SPEECH AND INFORMATION STRUCTURE
A comment on comment clauses: data from European Portuguese

Aida CARDOSO, Sandra PEREIRA, Sandra ANTUNES, Rita VELOSO
Centro de Linguística da Universidade de Lisboa
Av. Professor Gama Pinto, 2, 1649-003 Lisboa, Portugal
aidacard@gmail.com, spereira@clul.ul.pt, sandra.antunes@clul.ul.pt, rita.veloso@clul.ul.pt

Abstract

In this study, we analyze the prosodic realization of comment clauses in European Portuguese in a corpus of spontaneous speech: the Portuguese C-ORAL-ROM corpus. Focusing our analysis on comment clauses involving the verb ‘dizer’ (‘to say’), our main goal is to find if there is a pattern in the prosodic realization of similar comment clauses. Building on regular patterns found for the prosodic structure of these constructions, we discuss systematic relations between prosody and discourse structure in terms of semantic-pragmatic meaning. Our data evidences some regularities in the behaviour of comment clauses involving the verb ‘dizer’ (‘to say’), but we also found asymmetries between the prosodic realization of comment clauses constructed with different verb forms (the conditional form ‘diria’ – ‘I would say’ – and the subjunctive form ‘digamos’ – ‘let’s say’). We discuss these results considering three main points: (i) the results that have been described for parentheticals (and especially comment clauses) for other languages, (ii) the relation between prosodic structure and scope disambiguation, and (iii) the role of the concept of ‘cline of grammaticalization’ (Dehé & Wichmann, 2010) in the understanding of the status of comment clauses in the informational structure of the sentence.

Keywords: comment clauses; parentheticals; prosody-discourse interface; European Portuguese.

1. Introduction

Recently, parentheticals have been receiving a special attention in literature and have been studied from different perspectives. Nevertheless, establishing a typology of parenthetical structures or even describing its features can be challenging. One of the reasons is the fact that the designation ‘parenthetical’ covers a wide variety of structures that are heterogeneous in their nature.

Despite the complexity of the topic, many recent studies are relevant contribute towards a better understanding of the syntactic, semantic, prosodic and pragmatic features of parentheticals (e.g., Dehé & Kavalova, 2007). Moreover, as has been proved by the perspective adopted in several studies, parentheticals provide a very interesting subject for interface studies.

In this paper we focus our attention in a particular type of parentheticals – comment clauses (CC). Specifically we describe data from European Portuguese obtained from the prosodic analysis of CCs formed by verb ‘dizer’ (‘to say’) in a corpus of spontaneous speech. The discussion of the results of our prosodic analysis will take into account the relation between prosody and discourse. Our data allow us to indentify some patterns in the prosodic realization of the CCs analyzed and, thus, present some hypothesis regarding the relation between prosody and semantic-pragmatics, specifically in terms of scope disambiguation and grammaticalization.

2. Theoretical Background

Parentheticals have been traditionally described, considering the relation between syntax and prosody, as having some specific characteristics regarding phrasing and intonation, namely that they are separated by pauses from the rest of the utterance (e.g., Nespor & Vogel, 1986; Frota, 2000) and that they are most commonly produced with a lower pitch than the rest of the utterance (e.g., Crystal, 1969; Bolinger, 1989). Authors such as Wichmann (2000), Dehé (2007, 2009), Dehé & Wichmann (2010), on the contrary, argue that there is no one-to-one relation between syntax and prosody and present data (in particular data from spontaneous speech), showing that parentheticals are not obligatory set off by pauses and that they can be associated with different intonation contours.

In the case of European Portuguese, a few studies have described some prosodic features of parentheticals. Frota (2000, in press) describes parenthetical clauses as forming a major intonational phrase (set off by pauses) independent from the rest of the utterance. The author also indicates that parentheticals are associated with the intonation contour L*+H H%. In a study specifically about vocatives, Abalada, Cabarrão & Cardoso (2011) argue that these parenthetical elements do not always form a major intonation phrase and that both the phrasing and the intonation reflect a close relation between syntactic distribution, pragmatic value and prosodic realization of the vocatives. For example, the authors observed that initial vocatives had a stronger tendency to form major intonational phrases than the non-initial (media or final) vocatives and that there were differences in the intonation contours associated with initial and non-initial vocatives.

Regarding CCs, they are often analyzed grouped with other elements, and, accordingly, their characterization is made on a par with other types of parentheticals. Therefore, the prosodic features referred above have been applied to CCs as well. Moreover, the definition of CCs presents some challenges, since it is not always clear where to draw a boundary between them and other parentheticals, such as discourse markers or reporting verbs, as pointed out by Kaltenböck (2007) and Dehé (2009). Both authors present definitions of CCs based on syntactic and semantic criteria: the former identifies CCs with “asynthetic clauses (…) linked to the host in that they contain a syntactic gap (typically the complement of the verb) which is filled conceptually by
the host clause” (Kaltenböck, 2007: 4) and the latter defines CCs as consisting “of a first-person pronoun and a verb of knowledge, belief or conjecture or a corresponding adjectival construction” (Dehé, 2009: 14).

Furthermore, in what concerns the prosodic features of CCs and the prosody-pragmatics relations, the results discussed in studies such as Peters (2006), Kaltenböck (2007), Dehé (2007, 2009), and Dehé & Wichmann (2010) show that CCs tend to not form a major intonational phrase, being accentuated or not. In fact, these authors mention that there are several factors that can influence the prosodic phrasing of these elements, namely the length, the syntactic complexity and even the semantic-pragmatic scope of the parenthetical element.

Secondly, CCs seem to be associated to various intonation contours. Lowered pitch, higher pitch and rising contours are some of the prosodic realizations of parentheticals described by authors as Bolinger (1989), Wichmann (2000), Dehé (2009), Dehé & Wichmann (2010).

Finally, it is important to mention that Kaltenböck (2007) and Dehé & Wichmann (2010) take into account the interface between prosody and semantic-pragmatics meaning in their analysis. Kaltenböck (2007) focuses on the role of prosody in the disambiguation of the semantic-pragmatic scope of the CCs. In this context, the level of juncture between the CC and the sentence is a key factor to determine the scope of the first one and to decide whether the scope of a CC is clausal or phrasal. On the other hand, Dehé & Wichmann (2010) propose an analysis of ‘cline of grammaticalisation’, where the prosodic properties of CCs, along with their semantic-pragmatic status, place CCs in a continuum between ‘propositional’ and ‘formu- laic’ meaning. Hence, the authors argue that prosodic separation and prominence are indicators of CCs with a ‘propositional meaning’, but that CCs associated with disfluency and hesitations have more of a ‘formu- laic meaning’. In an intermediate position of this continuum, we can find CCs with prosodic integration and deaccentuation, which have “discursal, interactional and interpersonal purposes” (Dehé & Wichmann, 2010: 39).

3. Methodology

For this study, we analyzed the Portuguese C-ORAL-ROM corpus (Bacelar do Nascimento et al., 2005), a multimedia corpus of spontaneous spoken speech, in a total of approximately 300,000 words. This spoken corpus represents real communication acts collected among sociolinguistically diverse speakers and it is composed by 153 recordings, in a total of 30 hours. Each text/recording comprises: (i) the acoustic source; (ii) the orthographic transcription in CHAT1 format and enriched with the tagging of terminal and non terminal prosodic breaks, and (iii) session metadata containing essential information of speakers, recording situation and contents of each session; (iv) text-to-sound synchronization, based on the alignment with the acoustic source of each transcribed utterance; (v) a second orthographic transcription with lemma and PoS tags of each form in the transcribed texts, and (vi) frequency lists of forms and lemmas.

This corpus is constituted by different types of informal and formal speech acts, as shown in Table 1, below.

<table>
<thead>
<tr>
<th>INFORMAL REGISTER</th>
<th>Conversations</th>
<th>Dialogs</th>
<th>Monologs</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family / Private</td>
<td>24,449</td>
<td>62,738</td>
<td>46,005</td>
<td>133,192</td>
</tr>
<tr>
<td>Public</td>
<td>1,817</td>
<td>23,119</td>
<td>7,710</td>
<td>32,646</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>26,266</strong></td>
<td><strong>65,857</strong></td>
<td><strong>53,715</strong></td>
<td><strong>145,838</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FORMAL REGISTER</th>
<th>Conversations</th>
<th>Dialogs</th>
<th>Monologs</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural</td>
<td>10,215</td>
<td>9,750</td>
<td>8,923</td>
<td>66,274</td>
</tr>
<tr>
<td>Context</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interviews</td>
<td>14,570</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>News</td>
<td>1,859</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reportages</td>
<td>10,762</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scientific</td>
<td>9,923</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Press</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sport</td>
<td>5,676</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talk Shows</td>
<td>17,396</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weather</td>
<td>1,930</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forecast</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monologs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explanation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preaching</td>
<td>6,127</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Political</td>
<td>8,649</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teaching</td>
<td>9,822</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>24,365</strong></td>
<td></td>
<td></td>
<td><strong>152,755</strong></td>
</tr>
</tbody>
</table>

Table 1: Portuguese C-ORAL-ROM corpus constitution

In order to extract our sample of CCs from this corpus, we adopted a definition of CC along the same lines as what has been described in the literature referred above (Kaltenböck, 2007; Dehé, 2009). Then, we selected a sample of 30 occurrences of CCs involving the verb ‘dizer’ (‘to say’), namely the forms ‘diria’ (‘I would say’) – 1st person singular of the conditional – and ‘digamos’ (‘let’s say’) – 2nd person plural of the subjunctive present. This sample includes 26 CCs in interpolated contexts and 4 in final contexts. In what concerns the number of syllables, it must be mentioned that the CCs have a minimum of 3 syllables and a maximum of 6 syllables. This variation in the number of syllables is related with some slight differences in the composition of the CCs analyzed. Hence, it is worth noting that, in the case of the 1st person singular of the conditional form, the CCs can be formed: (i) by the verb form – ‘diria’ –, since European

---

Portuguese is a null subject language; or (ii) by the verb form plus the 1st person singular of the personal pronoun – ‘eu’ (‘I’), or (iii) by the verb form, the 1st person singular of the personal pronoun, and the adverb ‘quase’ (‘almost’), as in ‘quase diria eu’ (‘I would almost say’). In the case of the 1st person plural of the subjunctive present, on the other hand, the CCs included in our sample are formed either by the verb form only – ‘digamos’ – or by the verb form followed by the adverb ‘assim’ (‘this way’), as in ‘digamos assim’ (‘let’s say it this way’).

