L2 Prosodic transfer and priming effects: A quantitative study on semi-
spontaneous dialogues

Giuseppina Turco¹, Michele Gubian²

¹ Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
² Centre for Language & Speech Technology, Radboud University, Nijmegen, The Netherlands
giusy.turco@mpi.nl, m.gubian@let.ru.nl

Abstract
This paper represents a pilot investigation of primed accentuation patterns produced by advanced Dutch speakers of Italian as a second language (L2). Contrastive accent patterns within prepositional phrases were elicited in a semi-
spontaneous dialogue entertained with a confederate native speaker of Italian. The aim of the analysis was to compare learner’s contrastive accentual configurations induced by the confederate speaker’s prime against those produced by Italian and Dutch natives in the same testing conditions. F0 and speech rate data were analysed by applying powerful data-driven techniques available in the Functional Data Analysis statistical framework. Results reveal different accentual configurations in L1 and L2 Italian in response to the confederate’s prime. We conclude that learner’s accentual patterns mirror those ones produced by their L1 control group (prosodic-transfer hypothesis) although the hypothesis of a transient priming effect on learners’ choice of contrastive patterns cannot be completely ruled out.

Index Terms: L2 prosodic transfer, (semi)spontaneous speech, priming, Function Data Analysis, Principal Component Analysis

1. Introduction

In the psycholinguistics literature, studies on associative priming [4] support the hypothesis that contrastive intonation contours induce semantic processing that can lead to priming effects. For instance, a study on L2 priming [5] showed that in perception, L2 listeners processing of intonational meaning depends on the prosodic system of their L1. Another study [6] assessing priming effects in L1 production suggests that even if prosodic representations can be primed, this effect is only short-lived.

This paper investigates whether L1 prosodic-transfer effects and short-lived priming effects coexist in accentuation patterns produced by L2 advanced Dutch speakers of Italian. A confederate speaker was involved in a dialogue-game and had the role of eliciting contrastive phrases from participants. We chose to test learners with Dutch-Italian as L1-L2 language pairs because previous experimental studies [7] support the hypothesis [8,9] that Germanic and Romance languages differ in how prosody is exploited for marking information status. [7] found that Dutch native speakers typically tend to accent new words and deaccent given (‘repeated’) words within syntactic constituents. Such a prosodic-pragmatic relation, by contrast, was not found in (Tuscany) Italian (but see [10]).

Our choice of Italian and Dutch is thus aimed at finding measurable correlates of L1 prosodic transfer in systematic differences in F0 contour shapes realised by the different speaker groups. F0 contours, together with relative speech rate information, were processed using Functional Principal Component Analysis (FPCA), a modern statistical tool available within the framework of Functional Data Analysis (FDA) [19]. FPDA allowed to assess and visualize significative differences in the shape of F0 contours belonging to the different speaker groups. The results of this analysis formed the basis of the discussion that is reported in Sec.4.

FDA is a set of techniques relatively new to the speech research community. The interested reader can refer to lands.let.ru.nl/FDA.

2. Experiment

2.1. Materials
The material consisted of 32 semi-spontaneous utterances per speaker elicited in a picture-difference task, which was designed for other research purposes than the ones tested in this paper (for details, see [11]). All utterances always included the same prepositional phrases (“Nella mia immagine” – “Op mijn plaatje; In my picture”) consisting of a function word (“mia” - “mijn”, FW) followed by a content word (“immagine” - “plaatje”, CW). These phrases were always produced in a contrastive setting where confederate and participants had the role of contrasting each other’s picture. Within such a scenario, we assumed that the contrast (and therefore the new information) would be realized on the adjective “my” rather than on the noun “picture”, even if both words (“mia” “immagine”) were ‘formally’ repeated across the prime and the target phrases. Prepositional phrases were cut out of the utterances with a script and manually transcribed.

