

PROSODIC CONVERGENCE ACROSS VARIETIES OF ITALIAN

Michelina Savino^{1,3}, Francisco Torreira² & Martine Grice³

^{1,3}Dept. of Education, Psychology, Communication, University of Bari, Bari, ITALY

²Dept. of Linguistics, McGill University, Montreal, CANADA

³IfL-Phonetics, University of Cologne, Cologne, GERMANY

michelina.savino@uniba.it, francisco.torreira@mcgill.ca, martine.grice@uni-koeln.de

ABSTRACT

Polar question intonation in Italian is subject to regional variation. Here we investigate how far speakers of one variety (Bari Italian) converge in the production of such questions when playing a game with a partner from a different variety (Lecce Italian). Although these two varieties each have a number of contours in their repertoire, the distribution of these contours differs considerably. Bari Italian uses mainly Rise-Falls and Rise-Fall-Rises, whereas the Lecce variety uses mainly simple Rises. We show that speakers of Bari Italian produce fewer complex contours and more rises when interacting with a Lecce partner. They thus more frequently select the contour that is typical of the partner's variety. By contrast, an analysis of continuous properties of Rises and Rise-Fall-Rises reveals only minimal adjustments to the phonetic properties of each contour. Convergence thus appears to take place predominantly at the contour selection stage.

Keywords: spontaneous imitation, accommodation, question intonation, task-oriented dialogue, Italian.

1. INTRODUCTION

In research into prosody, sound-meaning mapping has been shown to be probabilistic and distribution-based, depending on linguistic and contextual factors [1], [2]. This is true for the intonation of polar questions in Italian, where a number of different contours are available, with highly different distributions across varieties [3]. Here we are concerned with two varieties, those of Bari (BA) and Lecce (LE), that, although they are in the same geographical region (Apulia), have been shown to have very different distributions. Analysis of task-oriented dialogues (map tasks) in [3] revealed that whereas BA speakers overwhelmingly prefer complex contours (rise-falls and rise-fall-rises) and rarely employ simple rises (98.3% vs. 1.7% respectively), LE speakers make use of simple rises most of the time (85.6%).

In this paper we investigate polar question contours, also from task-oriented dialogues, produced by speakers of Bari Italian when speaking to an interlocutor from the same variety as compared to when

speaking to an interlocutor from Lecce. For reference, we also investigate the contours produced by the interlocutors from Lecce. A comparison of these contours and their distributions will tell us how far Bari speakers accommodate their productions according to the provenance of their interlocutor.

Previous studies have shown that speakers of the Bari variety can adjust their question contours towards those of other varieties when explicitly instructed to imitate them, although they tend to produce a blend of the target contour with their own native contour. For example, [4] found that they modified their native peak alignment in question rise-falls when instructed to imitate the rise-fall question contours of a native Neapolitan Italian model speaker, by shifting the peak towards the later Neapolitan contour. As both contours were rise-falls, the imitation involved adjustments to phonetic implementation within this contour shape. In a later study using the same paradigm [5], in which Bari speakers were instructed to imitate questions with the Lecce Italian rising contour, two different adjustments took place. First, they shifted their distributions so as to produce predominantly rises and very few rise-fall(-rise)s. Second, since the Bari inventory contains a further rise conveying non-finality, this contour appeared to influence the phonetic implementation of the imitated rises, such that they appeared to be a blend of the Bari non-final contour and the Lecce question contour.

In the current study, participants took part in a game eliciting spontaneous polar questions. Any imitation of the contours produced by the Bari speakers when interacting with Lecce speakers was not explicitly requested. We explore whether and to what extent Bari speakers converge to the Lecce conversational partners' question intonation by:

- a) adjusting the distribution of the contour types;
- b) adjusting the phonetic implementation of their contours.

2. METHODS AND MATERIALS

2.1. Elicitation method and experimental setting

To elicit polar questions in a spontaneous interactional setting, pairs of speakers were required to play the popular game "Guess who?". In this game, each

participant receives a board with an array of pictures of characters along with a name written underneath. The game starts with each player selecting a card from a separate pile of cards containing the same pictures as the board. The aim of the game is to guess which card the partner selected, by asking questions about characters' features in order to eliminate candidates until only one is left. Since the answers allowed are either "yes" or "no", questions are necessarily polar.