Regarding the prosodic annotation, we used Praat (Boersma & Weenink, 2009) and our analysis focused on two aspects: (i) the break indices on the left and right boundaries of the CCs, and (ii) the nuclear pitch accent and boundary tone of each CC.

In the annotation of our data, we adopted an autosegmental perspective, accordingly with what is described in Pierrehumbert & Hirschberg (1990) and Beckman, Hirschberg & Shattuck-Hufnagel (2005). Hence, we followed the conventions defined by Viana et al. (2007) in the annotation system Towards a P_ToBI and took into consideration their description of pitch accents and boundary tones for EP. In what concerns break indices, we annotated the juncture level between the CC and the sentence using the break index values described in ToBI (Beckman et al., 2005) – 0, 1, 3, 4 – in which 0 represents the maximum level of juncture between words, 1 represents a normal level of cohesion inside of a prosodic constituent, 3 represents a minor intonational phrase boundary (in EP), and 4 represents a major intonational phrase boundary.

### 4. Data

Regarding the data, our analysis reveals important regularities in the prosodic realization of the sample of CCs considered in this study.

Firstly, it is worth discussing the level of juncture between the CCs and the utterance. Hence, the data revealed that the analyzed CCs do not tend to form a major intonational phrase, since only 10% of the totality of our CCd formed a major intonational phrase independent from the sentence. These results enable us to compare our data with some findings reported for other languages: the fact that a syntactic parenthesis does not obligatorily correspond to a prosodic parenthesis points to the non existence of a one-to-one relation between syntax and prosody, as has been stated before by Dehé (2007, 2009) or Dehé & Wichmann (2010). Furthermore, and taking into account the number of syllables of the CCs, we hypothesized that variables like the length of the parenthetical also play a role in the prosodic phrasing of the CCs analyzed in this study, in the same line as what is argued by Peters (2006) and Dehé (2009).

Additionally, our data can be related with the results found for vocatives in European Portuguese (Abalada et al., 2011), in terms of prosodic integration, in the sense that, despite of having a different nature than CCs, vocatives are also short parenthetical elements and do not always form a major intonational phrase, especially vocatives in medial and final position.

Nevertheless, we did find a high percentage of CCs that form a minor intonational phrase (73.3%), which suggests that, although CCs are more likely to not form an independent tone unit, this does not necessarily translates in a total prosodic integration of the CC in relation with the host sentence. In fact, we observed, particularly in what concerns the CCs formed by the conditional form (‘diria’), some differences in the strength of the break index on the left and right boundaries. As Kaltenböck (2007) remarked, the level of juncture between the utterance and the CC can be related to informational structure, which may represent a clue to identify the semantic-pragmatic scope of the CC. In our data, we also noticed that the phrasing differences referred to above can be related with the fact that the CC has a clausal or phrasal scope. Example (1) illustrates a case in which the phrasing evidences that the CC has a clausal scope, and not a phrasal one, since the break index on the left boundary of the CC (‘eu diria’) is stronger – [4] – than the one identified on the right boundary – [3].


(The other three evangelists [4] I would say [3] have such evident and unique features (...).)

By contrast, CCs formed by the subjunctive verb form (‘digamos’) evidence a greater level of juncture in relation with the utterance and, significantly, it is on the right boundary of these CCs that we find a higher frequency of break indices of level 0 and 1.

Similarly to what has been described for phrasing, there are also some relevant aspects regarding intonation that provide some clues to a better understanding of the prosodic behavior of the two types of CCs analyzed. First of all, is should be mentioned that there is a high percentage of CCs (86.6%) that are accented. Nevertheless, this percentage is higher in the case of CCs with the conditional verbal form ‘diria’. In fact, 18.8% of the CCs formed by the subjunctive form ‘digamos’ are un-accented (as shown in Table 2).

Regarding the distribution of pitch accents (cf. Table 2), we identified five pitch accents associated with the CCs included in our data. The fact that these parenthetical elements are characterized by various pitch accents allows us to draw a comparison between our data and what has been stated for other languages, namely English, by authors such as Wichmann (2000), Dehé (2009b), Dehé & Wichmann (2010). In spite of the importance of the non-existence of a obligatory association of these parenthetical elements to a certain intonation contour, it is also relevant that, considering both types of CCs (‘diria’ and ‘digamos’), there is a higher percentage of CCs associated with low pitch accents (L*), followed by the rising pitch accent L+H* and by the high pitch accents (H*).
In what concerns boundary tones, Table 3 shows that in both types of CCs we found a higher percentage of low boundary tones, but the subjunctive form ‘digamos’ has a higher percentage of cases with no boundary tone, accordingly to what has been previously discussed about the prosodic integration of CCs with this verb form.

Table 3: Distribution of boundary tones

<table>
<thead>
<tr>
<th>Boundary Tones</th>
<th>‘diria’ (‘I would say’)</th>
<th>‘digamos’ (‘let’s say’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H- / H%</td>
<td>35,7%</td>
<td>25%</td>
</tr>
<tr>
<td>L- / L%</td>
<td>57,1%</td>
<td>43,8%</td>
</tr>
<tr>
<td>No boundary tone</td>
<td>7,1%</td>
<td>31,3%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

We think that the results presented above can be interpreted along the lines of what Dehé & Wichmann (2010) have described as ‘cline of grammaticalization’. On the other hand, the fact that the prosodic realization of CCs can play a role in scope disambiguation and that CCs do not evidence a tendency to total prosodic integration seems to indicate that the CCs included in our sample do not have a ‘formulaic meaning’. On the other hand, we found differences between CCs with two different forms of the verb ‘dizer’. As a result, some of the prosodic characteristics of the subjunctive form ‘digamos’ contrast with what can be observed for the conditional form ‘diria’: (i) the former does not seem to play such an important role in scope disambiguation as the latter; (ii) the subjunctive form shows a greater tendency for prosodic integration; (iii) there is a higher percentage of low pitch accents associated with the subjunctive verb form, and (iii) there is a higher percentage of un-accented occurrences of CCs with the subjunctive form. Considering these results, we hypothesize that the two types of CCs are in different stages of a grammaticalization continuum. Hence, whereas CCs with the conditional form seem to have more of a propositional meaning, CCs with the subjunctive form are possibly closer to an intermediate stage between propositional and formulaic meaning, characterized pragmatically as having “discursal, interactional and interpersonal purposes” (Dehé & Wichmann, 2010: 39), and prosodically by prosodic integration and deaccentuation.

5. Conclusion

The results discussed in this paper are a starting point to the study of CCs in European Portuguese. By studying a sample of CCs formed by the same verb – ‘dizer’ (‘to say’) – we were able to detect patterns in the prosodic realization of these parenthetical elements.

The fact that CCs do not always form an independent tonal unit and that they are not obligatory associated with a single intonation contour is in agreement with the idea that (i) syntactic parenthesis do.
not necessarily correspond to prosodic parenthetical, as argued by Dehé (2007), and (ii) parenthetical elements can have intonation contours other than a lowered pitch accent, as have been shown in studies such as Wichmann (2000), Dehé (2009), and Dehé & Wichmann (2010).

On the other hand, we also found some asymmetries in the prosodic behaviour of CCs with different verb forms, namely the conditional form ‘diria’ (‘I would say’) and the subjunctive form ‘digamos’ (‘let’s say’). We interpreted such asymmetries in relation with CCs’ semantic-pragmatic meaning, in terms of scope disambiguation and grammaticalization. More specifically, our data suggested that the conditional verb form evidences more features associated with a propositional meaning than the subjunctive verb form.

6. Acknowledgements

This work was undertaken and supported by the funding of the project METANET4U (grant agreement no.: 270893), co-funded by the 7th Framework Programme and the ICT Policy Support Programme of the European Commission.

7. References


Emanuela CRESTI  
LABLITA – Dipartimento di Italianistica  
P.za Savonarola, 1 – Firenze  
elicresti@unifi.it

Abstract

This paper presents a new semantic definition of Focus within the LAcT’s assumptions and on the basis of corpus evidence. We discuss current theoretical frames based on the concept of Common Ground, such as Alternative Semantics, following which Focus is given a semantic definition inside a question-answer model depending on the context. According to the LAcT’s pragmatic interpretation of the information structure, the information unit (IU) of Comment, which is devoted to the accomplishment of the illocution, is considered its centre. Comment, as a pragmatic entity, and Focus, as a semantic entity, must be distinguished given that their respective levels of activity are illocution and locution in Austinian terms. Within the locutionary act, Focus is a semantic entity signalling the apex of a domain functioning in the illocutionary act such as an IU. Moreover, we claim the existence of two different semantic values. We derive this from the condition that a Focus must be marked by a prosodic prominence, and from corpus observations showing that a Topic-Comment information pattern is necessarily performed with two prosodic prominences.

Keywords: information structure; pragmatics; comment; focus; prosody; corpus data.

1. The pragmatic nature of Comment

In the framework of the Language into Act Theory (LAcT) the information structure of the utterance is pragmatically based. An utterance can be compounded of many information units (IU), developing different information functions and each must be identified by a prosodic unit (PU). Alternatively, it can be simple i.e. compounded of only one IU, necessary and sufficient, named Comment, whose specific function is the accomplishment of the illocutionary force (Austin, 1962). According to LAcT the starting point of the information patterning (IP) is the accomplishment of the illocutionary force by the specific Comment IU, because the action of the speaker, given its affective nature, has a subjective and internal origin and is continuously changing. In accordance, the information about the type of action the speaker will utter is expected as necessary and unforeseeable. This perspective on the structure of information departs from the track of traditional assumptions in relation to one particular feature: the pragmatic origin of information, which they do not consider, ignoring its illocutionary definition. Generally speaking, they also share two other aspects diverging from LAcT: the semantic nature of Focus, which is substantially identified on the basis of its novelty with respect to context and represents the only key to explaining the information structure, and the fact that Topic derives from the context. In that way the entire information organization of the utterance results is conditioned by the context.