2.2. Experimental setting
The picture-difference task is based on a picture comparison in dialogue form between a confederate speaker and the participant, thus allowing for the elicitation of a semi-
spontaneous production. The task was to spot differences across pictures: the confederate had to contrast her picture in relation to a reference baseline picture, always by holding the first turn in each trial; then, the participant took the turn and had to contrast his/her own picture in relation to the confederate’s one. The confederate speaker was a female native speaker of (Rome) Italian (23-years-old) and she was not directly instructed on which intonation contour to use but...
only told to produce very similar intonational realizations throughout all the sessions and for each speaker. A post-experiment analysis of her sentences revealed that she realized the contrast on the prepositional phrases by constantly using a pre-nuclear accent on “mia” and a nuclear falling accent on “immagine” (H* H+L* L% according to ToBI [12]).

2.3. Participants
We collected data from 8 Dutch native speakers (male=2, age average=21.2, sd=1.2) and 8 Italian native speakers (male=4, age average=23.3, sd=2.3) and from 9 Dutch learners of L2 Italian (male=2, age average=43.8, sd=9.7).

The L1 Dutch data-set (N, hereafter) consisted of 231 prepositional phrases; the L1 Italian (I) data-set of 263; the L2 Italian data-set (L) of 218. The confederate data-set (C) consisted of 120 prepositional phrases. Phrases containing ellipsis (e.g. “Nella mia”: On mine) or hesitations were discarded from the analysis.

Language proficiency of learners was classified as intermediate according to a writing assessment test.

3. Data Analysis

3.1. Pre-processing: Forced Alignment and F0 extraction
The prosodic analysis of this work is based on two types of input data. The first data type consists of sampled F0 contours, the second one consists of sequences of phone boundaries. The latter has two purposes, one is to align F0 contours according to the segmental material, the other is to infer information about local speech rate.

F0 contours were computed using the F0 tracker available in the Praat toolkit [13]. A default range of 70-350 Hz for males and 100-500 Hz for females was used. These ranges were adjusted for specific speakers in order to minimize obvious errors such as octave jumps. Values of F0 were then transformed into semitones and the mean value of each contour was subtracted out, in order to minimize vocal tract length effects.

Boundaries between adjacent phones were computed using ASR-based forced alignment. The Italian material consists of repetitions of the phrase “nella mia immagine”. Such material was assigned the broad phonetic transcription /nella mia imadZine/ (SAMPA notation [14]), a slight simplification of the canonical form that was chosen in order to match the acoustic models available in the automatic aligner for Italian. This aligner is based on the SPRAAK ASR toolkit [15] and the models are trained on eight-hours of Italian speech [16]. Similarly, the Dutch material consists of repetitions of “op mijn plaatje”, which was transcribed as /op mEin plat@/. Also the Dutch material was aligned using SPRAAK, the acoustic models are trained on the read speech part of the Corpus of Spoken Dutch (CGN[17]).

3.2. Principal Component Analysis of contours: overview
Principal Component Analysis (PCA) is a classic dimensionality reduction technique that can be applied to a vectorial, i.e. the input data elements are vectors (ordered lists) of numbers. In this work, we applied an extension of PCA that allows data elements to be contours or trajectories. This is called Functional PCA (FPCA) and it is one of the techniques available within the framework of Functional Data Analysis [18], a set of modern statistical tools for the analysis of data in the form of functions, where “function” refers to the mathematical representation of a curve (e.g. a polynomial).