To introduce variation in the lexical composition of the questions, as well as in the game itself, participants were given five different boards, each with a different set of character types. Once one of the two players has successfully guessed three characters on one of the boards, participants started a new game with another board. The winner was the participant who had accumulated the highest score at the end of the session.

In each recording session, participants sat opposite each other at separate desks. They wore an AKG C520 condenser microphone headset connected to a Marantz PMD 661 digital recorder. A cardboard screen between the two desks prevented eye contact during the game. This restricted the interaction to the audio channel and ensured that players could not see each other's boards. A silent experimenter (the first author) was present during the session but without interfering in the game. All sessions were recorded in a quiet room at the University of Bari, and their average duration was 35 min.

2.2. Participants and recording sessions

Twelve participants, six speakers of Bari Italian (henceforth BA) and six speakers of Lecce Italian (henceforth LE) participated in the game sessions on a voluntary basis. They were all born and living in Bari and Lecce metropolitan areas respectively, aged 22-24. They were also female university students and did not know each other. These parameters were kept constant since it has been shown that gender (e.g. [6]) and familiarity (e.g. [7]) can play a role in speech accommodation. Participants were given one exam credit for participating in the experiment.

BA speakers participated in two different recording sessions:

- BA-LE speaker pair session
- BA-BA speaker pair session

The BA-BA game sessions were all recorded three months after the BA-LE counterparts. This delay between recordings was introduced to avoid possible long-term accommodation effects – as a consequence of interaction with a LE partner – on BA speakers' intonation productions during their exchanges with a

same-variety interlocutor. A same-variety dyad condition was not recorded for LE speakers.

2.3. Preparation of speech materials for analysis

All game interactions were orthographically transcribed, and information-seeking polar questions identified based on listening to the recordings, taking the context, including the interlocutor's response, into account. At this stage one of the BA-LE recordings was excluded from further analysis because the spoken accent of the LE partner was judged as not genuinely LE by a native speaker of that variety. Utterances produced with accompanying disfluencies, laughter, or other kind of signal noise generating possible ambiguities in intonation analysis were also discarded.

After this initial screening, 1,439 utterances were included in the analysis: 487 questions produced by BA speakers in the BA-BA sessions, 478 by BA speakers in the BA-LE sessions, and 474 by LE partners in the BA-LE sessions.

3. DATA ANALYSIS

In the speech signal, time intervals corresponding to each question, the nuclear words and component syllables were annotated in Praat [8].

3.1 Phonological analysis of question contours

Polar questions were intonationally annotated by the first author in the Autosegmental-Metrical framework for varieties of Italian [9], [3], [10]. Within this framework, previous work has identified the predominant contour for polar questions in BA as a rising nuclear accent followed by either a low or a rising boundary tone sequence, analysed as L+H*L-L% (a rise-fall) and L+H*L-H% (a rise-fall-rise) respectively. In LE questions, the most widespread F0 contour has low pitch on the accented syllable and a rise at the intonational boundary, analysed as L*L-H% (a rise).

A number of additional polar question contours were identified and sorted as follows into macro-categories. Rises with a fall before the regular rise (transcribed as H+L* L-H%), were treated as a variant of the rise. Rise-falls with a late rise (annotated as L*+H L-L%) were treated as a variant of the regular rise-fall. The few truncated contours on final stressed syllables were treated as simple rises (n=3). The following macro-categories (based on overall F0 shape, as described above) entered into the analysis of the contour distribution: **Rise** (n=637), **Rise-Fall** (n=182), and **Rise-Fall-Rise** (n=620).

3.1.1. Results

Fig. 1 displays the distribution of question contour types produced by all BA speakers across the two conditions (represented by the interlocutor variety in parenthesis) - BA(BA) and BA(LE) - alongside those of their LE conversational partners in the BA-LE sessions - (BA)LE.