The reason for these differences is that no distinction is foreseen between different activities (illocution and locution) accomplished by the speaker simultaneously (Austin, 1962), but which diverge in their nature (affective/pragmatic vs cognitive). Given the lack of the illocutionary notion of Comment, in the literature there is no distinction between the semantic concept of Focus and the pragmatic one of Comment, and Comment and Focus are employed like terminological variants. But Comment develops a pragmatic role and cannot be defined in semantic terms, such as is done for Focus. Traditional semantic definitions foresee that Focus represents the “most important” or “new” information in an utterance. However, importance is a vague aspect and can hardly be verified. As regards the feature of novelty, on the basis of our corpus data we have already shown that a Comment can record old semantic content from a contextual point of view (becoming “new” for the illocutionary accomplishment), and that a Topic can record new semantic content (Cresti & Moneglia, 2010). If a Topic can be new and a Comment can be old, are importance and novelty opposing values? These questions don’t seem to have clear solutions.

We will discuss current theoretical frames based on the concept of Common Ground, such as Alternative Semantics, and assuming that Focus must be prosodically marked we will also argue against the perspective of...
Contrastive Focus on the basis both of our pragmatic framework and corpus data evidence. We will propose the existence of two Foci within an utterance corresponding to a Topic-Comment IP: a Topic-Focus and a Comment-Focus with a new semantic definition.

2. The model of Common Ground

Some acknowledged research on information structure employs the concept of Common Ground (CG) in the place of context. The concept was formulated by Stalnaker (1974) and can be described as «a way to model the information that is mutually known to be shared, which is continuously modified in the course of communication». But if a pragmatic perspective is adopted, no “mutually shared information” can exist. The fact that the context is real does not mean that it is an independent entity, knowable in its whole as a logic universe. Everybody knows it subjectively, following his mood and giving attention to what is interesting for his own attitude in that moment. There are no mandatory information prominences in the context, but only those inputs which are prominent for the speaker’s attention in that moment. Moreover, there is no determination from contextual inputs to the performance of a specific illocutionary speech act, because of the internal affective and mental origin of the latter. The speaker’s next speech act is unforeseeable despite every kind of contextual prominence. Mutually shared information could exist only in a platonically semantic or logic context existing outside of the speakers and in spite of their living actions.

In some sense a more concrete definition of Focus seems to be given within the framework of Alternative Semantics (Rooth, 1992; Krifka, 2006).

“Focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions. […] This distinction is relevant for information packaging, as the CG changes continuously, and information has to be packaged corresponding to the CG at the point at which it is uttered”.

This assumption could seem reasonable and of use, but the claim that information can be packaged «corresponding to the CG at the point at which it is uttered» seems to lead again to a semantic dependence of the information structure on the context. It means that some specific objective features in the context, identifying a point in the CG, condition Focus.

Advancing along these lines, Krifka explicates that the prominent use of Focus is the identification of context-questions in answers:

“The idea is that the meaning of a question identifies a set of alternative propositions, the answer picks out one of these, the Focus within the answer signals the alternative propositions inherent in the question”.

In substance, following Alternative Semantics, the core of an assertion i.e. the part adding a novelty to the CG, should be the answer chosen by the speaker among the possible ones, given a certain open question in the CG, that may be optionally reported in the theme/Topic.

A relevant extension of the question-answer model is due to theories assuming that a coherent discourse is structured by implicit questions (van Kuppevelt, 1994; Büring, 2003) and by Focus on the answers. The concept of implicit questions foresees that context is characterized by features that by themselves can constitute or suggest questions for the addressee. In this sense the activity of speech is reduced to answering in a coherent way the questions suggested by the world and the operation could be reduced to a logic schema. The semantic question-answer model transforms the context into an open variable and assumes its satisfaction in the answer, ensuring a result which is characterized by a propositional form. No pragmatic value of the utterance is even hypothesized and this ends the claim of equivalence between utterance and proposition and the allowance of the analysis of the former in the semantic terms of the latter.

However, corpus data supports the fact that real spontaneous spoken activity does not occur in this way. The framework of Alternative semantics, defining the pragmatic use of Focus as the point marking an alternative in an answer to an overt or covert CG question, does not seem adequate for explaining corpus data. Analyzing a stretch of spontaneous dialogue will demonstrate as impossible the continuous discovery of elements for consideration as the origin of covert questions in the CG, so that they are adequate input for speech behavior (Cresti, 2012). The normal manner of human spoken communication is about the context but has its origin in the speakers’ thoughts and in the affective dynamics among these speakers. They are not determined by the context and they continue on with subjective actions and reactions.

For instance, what could be the context-question generating the accomplishment of a specific illocutionary act like an alternative question, or an instruction, or an expression of obviousness? Given that at least 40% of the illocutionary values of utterances in spontaneous speech are not assertive (Moneglia, 2011), it is not clear what the covert question in the context could be, being the adequate input of one of these specific speech acts. In conclusion, a constant aspect of every utterance derives from its pragmatic nature and from its illocutionary types which very rarely can be connected in an incontrovertible way to an objective/contextual input, and on the contrary ties to an internal affective disposition.

---

4 It’s likely that research carried out on map-task data, or call center conversations, or other kinds of ruled spoken exchanges will allow a different perspective, because within a shared and limited context the task of the participants is exactly that of posing questions and giving appropriate answers. However, even in these instances it is easy to find continuous counter-examples.
3. The prosodic Prominence

At this point it must be stressed that real speech must also be studied with consideration for its sound counterpart and, especially, some prosodic cues like terminal and non-terminal breaks, prosodic forms with illocutionary values, prosodic prominences necessarily signaling focus, etc. In accordance with these premises, it is assumed by the greater part of the literature that Focus must correlate with a phonetic-prosodic prominence⁵. The taking into account of this prosodic cue causes new contradictions, because it cannot be ignored, too, that there are utterances bearing two prosodic prominences. Thus, the occurrence in the same utterance of a first prominence and a second one, corresponding to semantic Foci, are a phenomenon it becomes necessary to explain.

In reality, systematic controls on the corpus carried out in LABLITA make us certain that not only are the root⁶ PUs performing a Comment characterized by a prosodic prominence, but that also prefix PUs performing Topics are mandatorily concluded by a perceptual prominence, sometimes more relevant than that in the Comment. This means that the Topic-Comment IP is always performed with a prefix PU and a root PU, each of them recording a prominence, corresponding to the prosodic nucleus of the PU. In conclusion, each utterance corresponding to a Topic-Comment information pattern is characterized by two Foci.

Facing the case of the two Foci utterances, scholars have been, in some sense, obliged to make the hypothesis of a Contrastive Focus (Büring, 2003). This has been explained within the context question-answer model through the hypothesis of a double question which should motivate the double Focus (who stole what?)⁷.

It is obvious that if the finding of a mandatory question input in the context to explain a Focus in the answer hardly appears acceptable, the hypothesis that the context questions had to double to also explain a Contrastive Focus seems even less so. It must be considered, moreover, that corpus data records about 10% of non-simple topicalisation phenomena i.e. the IP of a lot of utterances is not composed of a Contrastive Focus information pattern, but of a Topic-Topic-Comment, or a Topic-Topic-Comment, or of a List of Topics and a Comment. In this case each of the prefix PUs performing the respective Topic bears its own prosodic prominence, marking a Focus. Thus, according to the question-answer model there has to be a new Contrastive Focus every time there is a Topic, and by consequence a multi-multi covert questions input has to be found in the context to justify that result⁸.

For instance, for (1) below, how do we formulate a triple covert questions input, implying a covert question for the first Contrastive Focus in Topic, one for the second Contrastive Focus in the second Topic, and finally one for the Focus in the assertive Comment?

(1) *MAA: la maggior parte /TOP [...] quelli che hanno portato Pinocchio /TOP va proprio bene quello che hanno //COM ‘the major part... those who brought Pinocchio, what they have is all right’ %ill: assertion [ipubcv02]⁹

We don’t see how it could be possible to justify as input such a triple covert question in the context, which seems to be a totally ad hoc solution.

In conclusion: in a lot of influential literature the notion of Focus is strictly semantic and has been considered the central point for the information structure of the utterance. The concept has traditionally been defined according to vague notions of importance and novelty. Starting from the assumption of Common Ground within the model of context question-answering, more recent approaches have proposed the function of Focus as highlighting a semantic alternative in the answer and have hypothesized the existence of Contrastive Foci to explain the occurrence of utterances with two Foci.

4. The LAcT definition of Focus

In the LAcT perspective the importance of the concept of Focus is strongly rescaled because the information structure is not conceived as a semantic entity with a propositional size/form, whose Focus has to be the center. Information patterning does not depend on it, but on the pragmatic accomplishment of an illocution by the Comment, and on the pattern of Topic-Comment. The overall structure is not semantic but is still informative.

Focus remains a semantic concept in LAcT too, but its domain spreads only to the boundary of a textual IU of Comment or Topic. Expressions are conceived to develop an information function of Comment or Topic within the illocutionary act. Simultaneously, the same expressions, produced with an information function, are performed, from a syntactic and semantic point of view, as islands within the locutionary act. The performance of a Topic-Comment IP constitutes the accomplishment of an utterance, whose definition is pragmatic, but it does not correspond to the performance of a semantic proposition or of a syntactic sentence at the simultaneous locutionary level (Cresti & Moneglia, 2010).