Given a dataset of contours, F0 in our case, FPCA offers a compact description of the different contour shapes that can be found in the dataset. Every input curve is represented by a combination of a small number of principal curves, which are the same for all input curves, but are combined in different proportions. Formally, each input curve f(t) (t refers to time) is represented, or better, approximated by a linear combination of fixed functions. One is the mean m(t), i.e. a curve whose value at any instant t is the arithmetic mean of all input curves at t. The others are the so-called Principal Components (PCs), which are found by the FPCA algorithm solely on the basis of the input dataset and are ordered by explanatory power (in terms of percentage of explained variance). If we limit ourselves to the first two PCs, then each curve f(t) is represented by:

\[ f(t) = m(t) + s_1 \cdot PC1(t) + s_2 \cdot PC2(t) \]

where s1 and s2 are called PC scores and are real numbers that determine the proportion with which PC curves have to be combined in order to reproduce the shape of f(t) as faithfully as possible. Fig.1 illustrates the mechanism for PC1 only, where we see how adding s1·PC1(t) to m(t) produces a better approximation of f(t) than using only m(t).

The form of FPCA output offers the possibility to bind a qualitative description of curve shapes to a numerical counterpart. Since each original input curve is associated to a set of numerical scores, classic statistical tools can be applied to those scores to produce inferences. At the same time, scores have a precise relation to curve shapes by virtue of eq.1, thus any statement on PC scores can be translated into a statement on contour shapes.

In this work we applied FPCA to the entire set of F0 contours extracted from the experimental material described in Sec.2.1/2.3. In this way, FPCA did not use the membership information of each contour to its speaker group I, N, L or C. The FPCA representation of the whole dataset was then used to extract a prototypical F0 contour for each of the four groups based on the values of the first two PC scores (s1, s2). Four distinct clusters were clearly identifiable in the (s1, s2) space, and their centroids where used to construct prototype curves by means of eq.(1). These prototypes became the basis for a qualitative description in terms of pitch accent characterization reported in Sec.4.

3.3. Landmark registration and representation of speech rate
The first step towards the application of FPCA to a set of sampled F0 contours is to represent each contour by a continuous function f(t), which is the required input form for FPCA. Every function has to approximate the shape suggested by its corresponding F0 sample sequence, but does not have to become too much rough or wiggly, because we do not want to include unnecessary detail due to errors of the F0 tracker or to microprosody. This is achieved by applying standard smoothing techniques (B-splines-based smoothing with roughness penalty [18]).
It is customary to analyse F0 contours by referring them to the underlying segmental material, as opposed to absolute time. However, each F0 contour has a different duration, and also each word or syllable is in general pronounced at a different rate across repetitions. To make FPCA work on F0 contours referred to the segmental material, an operation called \textit{landmark registration} is applied. This warps the time axis in such a way that it synchronises the position of a number of segmental boundaries selected by the user. In this way, all F0 contours appear to cross a certain boundary exactly at the same (clock) time, thus making the results of FPCA meaningful for a prosodic analysis [18].

To preserve the relative duration of corresponding segments, the second author proposed a way to recover and integrate this information into FPCA by attaching a corresponding relative speech rate contour to each F0 contour [19].

Since our material includes two different phrases, we had to decide on a common set of comparable segmental boundaries. We placed three internal boundaries as follows: /la | mia | ima | dżine/ for the Italian material (thus cutting the first un stressed syllable /ne/), and /op | mEin | plig | tj@/ for the Dutch material.

### 3.4. Results

The application of FPCA to the entire dataset of F0 contours produced the PC scores distribution plotted in Fig.2. Each point represents the values of s1 and s2 for each F0 contour as in eq.(1), and it is labeled according to the speaker group it belongs to. We note four distinct clusters. This means that the shape characteristics described by the first two PCs, together explaining 54.2% of the variance, strongly correlate with speaker group membership. Since FPCA does not make use of the group membership information, i.e. the labels were added after FPCA was carried out on the entire dataset, the appearance of those distinct group-related clusters in the (s1, s2) space provides evidence that the four speaker groups differ from one another in the way they produce their F0 contours. To verify this, an ANOVA followed by a Tukey HSD test was performed on each of the two complete sets of PC scores separately, with groups I, N, L and C. Results revealed that the means of each group (marked as triangles in Fig.2) are statistically different from each other (p<0.0001) in at least one of the two PC scores, which are not correlated by construction. This result allowed us to plot four prototypical curves and safely discuss their shape traits, given that these traits represent significant differences among F0 contour realisations for each speaker group. Fig.3 shows the four F0 prototypical curves obtained by applying eq.(1) to the four cluster centroids (i.e. substituting s1 and s2 with the centroids coordinates).