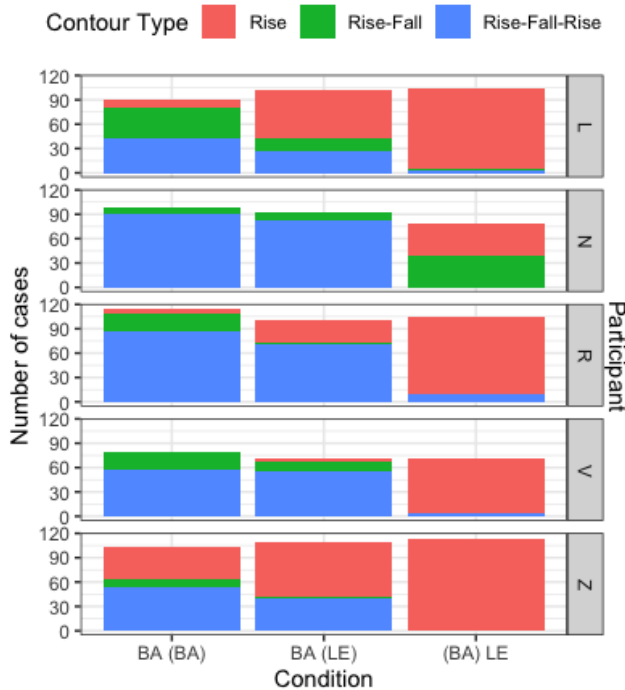


Figure 1: Distribution of contour types for questions produced by all BA speakers in BA(BA) vs BA(LE) sessions, and by all LE speakers in the (BA)LE sessions (BA= Bari Italian, LE=Lecce Italian, on x axis variety of interlocutor in parenthesis). The capital letter on the right indicates the BA speaker in the dyad referred to in the results

For BA speakers (in Fig.1 identified by the letters L, N, R, V, Z), results for the BA(BA) condition confirm the picture provided by past studies in the predominance of Rise-Fall-Rises (68%) and Rise-Falls (20%) over Rises (12%) in polar questions in this variety. For LE speakers, results also confirm the picture from previous studies, with considerably more simple Rises (88%) and very few Rise-Fall-Rises (4%) and Rise-Falls (9%). Crucially, when interacting with a LE interlocutor, BA speakers produced fewer rise-fall-rises (57%) and rise-falls (9%) but more rises (34%). We fitted a mixed-effects logistic regression model to the BA speaker data predicting whether the contour produced was a RISE (i.e., typical of LE) or either of the other two contours (typical of BA). CONDITION was the only fixed predictor (BA(BA) vs. BA(LE)). Random intercepts were fitted for SPEAKER and for the interaction between

SPEAKER and CONDITION. The fitted model predicts Rises to be more likely in the BA(LE) condition than in the baseline BA(BA) condition ($\beta = 2.6$; $z = 2.7$; $p < .01$). This indicates that, when asking questions, BA speakers tended to accommodate intonationally to their LE interlocutors in terms of contour type selection.

Although the distribution of contour types realised by LE speakers in the BA-LE condition shows a predominant use of Rises, roughly corresponding to the distribution reported for LE speakers in native interactional settings [3], one LE speaker (BA speaker N's interlocutor) did make considerable use of Rise-Falls, indicating that this speaker might have accommodated to the BA partner in her contour selection. Interestingly, the BA partner in this dyad is the only one who did not adjust her contour selection, producing no Rises at all.

3.2. Quantitative analysis of question contours

We then examined the phonetic realization of nuclear contours in a subset of the data. In this subset, most combinations of factor levels of interest contained a fair number of observations. This reduced dataset consisted of all Rises and Rise-Fall-Rises featuring a word with penultimate stress spoken in modal voicing ($n = 1335$; 93% of the complete dataset).

We extracted pitch contours from the acoustic data using the autocorrelation method in Praat set to default parameters except for 'floor' (160 Hz for all speakers except one, 150 Hz), 'ceiling' (400 Hz for all speakers), and 'voicing threshold' (0.85). A more conservative voicing threshold allowed us to exclude microprosodic F0 perturbations occurring at the onset and offset of voiced intervals. We then smoothed and interpolated the extracted contours in Praat using default settings. Finally, we transformed the resulting pitch values to semitones (re 100 Hz), and normalized them relative to each speaker's median, as computed from the complete recordings. Fig. 2 shows the data broken down by SPEAKER and CONDITION as a function of CONTOUR. Note that, although the data for BA speakers N and V were not included in the statistical analyses below, their data are shown in Fig. 2.