Specifically, the semantics both of Topic and of Comment record kinds of relations regarding regency, quantification, modality, and Focus. Focus is a high semantic level of composition occurring both in Comment and Topic IUs. So, even if Focus is still a

---

⁵ See studies on prosodic Focus (Avesani & Vayra, 2003; D’Imperio, 2001).
⁶ For this terminology see (‘t Hart et al., 1990; Firenzuoli, 2003).
⁷ It must be noted that the hypothesis of Contrastive Focus does not assume that there is a Focus in the Topic but only that there are utterances with two Foci, one of which is considered “contrastive”.
⁸ On the contrary, with our information perspective the speaker can duplicate or triplicate the field of application of the illocutionary force, Topic, just adding linguistic details.
⁹ Examples are taken from Cresti & Moneglia, 2005 and from the LABLITA archive and are transcribed with an implemented version of the CHAT format (McWhinney, 2000; Moneglia & Cresti, 1997).
semantic notion, its domain is related not to an entire utterance, or presumed proposition, but to the semantics of a Topic or a Comment IU, which often copes only with phrasal constituents from a syntactic point of view.

Thus a general semantic definition of Focus in LAcT:

“A Focus signals the apex of a semantic domain which develops a Topic or a Comment information function within the information pattern of an utterance”.

The semantics of the domain behaving like a Topic or a Comment is conditioned by the information function that the expression is developing: in the case of Topic that of a field of application of an illocutionary force (T-Focus) and in the case of Comment that of the expression of an illocutionary force (C-Focus).

The Foci of the two IUs are apexes of semantic domains which systematically diverge for their semantic content and their respective lexical and morpho-syntactic composition: for Topic 75% of the linguistic content corresponds to Noun phrases and Prepositional phrases and for Comment 61% to Verb phrases. Generally speaking, T-Focus has a semantic identification function within a non-action domain and C-Focus has a semantic specification function within an action domain.

Therefore we claim that there are Topic Focus (T-Focus) and Comment Focus (C-Focus).

### 4.1 Topic-Focus (T-Focus)

For T-Focus, following its function must lead to the apex of a domain adequate in identifying the field of application of the illocutionary force. In an utterance like (2) with a total question force, but with two Topics, each Topic must identify a field of application for the question illocution in the Comment while functioning as a Topic by itself.

(2) *CEC: di là /TOP gli acidi /TOP tutto pronto ?COM “there, (for what concerns) the acids, everything ready?”

%ill: total question [ifamdll17]

The right part of the *prefix* PU is the seat of its prosodic nucleus, containing a prominence, and the majority of times it is performed with a relevant rising or a rising-falling movement. This position copes with the last semantic word of each Topic. So the adverb ‘là’ (*there*) and the noun ‘acidi’ (*acids*) can be considered the respective semantic Focus marked by the prosodic prominence.

Very often the expression, functioning as Topic, is from a syntactic point of view a well formed phrase (Noun, prepositional, adverbial), whose last word is also the head of the phrase. But it can happen that this coincidence doesn’t occur, like in the second Topic of (1), where the proper name ‘Pinocchio’ is the last word but it is not the head of the noun phrase. It should have been in doubt regarding the semantic or syntactic condition for being the Focus of a Topic domain, but corpus examples allow us to verify that it is always the final seat that correlates with the role of Focus in spite of the syntactic head position.

In speech we have the habit of expecting the end of something in recognizing it as a whole and the signal of ending or starting is given primarily by prosody. As a result, the last semantic word of the Topic marked by a prosodic prominence is recognized by the hearer as the expression closing the domain and identifying it as the semantic entity to be considered in its whole as the application field of the illocution i.e. the Topic.

### 4.2 Comment-Focus (C-Focus)

On the contrary C-Focus has no fixed seat, even if it occurs very often for a semantic word in the right side of a Comment IU or only at the last word. It depends on the fact that the C-Focus is also marked by the prosodic nucleus of the *root* PU, but this can occur in different seats within the PU depending mostly on the illocutionary type accomplished. Below are some examples with different illocations where the C-Focus doesn’t occur at the last word of the IU.

(3) *PAO: il resto /TOP non voglio sapere che cosa pensano //COM ‘for the rest, I don’t want to know what they think’

%ill: refusal [ipubcv01]

Figure 1: Utterance (2)

Figure 2: Utterance (3)
184

(4) *VAL: perché io sono stata nominata /SCA prima
   di’ trenta di’ giugno //COM ‘cause I have been
   appointed, before the thirtieth of June’
   %ill: answer [ifamvc18]

   Figure 3: Utterance (4)

For what it has been possible to verify, C-Focus, coping with the prosodic nucleus of the root PU, represents the phonetic part necessary to express and specify the illocutionary type of the Comment. This means that the recoverability of the illocutionary type is assured if the prosodic nucleus is the only sound conserved within the root PU.

Below are some examples where listening to the bare nucleus of the root PU allows the recognition of the illocutionary value. See Figure 2, where the prosodic shape of a total question is clearly recognizable from the last two syllables.

In (5) a partial question is performed.

(5) *PRO: [ l’unit linked] /TOP praticamente /TOP
   che cos’è ?COM ‘the linked unit, actually, what is it?’
   %ill: partial question [ipubdl04]

   Figure 4: Utterance (5)

(6) is an example of the expressive illocution of Contrast with a high jump on you.

(6) *PAO: che tu me l’avevi detto te /COM i’cream
caramel //APC ‘cause it was you that said it to me,
the cream caramel’
   %ill: assertion of contrast [ifamd12]

   Figure 5: Utterance (6)

Evidently what is relevant to perform with a C-Focus, more than the recoverability of an entire semantic domain (as in the case of T-Focus), is the sense of an expression through which a specific illocutionary act is accomplished. Then the goal of C-Focus emerges as supporting the word (s) and bettering the sense with which a specific illocution can be recognized, and in doing so prompts the addressee’s attention to the latter.

C-Focus marks the expression, allowing us to specify what type of illocution is performed within the semantic domain, dedicated in its whole to the accomplishment of the illocutionary force.11

In conclusion, according to LAcT the IP of the utterance has a pragmatic nature and its origin is in the accomplishment of an illocutionary force by the Comment. The IP does not correspond to a semantic structure whose center is the Focus. IP does not depend on Context. Focus corresponds to a semantic level of composition within the domain both of a Topic and of a Comment IU, and while T-Focus develops the semantic function of allowing the recoverability of the entire field of application for the illocutionary force, in its turn the semantic function of C-Focus is bettering the sense of an expression through which a specific illocution can be recognized. Both are mandatorily signalized by the nuclear prominence of their respective prefix PU and root PU.

5. References


11 Often the prosodic prominence marking the nucleus, dealing with the C-Focus, is of little relevance and specifically it can be weaker than that of T-Focus. This is not so strange, depending on the fact that what is necessary is that the root PU of Comment must clearly manifest a specific illocution. So this task is accomplished more by the form of the root PU than by the scale of the prosodic prominence, while in the Topic the only way to signal the Focus is through the relevance of its prominence.


LABLITA. http://lablita.dit.unifi.it/


Prosodic marking of referential status in Brazilian Portuguese: a preliminary study

Alexandre DELFINO, Maria Luiza CUNHA LIMA, Pablo ARANTES
Faculdade de Letras, Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627 - Belo Horizonte/MG - Brasil
E-mail: aldelfino@gmail.com, marialuiza.cunhalima@gmail.com, pabloarantes@gmail.com

Abstract

Current accounts of referential status (Gundel et al., 1993, inter alia) propose that the speaker’s mental model is reflected by a variety of linguistic forms during discourse production. In this framework, a scale of numerous informational statuses is related to the degree of givenness of the referents. Languages may mark these different degrees using prosody. For instance, in West Germanic languages such as German and English, it is said that new referents tend to be marked with an intonational prominence, whereas given referents tend to be deaccented. Accessible referents are marked with an intermediate marking, depending on its semantic relationship with previous referents (Baumann, 2006). The aim of this paper is to investigate how different degrees of informational status are acoustically marked along the speaker’s discourse in Brazilian Portuguese (BP). This study analyzed word duration, global F0 measures and time-normalized F0 contours of target words in three conditions: new, given and accessible referents. Results show that despite variability across speakers, both duration and F0 are used to mark different statuses. New and given statuses have the most different prosodic patterns and accessible is usually in between the two.

Keywords: information structure; referential status; prosody; Brazilian Portuguese.

1. Introduction

During discourse production, interlocutors refer to entities and events from the real world, and a mental model is built as new information is added and integrated to given, previous information. These entities and events are surfaced as linguistic forms, typically under a form of referential expressions e.g. determiner phrases.

A view widely accepted (Prince, 1981; Gundel et al., 1993; Chafe, 1976, 1994; Almor, 1999; Baumann, 2006; Baumann & Riester, 2010) posits referential expressions as taking a whole range of referential statuses, despite a traditional division of referential expressions into given and new information. Consider the following sentence:

John had to call the tow service because the engine had broken down on the road.

In the example above, the referent the engine cannot be taken simply as a given referent. First, a strict morphosyntactic analysis indicates the determiner as typically related to familiar referents yet the referent has not been previously mentioned. On the other hand, it cannot be also considered a new referent, as its meaning can be taken through context. One might conclude that referential forms do not only possess a basic lexical meaning, but also an information status regarding cognitive and contextual factors (Baumann & Riester, 2010, 1). In fact, it seems that ‘given’ and ‘new’ information describes both ends of a continuum of referential statuses.

One central question on prosodically-encoded information structure relates to determining which referential statuses can also receive a specific prosodic counterpart. Under a phonological perspective, Baumann & Grice (2006) and Baumann (2006) show that the informational status or activation level of a referent can be either lexically or acoustically marked. In West Germanic languages such as German, a three-way distinction - new, given and accessible - is said to be expressed in terms of prosodic encoding. New referents tend to be marked with a phrasal accent (H*), whereas given referents tend to be deaccented. The acoustic marking of the accessible status is very sensitive to various semantic relationships (e.g. hyponymy, synonymy, meronymy) between the previous item and the referent, and it does not seem to have a well-defined acoustic marking. Baumann (2006) observed that for accessible referents whose semantic relationship to its prime is whole-to-part tend to present an intermediate phrase accent (H+L*). Fowler & Housum (1987) carried out an acoustic analysis of first and second occurrences of words in English and concluded that second occurrence words were shorter than first occurrences.