The same operation was done for the associated relative speech rate curves in Fig.4. The latter plot reports relative speech rates, i.e. the value 1.0 means average rate, 2.0 means twice as fast as average, or duration two times shorter. Average rates are computed from the durations of matching segments in the alignment described in Sec.3.3. The prototypical curves in Fig.3 and 4 are used in the discussion that follows.
This cannot be teased apart. C and N, the two effects (prosodic given the similarity of the L accent pattern and shape to both smaller (2 st in N) up to the start excursion is excursion of the rise accent pattern configuration fact that reported in Fig.4 in Fig.3 realized by the and th.

Before discussing our results in relation to the learners' data, it is worth talking about the different accentual configurations produced across the mini-dialogues between groups I and C. In I, the peak culminates in the stressed syllable of “immagine” (as indicated in Fig.3, 3rd boundary), whereas C realizes the peak on “mia” (Fig.3, 2nd boundary), followed by a very steep fall on the syllable “-mna” (Fig.3, 3rd boundary). This difference is a clear indication that I and C are using two different pitch accents in signaling the contrast to each other's picture. We can speculate that sequential effects of turn-taking in the dialogue might have caused a different accentual configuration choice across turns. Recall that C always speaks first. Regarding L, we noticed that they behaved very differently from I in the choice of contrastive accent pattern (i.e. the location of the peak is on “mia”, Fig.3). This probably suggests that they have not learnt yet this ‘native-like’ possibility of ‘tuning’ and accommodating their contrastive pattern in relation to a preceding one. This might be due to 1) L1 transfer of prosodic function from L’s mother-tongue, 2) by a transient effect of prosodic priming induced by the C. The first hypothesis is supported by the fact L have chosen the same accentual configuration as the N (i.e. a peak on “mijn”) and that the shallow slope of L’s fall is very similar to the one realized by the natives. The reader should bear in mind that the F0 curve of N is actually steeper than how it is represented in Fig.3, by virtue of the corresponding increase of speech rate reported in Fig.4. The second hypothesis is supported by the fact that L curve is similar to the C one in the choice of the accent pattern configuration (both peaks are on “mia”), in the excursion of the rise (from the 1st to the 2nd landmark the excursion is 4 st for both C and L) and in the shape of the peak up to the start of the fall. By contrast, the excursion in N is smaller (2 st in N) and characterized by a plateau. However, given the similarity of the L accent pattern and shape to both C and N, the two effects (prosodic-transfer and priming) cannot be teased apart.

4. Discussion

Regarding L, we noticed that t effect of prosodic priming induced by the s in relation to a preceding . This probably might have cause . Our findings are in line with previous studies [7, 9].

5. Conclusions

This study accent-patterns in L2 Italian produced by Dutch learners in semi-spontaneous dialogues were compared to control groups and explored by using Functional Data Analysis. Our findings are in line with previous studies [7, 9] on cross-linguistics differences in prominence patterns and reveal that 1) learners differed from L1 natives in how they realized their accent configuration in relation to a previous one in the dialogue; 2) learners’ prototypical curves had features similar to the confederate’s and to the L1 Dutch. However, we could not attribute these similarities to a case of prosodic-transfer or to a priming effect. Future ad-hoc experiments could be designed in order to answer the question of whether prosodic transfer and/or prosodic priming influence L2 learners ability of accommodating their accentual configurations in the complex chain of relations entailed by a dialogue.

5. Acknowledgements

The research of the first author is framed in the ANR-DFG project ‘LANGACROSS’ (DFG Grant) awarded to C.Dimroth.

6. References


