the H peak (and is hence analyzed as part of the pitch accent), broad focus accents are followed by a L tone which occurs at a variable distance from them (and is hence analyzed as a phrase accent).

An alternative analysis would be to consider the fall following LH as part of the nuclear accent, which would necessarily be tritonal (LHL). There are at least two lines of argument that can prompt to avoid such tritonal configurations. Apart from not being frequent in the languages of the world, tritonal accents unnecessarily complicate the theory. In fact, one could predict the existence of a number of possible configurations obtained from permutating tones at the three different positions, with the risk of overgenerating. Even in those theories where such configurations are underlyingly allowed (Grice 1995b), constraints on the surface insure that only two tones per pitch accent be realized.

5. Conclusion

In summary, narrow focus statements of Neapolitan Italian appear to be characterized by a pitch accent that can best be described as a rising LH. On one hand, slope values of the region falling from the first F0 peak to the medial valley within the focus constituent, as well as F0 measurements, point to the existence of a medial low tonal target (L), whose value is independent of constituent size. On the other hand, slopes of the region rising from the medial valley to the second peak, as well as the tight timing relationship between these locations, support the hypothesis that the medial L is the leading tone of a LH pitch accent. These results are instrumental in revealing similarities between the tonal structure of questions and statements. First, the pitch accent of questions appears to be also a LH rise. Second, the focus constituent final fall must be analyzed separately from the nuclear pitch accent and might have a similar demarcative function in both questions and statements.

References

- Avesani, Cinzia (1995). «ToBI: un sistema di trascrizione per l'intonazione italiana». *Atti delle V Giornate di Studio del Gruppo di Fonetica Sperimentale* (A.I.A.). Povo (TN), Italy, pp. 85-98.
- Avesani, Cinzia; Vayra, Mario (2000). «Strutture marcate e non marcate in italiano. Il ruolo dell'intonazione». *Atti delle X Giornate di Studio del Gruppo di Fonetica Sperimentale*. Napoli, Italy, 13-15 Dec. 1999.
- D'Imperio, Mariapaola (1995). «Timing differences between prenuclear and nuclear pitch accents in Italian». *Journal of the Acoustical Society of America* 98(5): 2894.
- (1997). «Breadth of focus, modality and prominence perception in Neapolitan Italian». In: Ainsworth-Darnell, K.; D'Imperio, M. (eds.). *The Ohio State University Working Papers in Linguistics - Papers from the Linguistics Laboratory*, vol. 50, Ohio State University, pp. 19-39.
- (2000). *The Role of Perception in Defining Tonal Targets and their Alignment*. The Ohio State University, doctoral dissertation.
- (2001). «Focus and tonal structure in Neapolitan Italian». *Speech Communication* 33: 339-356.
- D'Imperio, Mariapaola; House, David (1997). «Perception of questions and statements in Neapolitan Italian». *Proceedings of Eurospeech '97*. Rhodes, Greece. Kokkinakis, G., Fakotakis, N.; Dermatas, E. (eds.), vol. 1, pp. 251-254.
- Gili Fivela, Barbara (2002). «Tonal Alignment in Two Pisa Italian Peak Accents». *Proceedings of the 1st International Conference on Speech Prosody (Speech Prosody 2002)*. Bel, Bernard; Marlien, Isabelle (eds.), Aix-en-Provence, France, 11-13 April, 2002, pp. 339-342.
- Grice, Martine (1995a). *The intonation of interrogation in Palermo Italian: implications for intonation theory*. Tübingen: Niemeyer.
- (1995b). «Leading tones and downstep in English». *Phonology* 12: 83-233.
- Grice, Martine; Ladd, Dwight R.; Arvaniti, Amalia (2000). «On the place of phrase accents in intonational phonology». *Phonology* 17:143-185.
- Grice, Martine; D'Imperio, Mariapaola; Savino, Michelina; Avesani, Cinzia (forthcoming). «Strategies for intonation labelling across varieties of Italian». In: Jun, Sun-Ah (ed.). *Prosodic Models and Transcription: Towards Prosodic Typology*. Oxford: Oxford University Press.
- Pierrehumbert, Janet (1980). *The Phonology and Phonetics of English intonation*. MIT, doctoral dissertation.
- Pierrehumbert, Janet; Beckman, Mary (1988). *Japanese Tone Structure*. Cambridge, MA: The MIT Press.