By means of an electrophysiological measurement experiment (EEG), Schumacher & Baumann (2010) tested how prosodic information can affect reference processing. Two components were analyzed, the N400 and a late-positivity. The results show that reference processing takes prosodic information into account, together with semantic or morphosyntactic marking. ‘The data thus show that prosodic information guides the computation of a referent’s accessibility and can result in integration costs when less appropriate accent types are encountered’ (Schumacher & Baumann, 2010, 620-1). The experiment results also lead to the conclusion that the three-way classification of the referential status is significant not only for production, but also for perception.

The possible acoustic correlates of informational status have not been well studied in Brazilian Portuguese (BP). In a descriptive experimental study of test words nested within noun phrases, Arantes et al. (in preparation) show that there are some prosodic differences between new and given referents: (i) new
The present experiment we apply the same descriptive tools used by Arantes and collaborators to expand the investigation about the prosodic correlates of referential status in BP beyond the new-given dichotomy. More specifically, we tried to see if the proposed threefold distinction found in German (new, given and accessible) can be found in BP as well. In contrast to Baumann (2006), we focus on the description of the acoustic patterns found and do not provide a phonological interpretation of them.

2. Experiment

For this study, we designed a corpus of 90 short paragraphs, which were divided distributed into three conditions: given, new and accessible. Each paragraph had one target word, which was embedded in a control phrase.

All target words are four-syllable, penultimate stressed. Relatively long words were chosen as targets because Arantes et al. results suggest that prosodic effects due to referential status are more evident when more phonetic material is available to the speaker. Sentences preceding the control phrase provide context that determine if the target NP is given, new or accessible.

The following paragraphs are examples of the three conditions investigated in the experiment. Control phrase in italic and target word in bold italic.

New

Um terremoto causou destruição em boa parte da costa leste. Várias cidades não tinham um programa de evacuação, o que deu trabalho para as equipes de resgate.

(An earthquake caused destruction in a huge part of the East coast. Several cities did not have an evacuation program, which caused problems to the rescue teams.)

Given

O governo decidiu fechar a usina nuclear após o terremoto ocorrido no mês passado. O terremoto causou destruição no núcleo do reator, aumentando o risco de contaminação. (The government decided to shut down the nuclear plant after the earthquake occurred last month. The earthquake caused destruction to the reactor nucleus, increasing the risk of contamination.)

Accessible

Estudiosos da Sismologia têm procurado analisar os dados de tremores para prever novas ocorrências. O terremoto causou destruição sem que ninguém pudesse se prevenir.

(Seismology experts have been trying to analyze the tremors data to predict new occurrences. The earthquake caused destruction without anyone being able to prepare themselves.)

Four subjects (one male) read the 90 paragraphs, presented one by one on a computer screen in randomized order. Subjects were instructed to read each paragraph silently before reading it out loud to ensure they would be aware of the content of the paragraphs and minimize hesitation. Subjects were recorded in a sound treated room in separate sessions. After each recording session, sound files were edited and labeled. For each sound file the target NP (determiner plus noun) was manually segmented into syllables and the boundaries were stored in Praat metadata files. All acoustical analyses were performed with the help of Praat scripts.

Arantes et al. (in preparation) investigated a wide range of acoustic correlates traditionally linked to prosodic functions in order to find the ones that correlates the best with referential status differences. The authors measured acoustic duration, fundamental frequency, spectral emphasis and long-term average spectrum and suggest that at least for BP duration and fundamental frequency are the best correlates of referential status. Following their suggestion those were the acoustic parameters analyzed in the current study.

Because definite and indefinite determiners have different number of syllables in BP, we decided to measure the duration of the noun instead of the whole NP. Duration values were extracted and the means across the different values of referential status were then calculated.

Fundamental frequency was analyzed in two ways. First, mean values of central tendency (mean) and variability (standard deviation and range) of the test NPs F0 contours were compared among the referential status values. For the F0 central tendency analysis, we extracted the mean F0 value of all target NPs (determiner plus noun) and then calculated the mean of the means grouped by referential status value. This measure can be interpreted as the pitch level of the F0 contours. Mean standard deviation and range were obtained applying the same procedure. These two measures were used as estimates of the flatness or “bumpiness” of the F0 contour in each status condition. Range in semitones was calculated for each contour by applying the formula below. F0max and F0min are respectively the maximum and minimum F0 values in the contour:

$$12 \log_{10}(\text{max}_{\text{Hz}}/\text{min}_{\text{Hz}})$$

For the second analysis, individual F0 contours were time-normalized following the procedure described in Arantes (2011), which allows the comparison of F0 contours having different duration. Mean time-normalized contours for each referential status value
were obtained and then visually compared. The main hypothesis being tested is that ‘new’ referents, being the most salient, will have longer duration, higher F0 mean, standard deviation and range when compared to ‘given’ referents. Following Bauman’s findings, it’s also possible to predict that ‘accessible’ referents will be in between the two others in terms of the values of the acoustical parameters.

The data generated by the four subjects were analyzed separately. Referential status with three levels (given, new, accessible) was the independent variable for all analysis. Analysis of variance (ANOVA) was used to determine if differences in mean values of the acoustic parameters were statistically different among the levels of the independent variable. An alpha level of 5% was adopted for all analyses. When post-hoc multiple comparisons were performed the alpha level was adjusted by the Bonferroni correction.

3. Results

3.1 Duration

Figure 1: Mean target word duration (in milliseconds) grouped by referential status (accessible, given and new) and subject. Whiskers indicate 95% confidence intervals

Mean target word duration grouped by referential condition and subject is shown in Figure 1. The statistical analysis supports the main hypothesis being tested: for all subjects, ‘new’ referents were longer than ‘given’ ones. For subjects f3 and m1, ‘new’ was also longer than ‘accessible’. Differences between ‘given’ and ‘accessible’ are never statistically significant. Statistical results by subject are reported below:

- f1: $F(2, 84) = 10.4 \ p < 0.001$; N-G: $p < 0.001$; N-A: $p < 0.01$; A-G: n.s.
- f2: $F(2, 84) = 7 \ p < 0.01$; N-G: $p < 0.01$; N-A: n.s.; A-G: $p < 0.1$.
- f3: $F(2, 84) = 9.9 \ p < 0.001$; N-G: $p < 0.001$; N-A: $p < 0.05$; A-G: n.s.
- m1: $F(2, 84) = 20.5 \ p < 0.001$; N-G: $p < 0.001$; N-A: $p < 0.001$; A-G: n.s.

3.2 Mean F0

Mean target NP mean F0 grouped by referential condition and subject is shown in Figure 2. The average value of F0 contours of ‘new’ referents is significantly greater than ‘given’ and ‘accessible’ for subjects f1, f2 and m1. There is no difference between ‘given’ and ‘accessible’ ones. Subject f3 presented no significant differences among statuses. Statistical results by subject are reported below:

- f1: $F(2, 84) = 6.64 \ p < 0.01$; N-G: $p < 0.01$; N-A: $p < 0.05$; A-G: n.s.
- f2: $F(2, 84) = 7.61 \ p < 0.001$; N-G: $p < 0.01$; N-A: $p < 0.05$; A-G: n.s.
- f3: $F(2, 84) = 0.7 \ n.s.$
- m1: $F(2, 84) = 123.73 \ p < 0.001$; N-G: $p < 0.001$; N-A: $p < 0.001$; A-G: n.s.

Figure 2: Mean target NP mean F0 (in Hz) grouped by referential status (accessible, given and new) and subject. Whiskers indicate 95% confidence intervals

3.3 Mean SD and Range

Referential status affects standard deviation and range only for one of the speakers, namely m1. Statistical results by subject are reported below:

- f1: $SD \ F(2, 84) = 3 \ p < 0.1$; range $F(2, 84) = 0.32$ n.s.
- f2: $SD \ F(2, 84) = 1.8$ n.s.; range $F(2, 84) = 0.03$ n.s.
- f3: $SD \ F(2, 84) = 0.8$ n.s.; range $F(2, 84) = 0.6$ n.s.
- m1: $SD \ F(2, 84) = 86 \ p < 0.001$; N-G: $p < 0.001$; N-A: n.s.; A-G: n.s.; range $F(2, 84) = 56$ p < 0.001; N-G: p < 0.001; N-A: n.s.; A-G: n.s.
3.4 Time-normalized F0 contours

For all subjects, given and accessible contours overlap significantly. For subjects f1 and f2, new referents contours differ from the other statuses mainly because they present a peak aligned to the first two syllables of the target NP that is absent in the other statuses. Subject m1 given and accessible contours are mostly flat and in a lower register when compared to the new referents, that also have an initial high F0 peak. Subject f3 is unlike the others because referential status does not seem to influence the shape of the F0 contours at all.

On the whole, the results revealed that F0 contours of new referents are different from the given and accessible ones. Despite the individual variability, new referents are characterized by the presence of two major pitch peaks, one extending over the chain of pre-stressed syllables and other aligned to the stressed syllable. Given and accessible referent contours are very similar to each other.

Figure 5: Target NP time-normalized F0 contours grouped by referential status (accessible, given and new) and subject. Syllable boundaries indicated by vertical lines

4. Discussion

In general, there is positive evidence that acoustic parameters are affected by the referential status contrast. Duration seems to be the most robust correlate of the status distinction because it is the only parameter that is affected by the status variable in all subjects.

In addition to duration, mean F0 value is also affected by the referential contrast, with new referents being spoken in a higher register. Except for subject m1, there seems to be little difference in terms of F0 variability between the status categories investigated.