We tested the effect of CONDITION on the shape of nuclear contours by fitting mixed-effects linear models predicting F0 scaling of the following tonal landmarks: (1) FINAL F0 in both Rises and Rise-Fall-Rises, (2) the accentual F0 PEAK, and the post-accentual F0 VALLEY in Rise-Fall-Rises. To simplify the interpretation of the models, we recoded CONDITION so as to capture the two main differences of interest in our study: (1) between BA speakers when speaking to a LE or a BA conversational

partner, i.e. BA(LE) vs. BA(BA), and (2) between LE speakers and their BA interlocutors, i.e. (BA)LE and BA (LE). Because of model convergence issues, we could not include interactions between CONDITION and SPEAKER in the models (but see discussion of speaker variation below).

For Rises, FINAL F0 was found to be significantly higher for LE speakers than for their BA interlocutors (Contrast 2: $\beta = 1.78$; $t = 2.6$; $p < .05$), but we observed no difference between conditions BA (LE) and BA (BA) (Contrast 1: $\beta = 0.18$; $t = 7.6$; $p = .44$). Separate linear models fitted to each participant's data confirmed these findings at the individual level (note that participants L and V were excluded from this analysis due to lack of data).

For Rise-Fall-Rises, the mixed-effects model fitted to all participants yielded a statistical difference in F0 VALLEY between BA speakers when speaking to a LE or a BA conversational partner. Rise-Fall-Rises reach slightly lower F0 values in the BA (LE) condition relative to the control BA (BA) condition (Contrast 1: $\beta = -0.24$; $t = -3.07$; $p = .005$). Linear models separately fitted to each participant revealed however that this effect was exclusively driven by speaker N, as the contrast (BA(LE) vs BA (BA)) did not reach or approach significance in any of the models fitted to the other participants.

Analysis of individual participants' data further showed that speaker R's F0 PEAK and FINAL F0 were significantly higher when interacting with her LE than her BA partner (F0 PEAK: $\beta = .42$; $t = 3.1$; $p < .005$; FINAL F0: $\beta = -0.7$; $t = 4.52$; $p < .0001$). For BA speakers N and R, therefore, our results point to some phonetic adjustment of the RFR questions contours when exposed to the intonation of their LE conversational partner. Individual analyses also showed, however, that speaker V produces *lower* F0 PEAK ($\beta = -0.54$, $t = -2.6$, $p < 0.01$), and that speaker N produces *lower* FINAL F0 ($\beta = -0.57$, $t = -3.13$, $p < 0.005$) when speaking to a LE speaker than when speaking to BA speaker. Together, these results indicate that if speakers make any phonetic adjustments, they are very subtle.

4. DISCUSSION AND CONCLUSION

Results from our explorative study have shown that, in spontaneous game-based dialogue with a LE conversational partner, BA speakers tend to converge to the LE intonation by shifting their native distribution of the question contour types. In particular, they produce more Rises (the predominant question contour in Lecce Italian) and fewer Rise-Falls and Rise-Fall-Rises (the contours with the highest probability in their native variety) when compared to interactions with a same-variety partner. Results indicate that prosodic accommodation is mainly realised at the contour selection stage, since we observed that distribution of contour type was not necessarily accompanied by an adjustment in the phonetic implementation (e.g. of the final rise in Rises and Rise-Fall-Rises).

Importantly, our preliminary results provide further support to the idea that, in prosody, the sound-meaning mapping is probabilistic and distribution-based [1, 2]. In this view, prosodic convergence can be accounted for as shifts in distributions of individual contour types (composed of pitch accent and edge tone sequences). Bari Italian speakers exposed to Lecce Italian question intonation appear to increase the production of Rises not only when they are explicitly instructed to imitate the Lecce Rises uttered by a model speaker, as reported in [5], but also in cases where they are not instructed to pay attention to intonation at all, as in the current study. However, differences in the task – explicit imitation vs spontaneous interaction – implying different conditions in terms of allocation of attention and memory in imitation [11], [12] lead to a greater proportion of Rises in the imitation task than in the conversational setting.