Besides the observed differences in pitch register, the time-normalized contours analysis suggests that the presence of a NP-initial F0 peak can be used as a correlate of the 'new' status in contrast to the other two statuses. In BP, the chain of pre-stressed syllables (including the NP determiner) seems to be an important locus of F0 differences among the status levels.

The lack of a clear distinct acoustic pattern for the accessible status can be evidence that the prosodic marking works, in a general way, associated to other types of information (syntactic position, semantic relationship, register, focus, etc.). Baumann’s (2006) results showed that prosodic marking of accessible status were consistently observed in one type of relationship (whole-to-part), and may be of limited use.
5. Conclusion
The aim of this study is to investigate the relationship between referential status and its prosodic manifestation from the production point of view. Moreover, we intended to observe if any of the analyzed acoustic parameters could set significant differences among the three statuses, i.e. given, new and accessible statuses.

Word duration is the most expressive parameter, followed by the pitch level. Pitch variation and range does not seem to play an important role. New and given status are pretty distinct in most parameters; however, the accessible status is either too sensitive to its semantic relationship to the prime word or it is not relevant for BP speakers as for German speakers.

The current results lead to conclude that there is an interface between referential status and prosodic information, and that relationship is variable in different languages.

6. Acknowledgements
This work was supported by Capes Grant 2552/2010 to the third author.

7. References


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

Edoardo LOMBARDI VALLAURI¹, Fabio TAMBURINI²

¹Department of Linguistics, University of Roma Tre, Italy; ²Department 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:

\[
\text{Prom}^i = \text{Prom}^i_{\text{fa}} \cdot \left[ \text{SpEmph}_{\text{sML-SPL}} \cdot \text{dur} \right] + \text{Prom}^i_{\text{fa}} \cdot \left[ \text{en}_{\text{fa}} \cdot \left( A_{\text{event}(atM, atM)} - D_{\text{event}(atM, atM)} \right) \right]
\]

where \(\text{SpEmph}_{\text{sML-SPL}}\) is the spectral emphasis, \(\text{dur}\) is the nucleus duration, \(\text{en}_{\text{fa}}\) is the overall energy in the nucleus and \(A_{\text{event}}\) and \(D_{\text{event}}\) are the parameters derived from the TILT model (Taylor, 2000) as a function of the maxima alignment type – \(a_{\text{M}}\) – and the minima alignment type – \(a_{\text{m}}\). All parameters are referred to the generic syllable nucleus \(i\). See Tamburini (2006) for further details on parameter computation.

The body of the function \(\text{Prom}\) contains nine parameters. Five of them can be considered as supporting the prominence phenomenon from a cross-linguistic point of view (\(\text{SpEmph}_{\text{sML-SPL}}\), \(\text{dur}\), \(\text{en}_{\text{fa}}\), \(A_{\text{event}}\) and \(D_{\text{event}}\)), while the other four, represented in the vector \(\mathbf{W} = (W_{\text{fa}}, W_{\text{pa}}, a_{\text{M}}, a_{\text{m}})\), can be seen as language specific. In our model, \(W_{\text{fa}}\) and \(W_{\text{pa}}\) weigh the contribution of the two different accent types, while \(a_{\text{M}}\) and \(a_{\text{m}}\) model the different pitch accent alignments specific for each language.

All the parameters involved in the \textit{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 invariances 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.

<table>
<thead>
<tr>
<th>Var.-</th>
<th>Main Prominence on the...</th>
<th>No Main Prom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corp.</td>
<td>(LsT)</td>
<td>(LsF)</td>
</tr>
<tr>
<td>R-B</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>R-C</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>F-C</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>N-C</td>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1: Number of utterances divided by Variety-Corpus pairs (R=Rome, F=Florence, N=Naples; B=Bonvino, C=CLIPS) and configurations (e.g. \(LsT=\text{Last syl. of Topic, IsF=Internal syl. of Focus}\)). Some combination pairs are not possible; in those cases we have inserted a ‘-’ in the corresponding cells.
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:

<table>
<thead>
<tr>
<th>Evaluated IS</th>
<th>Measured position of MP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic-Focus</td>
<td>MP at Right end of Topic</td>
</tr>
<tr>
<td>Focus-Appendix</td>
<td>MP at Right end of Focus</td>
</tr>
<tr>
<td>Broad Focus</td>
<td>MP at Right end, or no MP</td>
</tr>
<tr>
<td>Topic-Focus or Broad Focus</td>
<td>No evident MP</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>corresponding to the description</th>
<th>not corresponding to the description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rome – Bonvino</td>
<td>43 (91.49%)</td>
<td>4 (8.51%)</td>
</tr>
<tr>
<td>Rome – Clips</td>
<td>55 (85.94%)</td>
<td>9 (14.06%)</td>
</tr>
<tr>
<td>Florence – Clips</td>
<td>53 (89.83%)</td>
<td>6 (10.17%)</td>
</tr>
<tr>
<td>Naples – Clips</td>
<td>53 (86.89%)</td>
<td>8 (13.11%)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>170 (87.88%)</td>
<td>28 (12.12%)</td>
</tr>
</tbody>
</table>

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.

![Diagram](image.png)

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

7. References


The work we present is an analysis of references in a corpus of French sportscasts. The successive naming of players and their actions is very much a part of the descriptive part of sport comment. By analysing a corpus of descriptive speech (a rugby match), we want to prove that it is not just a string of player names who currently perform the action but a construction of the referential structure, all along the discourse and in every descriptive period, which makes the discourse coherent. For the 120 descriptive periods of the chosen sportscast, we marked the first introduction of every referent and its further references and resolutions. More precisely, we distinguished the referent activation in a given descriptive period from the coreferent expression(s) or reactivation(s) that followed its introduction, and the properties of all these elements (part of speech, prosodic and syntactic realisations) were noted.

Keywords: reference; French; sportscast; information structure; syntax; prosody.

1. Conceptual and methodological frameworks

The work we present consists in analysing references in a particular type of spoken corpus: live sports comment.

This study is based on a corpus of spoken French, a rugby match sportscast, and we are studying the descriptive periods, directly produced in relation to actions taking place on the field, under the speaker’s eyes.

We study the referential structure of these descriptive periods: how referents are introduced and reactivated (information structure) in this particular speech situation at syntactic and prosodic levels?

1.1 The corpus

Our corpus main characteristics (see Lortal & Mathon, 2008 for more information) are:

- Sports event: France-Argentine (Rugby World Cup 2007)
- Sportscasting language: French
- Recorded on TV (TF1)
- Number of speakers: 3
- Duration: 108 mn (total record), 55 mn (total speech), 40 mn (speaker 1), 13 mn (speaker 2)

We distinguish two types of periods:

- Descriptive periods (DP) that are in direct relation to actions
- Comments periods not (directly) related to actions (information about strategy, players’ career…)

For this study, we looked at the corpus 120 DP and their structures adapted to a particular production context and motivated by various parameters (see Boulakia & Mathon, 2011).

1.2 Corpus referential structure

At referential level, we consider three types of referents: those introduced for the first time in the discourse (ID), those activated for the first time in a given descriptive period (IP) and the coreferent expressions or reactivations of these referents (R).

Following Chafe (1976: 30) or Lambrecht (1988: 144), we distinguish “newly activated” or “unactivated” referents from “(already) activated” ones.

In the entire match comment, we have 40 different named-entities which are activated as:

- 510 cases of “new” referents in the discourse (ID)
- 470 cases of “new” referents in the DPs (IP)
- 346 cases of “given” referents (R)

In terms of part of speech, 76% of the ID & IP are proper names and more than 50% of the R are pronominal forms (relative pronoun, subject clitic…).

The three examples below illustrate the IP/R distinction:

1. IP + R (relative pronoun):
   les français sont debout avec la balle au fond dans les bras de Rémy Martin ouais ah qui perce au coeur

2. IP + 2 R (relative & strong pronouns):
   David Skrela qui j(oue) lui aussi joue dans les airs

3. IP + Intermediate referents (4) + 2 R (subject clitic and first & last name):
   avec Rémy Martin oh les jeux avec Pieter de Villiers avec Heymans maintenant c'est la grande relance française jusqu'à Rougerie Mignoni du rythme du rythme il est pris Rémy Martin
Our goal is to identify and correlate properties of these three types of referential expressions: syntactic category (NP, PP, Pro, etc.), position, utterance structure, prosodic realization...

1.3 Application to First name (FN) + Last Name (LN) referential expressions

All the corpus referential expressions were annotated, but we began our research by analysing specifically the ones that have the form “first name + last name” and the associate coreferent expressions.

Our test corpus contains 47 ID/IP referential expressions of the form [First Name + Last Name] corresponding to 26 R for these ID/IP.

Our study aims to realize a parallel analysis of these ID/IPs and R and identifies the syntactic and prosodic properties of each kind of referential expressions.

Three levels are taken into consideration:

- A pragmatic level, with three types of referential expressions: ID-IP, IP and R
- A syntactic level, with the identification of syntactic category, function, position, utterance structure, autonomy/dependency…
- A prosodic level, with measures of F0 patterns, F0 register (Low, Medium-Low, Medium-High, High) and F0 range (Delta F0, Max, F0 Min.).

2. Pragmatic and syntactic structure

In the DPs, discourse organisation is in part related to information progression but also to the iconicity of the production situation:

- Referential expressions’ properties depend on the kind of referent: ID/IP are typically introduced by definite descriptions or names, followed by one or more coreferent pronouns...
- Referential expressions properties depend on the sportscaster’s relation to the action on the field: speaker’s implication and action intensity weigh on the speaker’s syntactic and prosodic productions (Boulakia & Mathon, 2009).

At syntactic level, we analysed the referential expressions in terms of syntactic category, function, position, structure and autonomy/dependency. Our goal, here, is to evaluate some tendencies in our corpus such as:

- Postverbal expressions (subject or complement) are typically “new” information and have a particular prosodic realisation.
- Structures like cleft-sentences or dislocations are used to transmit specific information and associated with specific prosodic structures.
- Prepositional phrases and verbless sentences, very present in sportscasts, can be realised as independent groups. To check this last point, we used macrosyntactic categories (Blanche-Benveniste, 1997), to distinguish referential expressions that constitute a unit by themselves (nucleus, prefix, suffix, postfix) from referential expressions that are embedded in a macrosentence.

The tables below present first the results concerning FN+LN sequences and then those concerning the resolutions of these sequences:

<table>
<thead>
<tr>
<th>First Name + Last Name</th>
<th>47</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referential category</td>
<td>11 ID/IP &amp; 36 IP</td>
</tr>
<tr>
<td>Phrasal part of speech</td>
<td>27 PP &amp; 20 NP</td>
</tr>
<tr>
<td>Macro-syntactic category</td>
<td>39/47 embedded 8/47 whole macrophrase</td>
</tr>
<tr>
<td>Function</td>
<td>14 Verb complement 11 Noun complement 8 ‘utterances’ 6 NP’s heads 4 subjects 1 ATS 3 juxtapositions</td>
</tr>
</tbody>
</table>

Table 1: Pragmatic and syntactic properties for FN + LN sequences

Let’s take some examples of FN+LN expressions:

4. FN+LN at the end of a unit
   "à la lutte avec Nani Corletto"

5. FN+LN at the beginning of a unit
   "Felipe Contepomi pour ce drop"

6. FN+LN that constitute a unit
   "avec Damien Traille"

7. FN+LN in the middle of a unit
   "qui a décalé Lucas Borges face à..."

As presented in Table 1, referents are mostly not placed at the beginning of a syntactic and prosodic unit. They are mostly not introduced as topics (Player + Action, 9/47), but rather as (part of) focus (Action + Player, 38/47)

For the 8 FN+LN sequences realised as independent units, the principal formal characteristic is that 6 of these expressions are PP, introduced by ‘avec’ (‘with’) and ‘pour’ (‘for’).
Table 2 shows corpus resolutions’ properties:

<table>
<thead>
<tr>
<th>Resolutions</th>
<th>Number of occurrences</th>
<th>Average of intermediate referents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative subject</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Last name or First name + Last name</td>
<td>10</td>
<td>2-3</td>
</tr>
<tr>
<td>Clitic subject</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>NP</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Strong pronoun</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

As presented in this second table, there is no intermediate referent between a IP (FN+LN) and a (pro)nominal coreferent expression, whereas there is an average of 2-3 intermediate referents between a IP (FN+LN) and a direct anaphora (FN+LN or LN).

This data indicates that we have to distinguish two types of coreferent expressions: simple anaphora (pronominal forms) without intermediate referents and referents’ reactivations ((FN)+LN) with intermediate referents.

Example 8 is a case of reactivation of a referent with FN + LN, since there are intermediate referents:

8. FN+LN(1) + Intermediate Referents + FN+LN(1)
Hernandez le drop avec le pied gauche qui va mourir sous les poteaux où se trouve Cédric Heymans avec un arrêt de volée accordé par Monsieur Spreadbury tentative qui a échoué d'un rien hein de la part de Juan Martin Hernandez et le pied gauche de Cédric Heymans

In this first example, a referent (Cédric Heymans) is introduced by a FN+LN sequence. After this, two other referents (Monsieur Spreadbury and Juan Martin Hernandez) are mentioned. In order to reintroduce the first referent, the speaker uses a FN+LN sequence, not a pronominal expression that would have been ambiguous.

9. Particular case FN+LN(1) + Ø + FN+LN(1) ah c'est bien Rémy Martin (spk1) Rémy Martin qui l’a chipé c’est bien (spk2)

This second example is quite particular as there is no intermediate referent between two coreferent FN+LN expressions. There is only this case of direct anaphora [FN+LN] + [FN+LN] in the corpus: the juxtaposition can be explained by the simultaneity of two speaker turns. Speaker 1 introduces the referent and at the same time Speaker 2 (re)introduces the referent for his own production.

3. Prosodic structure

At prosodic level, our study aimed to show that prosody is an indicator of referent activation status, by analysing referential expressions in terms of melodic patterns, F0 range, F0 registers.

Furthermore, prosodic structure helps in identifying corpus macrosentences by distinguishing prosodically autonomous phrases dependent ones.

Our primary interest at this level was to measure the efficiency of prosody as an indicator of the activation status of referents. We focused on melodic variation, especially F0 Range, and F0 registers (Low, Medium-Low, Medium-High, High).

Figure 1 shows F0 range values depending on referents type (ID, IP, R).

![Figure 1: F0 range values depending on referents type (ID, IP, R)](image)

70% of Rs show a very narrow range of variation (50Hz or less).

50% of the IPs present a narrow range of variation (50Hz or less), while 35% show a wider range (100Hz or less).

IDs are more often realised with a wider range of variation (100Hz or less), and it’s the only category to show in some cases (20%) a wide range of variation (from 200Hz to 250Hz).

Figure 2 shows the repartition of referent realisations depending on F0 registers and referents type (ID, IP, R).

![Figure 2: F0 registers depending on referents type (ID, IP, R)](image)

Most of the referents are realised in a Medium-Low register independently from the referential category.
ID referential category is the only one to be realised in some cases in High register, and it is the less to be realised in a Low register.

There is no significant difference between IP and R concerning the F0 registers.

Results showed that there is no statistical evidence for a correlation between F0 range and activation status of referent for one hand, and between F0 registers and activation status of referent for the other. We just note a tendency for High register and wide F0 range being correlated with ID.

4. Prosodic realisations: some examples

We selected some examples of referents prosodic realisations, in order to understand what could be the reason of melodic variations, if not the referent activation status.

We selected utterances presenting the following two referential structures:

- ID-IP + R1 (Clitic subject) + R2 (First Name + Last Name)
- IP1(First Name + Last Name) + IP2 (Nominal Phrase) + R1(PN) + R2 (Relative Pronoun)

We also present a case of prosodic realisation motivated by action on the field.

4.1 ID-IP

Our first assumption was that the referent's degree of activation, at both discourse and descriptive period levels, could be prosodically patterned. The analysis shows that a referent’s new introduction in the discourse is not systematically characterized by melodic prominence (see 4.3.).

Figure 3, for example, shows the melodic variations for the ID-IP Nani Corletto and its reactivation as a proper name.

![Figure 3: Melodic variations for the utterance force Nani Corletto (IP) à une relance / il (R) y excelle / Nani Corletto (R) une chandelle](image)

The ID-IP Nani Corletto is introduced in a ML Register in an ascending-descending melodic pattern. The referent is first referenced by a clitic subject, which is not accented, then reactivated in a ML Register as well, and in the same pattern, but 50 Hertz lower. This reactivation is a postfix, i.e. an addition of information, given afterward, as a right dislocation. The macrosyntactic function could explain the non-prominent melodic realization of the reactivation (FN + LN).

4.2 IP1 + IP2 + R1 (LN)

Figures 4 and 5 show melodic variations for the utterance *dans le dos de Damien Traille (IP) c’est une touche trouvée par l’Argentine (IP) avec Traille (R) qui (R) saigne hein*

which is composed of the following referential structure:

IP1 + IP2 + R1 (PN) + R2

![Figure 4: Melodic variations for the utterance dans le dos de Damien Traille (IP) c’est une touche trouvée par l’Argentine (IP) avec Traille (R) qui (R) saigne hein](image)

On Figure 4, the pitch prominence is on the word *dos* (back), while the new referent in the period is activated in a ML register in a plane pattern.

![Figure 5: Melodic variations for the utterance avec Traille (R) qui (R) saigne hein](image)

A new referent *l’Argentine* is introduced. Then the referent *Damien Traille* is reactivated as a Proper Name followed by a relative pronoun, in a type of parenthesis, in a ML-L register and in a plane pattern.

Prosodic realisation does not seem to depend on the degree of activation of the referent, but rather on the impact of the situation on the ground.

4.3 Influence of action

Figure 6 shows melodic variations for IP *avec Pieter de Villiers* in the utterance:

*les voilà les meilleurs ballons à jouer avec Jauzion / avec Rémy Martin / [oh les jeux] / avec *Pieter de Villiers* (IP) / avec Heymans maintenant c’est la grande relance / française jusqu’à Rougerie*
This descriptive period corresponds to a specific offensive action from the French team, whom the speaker supports. He enumerates the players who are part of the offensive action. Each new referent is introduced as an independent unit, a nucleus. The speaker’s excitation, his enthusiasm is conveyed by by the use of the MH register, and the F0 range of 100Hz.

In this situation, the iconic function of prosody strongly influences the prosodic realisation of the referent.

5. Conclusion
Our work aimed at describing sport comment’s referential structure and more particularly the case of FN+LN sequences and their coreferent expressions.

First of all, the analysis led us to distinguish three types of referential expressions:

- a referent’s first introduction by a proper name or a nominal expression,
- a resolution by a pronominal form when the anaphora is direct,
- a reactivation by a proper name or a nominal expression. A referent’s reactivation is required as soon as some intermediate referents are introduced between the first introduction of a referent in a descriptive period and its resolution.

Concerning the role of iconicity and sport comment’s syntactic and macrosyntactic properties, we can conclude that:

- discourse structure is highly dependent on game actions (iconicity), rather than on information progression principles,
- referents’ introduction typically follows action’s introduction (81% of the cases) and players-agents are presented as new discourse information,
- specific syntactic structures as preposition + Proper name tend to be realized as independent macrosyntactic units.

At prosodic level, we saw that there is no prosodic difference between the first introduction of a referent at discourse level, its reactivation(s) in the descriptive periods, and its resolution(s) within the periods. In fact, at descriptive period level, the referential structure is highly related to the action on the field and the iconicity has a more important impact on prosodic realisation than the degree of activation of referents in this kind of discourse.