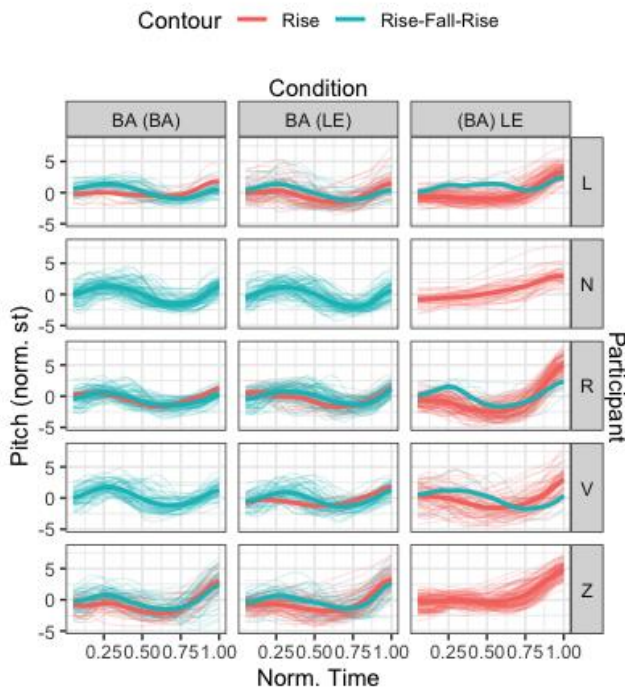


Figure 2: Speaker-normalized pitch (st) as a function of normalized time (1 = duration of nuclear region) for each CONDITION and PARTICIPANT in Rises and Rise-Fall-Rises featuring words with penultimate stress (see text for details).

6. ACKNOWLEDGEMENTS

For funding we are grateful to the German Research Foundation (DFG) – Project-ID 281511265 – SFB 1252 “Prominence in Language” (project A01) at the University of Cologne.

7. REFERENCES

- [1] Calhoun, S. 2010. The centrality of metrical structure in signalling information structure: A probabilistic perspective. *Language* 86(1), 1–42.
- [2] Kurumada, C., Roettger, B. T. 2021. Thinking probabilistically in the study of intonational speech prosody. *WIREs Cognitive Science* 13(1):e1579.
- [3] Savino, M. 2012. The intonation of polar questions in Italian: Where is the rise? *Journal of the International Phonetic Association* 42(1), 23–48.
- [4] D’Imperio, M., Cavone, R., Petrone, C. 2014. Phonetic and phonological imitation of intonation in two varieties of Italian. *Frontiers in Psychology* 5, 1–10.
- [5] Savino, M., Sbranna, S., Ventura, C., Albert, A., Grice, M. 2022. Imitating intonation in a non-native variety: the influence of the native repertoire. *Proceedings of Speech Prosody 2022*, Lisbon, 739–743.
- [6] Pardo, J. 2006. On the phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119 (4), 2382–2993.
- [7] Truong, P. K., Heylen, D. K. J. 2012. Measuring prosodic alignment in cooperative task-based conversations. *Proceedings of Interspeech 2012*, Portland, 843–846.
- [8] Boersma, P., Weenink, D. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5(9/10), 341–345.
- [9] Grice, M., D’Imperio, M., Savino, M., Avesani, C. 2005. Strategies for intonation labelling across varieties of Italian. In Jun, Sun-Ah (ed.), *Prosodic Typology: the Phonology of Intonation and Phrasing*. Oxford University Press, 362–389.
- [10] Gili Fivela, B., Avesani, C., Barone, M., Bocci, G., Crocco, C., D’Imperio, M., Giordano, R., Marotta, G., Savino, M., Sorianello, P. 2015. Intonational phonology of the regional varieties of Italian. In Frota, S., Prieto, P. (eds.), *Intonation in Romance*. Oxford University Press, 140–197.
- [11] Goldinger S.D. 2013. The cognitive basis of spontaneous imitation: evidence from the visual world. *Proceedings of Meetings on Acoustics*, vol. 19, 1–6.
- [12] Heath J. 2017. How automatic is phonetic convergence? In *Proceedings of the Linguistic Society of America* 2-35, 1–10.