6. References


Prosodic features of the topic information unit in BP and EP: a corpus based study

Maryualê MITTMANN, Bruno ROCHA
LEEIM Laboratory – UFMG
Av. Antônio Carlos, 6627 – Belo Horizonte
maryuale@gmail.com, bbruno791@gmail.com

Abstract

The aim of this study is to describe the topic intonational forms for European and Brazilian Portuguese (EP and BP). The dataset comes from two comparable spontaneous speech corpora: C-ORAL-ROM (EP) and C-ORAL-BRASIL (BP). The theoretical framework is the Language into Act Theory (Cresti, 2000), according to which the utterance corresponds to the shortest linguistic unit that can be pragmatically interpreted. The speech flow is segmented into utterances and its internal units by prosody. One of the most important informational units is the “topic”, which function is to identify the domain of relevance for the illocution conveyed by the utterance. Corpus based studies (Firenzueli & Signorini, 2003) identified three different intonational forms for the topic unit (types 1, 2 and 3) in Italian. This work shows that BP and EP present the three intonational forms found before and also a fourth one. Prosodic parameters of types 1 and 2 are highly similar in all three languages. Type 3 is the least common form in BP and type 1 is rarely used in EP. Type 4 topics seem to differ in BP and EP regarding the F0 values of the first tonic syllable onset.

Keywords: topic; informational structure; spontaneous speech; corpus; C-ORAL-BRASIL.

1. Introduction

The aim of this study is to describe the topic intonational forms for European and Brazilian Portuguese. The dataset comes from two comparable spontaneous speech corpora: C-ORAL-ROM (European Portuguese – EP) and C-ORAL-BRASIL (Brazilian Portuguese – BP). The theoretical framework is the Language into Act Theory (Cresti, 2000).

2. C-ORAL-BRASIL and C-ORAL-ROM corpora

The C-ORAL-ROM (Cresti & Moneglia, 2005) is a multilingual corpus for Italian, European Portuguese, French and Spanish. It was compiled by a consortium coordinated by the University of Florence. The EP session of the C-ORAL-ROM consists of 152 recordings and 317,916 words. The transcriptions are segmented in prosodic/pragmatics units, in order to provide the MEANS ADEQUADO for pragmatic studies. The corpus is integrated with Win Pitch software (Martin, 2004) files, which allows the simultaneous exploitation of the transcription and the acoustic data. The C-ORAL-ROM architecture is designed to cover a large amount of different recording situations in order to document a great variety of speech acts present on spontaneous speech.

The C-ORAL-BRASIL corpus (Raso & Mello, 2012) presents the same architecture of the C-ORAL-ROM and is completely comparable to it. C-ORAL-BRASIL gives a special emphasis on the diaphasic variation of the recordings and it has a very small number of interviews and chats.

3. Language into Act Theory (LAcT)

According to LAcT, the linguistic behaviour is accomplished through speech acts (Austin, 1962). A speech act is understood as the simultaneous performance of three acts: locutionary, illocutionary and perlocutionary. The locutionary act corresponds to the utterance, defined as "the linguistic entity accomplished by the speech act". The utterance is considered the reference unit for the analysis of spoken language and is the shortest linguistic unit that can be pragmatically interpreted (Cresti, 2000). According to LAcT, there isn’t a necessary correlation between utterances and propositions and corpus based studies have shown that a high percentage of linguistic units that are pragmatically autonomous don’t express a proposition (Crest, 2005).

In this framework, prosody works as an interface between the locutionary and illocutionary acts and it has three important functions: (i) to delimit the utterances in within speech flow; (ii) to assign the illocution conveyed in the utterance; and (iii) to organize information within the utterance.

As for the first function, the utterance is delimited by prosodic breaks perceived by the hearer as conclusive (terminal breaks) and can be parsed into smaller prosodic units (tone units), delimited by prosodic breaks perceived as non-conclusive (non-terminal breaks). The difference between terminal and non-terminal breaks can be seen in example (1), in which “//” marks a non-terminal break and “/” marks a terminal break.

(1) *BAL: as recarregáveis // tão aqui //

When hearing the sequence until the non-terminal break - // - , the fluent speaker doesn't perceive the tonal unit as a conclusive and autonomous sequence. It happens only when the speaker hears the whole sequence until the terminal break.

When a utterance is composed by a single prosodic unit it's considered a simple utterance. If the utterance is parsed into two or more prosodic units is, then it is considered complex. The example (1) shows a complex utterance and example (2) shows a simple one.

When 

The possible internal units are associated with informational functions, through which information is patterned within the utterance. According to LAcT, each prosodic unit corresponds, in principle, to an informational unit. The core of the utterance corresponds to the prosodic unit that bears the utterance's illocutionary force. This unit is called “comment”, and it is necessary and sufficient to form an utterance.

Other prosodic units correlate with different information functions, that can be textual (i.e. units that either compose or act on the text), or dialogic (i.e. units that are directed to the addressee and regulate the communicative channel) (Cresti, 2000; Cresti & Moneglia, 2010).

3.1 Topic information unit

Among the textual information functions, the most important and frequent (about 50% of textual units in a BP sample) is the “topic”. The topic information function identifies the domain of relevance for the illocutionary act, allowing for the illocution to be distanced from the direct situational context of speech production. The topic provides a linguistic context for the illocution carried by the Comment, when the situational context is not sufficient for the proper interpretation of the speech act (Signorini, 2005). The following example illustrates this.

(3) *CLA: come lei va via la sera /\=TOP= nell’ ascensore ‘un c’è più luce //

Corpus based studies (Firenzuoli & Signorini, 2003) identified three different intonational forms for the topic unit (types 1, 2 and 3). An intonational form is defined as a set of prosodic features that occurs consistently within an information unit and correlates with its informational function: pitch contour, timing, duration and F0 values. An intonational form is constituted by three distinct tonal portions: preparation, nucleus and coda. The nucleus carries the perceptual prominence associated with the informational function and is, therefore, mandatory. If the syllabic material is greater than what is necessary to accomplish the nucleus, it is distributed in the preparation and/or the coda, which doesn’t play any functional role in the topic.

3.1.1. Type 1 form of topic

Type 1 topic is characterized by a rising-falling F0 movement on the nucleus. The rising movement is on the last tonic syllable, and the falling is on the post-tonic syllable(s); tonic and the post-tonic(s) syllables are lengthened.

(2) *PAU: e cê acha que vai gastar mais um //

3.1.2 Type 2 form of topic

Type 2 has a rising intonation profile that begins in the last tonic syllable and continues in any potential post-tonic syllables. Tonic and post-tonic(s) syllables are lengthened.

3.1.3. Type 3 form of topic

Type 3 can be considered holistic, since the nucleus is distributed in two semi-nuclei, together building the topic functional focus. The first semi-nucleus has a falling profile while the second has a rising one, always corresponding to the last syllable of the topic, whether tonic or post-tonic. This syllable is also lengthened.

3.2 Methods

In this study, the intonational analysis of topics was carried out in two samples (BP and PE) of speech corpora that were previously prosodically segmented into utterances (simple and complex) according to the methodology developed from the framework of the LAcT (Moneglia & Cresti, 1997). The utterances containing at least one topic unit were extracted from the sample and then it was formed a second sample containing 110 utterances for BP and 72 utterances form EP (EP numbers
are smaller because this corpus wasn't previously informationally tagged). Then we proceeded the analysis through Praat software (Boersma & Weeninck, 2011) through the following steps: (i) identification of the nucleus of the topics; (ii) extraction of F0 values, intensity and duration (syllabic and vocalic) of the nucleus; (iii) stylization of F0 contour of the nucleus; (iv) grouping the topics according to their prosodic parameters; (v) manipulation of prosodic parameters through speech resynthesis. The aim of the manipulations was to identify the most relevant prosodic parameters to the nucleus/semi-nuclei of each intonational form of topic.

4.1 Demonstration of manipulations

In order to clarify the manipulations we did on this research, we present a manipulation of the course of F0 and a manipulation of the duration of the topic of example (4).

(4) *MAR: e estes espaços /=TOP= por exemplo / em autores como Camilo Castelo Branco /=TOP= ou Garrett / são determinantes para a interpretação / &d / dos acontecimentos //

4.1.1. Manipulation of the course of F0

With this manipulation we wanted to verify if the rising F0 movement on the topic's second semi nucleus was related to the perception of the topic's function. The original rising F0 movement goes from 182.4Hz to 349Hz. It was manipulated in order to become a flat movement of 182.4 Hz. Hearing both audio files a fluent speaker can notice that the informational unit function is preserved even if the difference between the two audio files can be easily perceived. This manipulation shows that, in this case, the F0 contour of the second semi-nucleus isn’t related to the topic’s function.

4.1.2. Manipulation of the syllabic duration

On the other hand, the manipulation of the last tonic's and post-tonic's duration has severe impacts on the unit's function. The original length of these syllables was 0.278s and 0.212s. When they both were reduced to 0.158s (the same duration of another tonic syllable not in the nucleus), the unit is no more recognized as a topic information unit. It means that, for this topic, the duration is a relevant parameter to the perception of the unit’s function.

5. Results

PB and PE presents the three intonational forms described for Italian (Firenzuoli & Signorini, 2003), but the EP type 3 form of topic is slightly different from the forms found in BP and Italian: in EP, the rising movement and the lengthening of the second semi-nucleus starts on the last tonic syllable of the topic, and not on the last syllable of the topic. This property can be noticed on example (5).

(5) *MAR: e / o aspecto dito claramente durativo /=TOP= é aquele / que / &he / refere / a relação entre / o discurso do narrador / e / a história //

However, only one type 3 topic of our BP sample has the stress on the last syllable, which means that further data are needed in order to confirm that BP type 3’s topic is similar to Italian’s one.

A fourth intonational form of topic (type 4) was found in both Portuguese varieties. Type 4 topic presents two semi-nuclei. The first one is characterized by an extra high onset on the first tonic syllable, with high duration and sometimes intensity as well, followed by a quick pitch fall. The second semi nucleus presents a lengthening and increase of intensity on the last tonic syllable. The F0 contour on the final portion can be either flat, or slightly falling or rising. Example (6) shows a BP type 4 topic with a rising F0 movement on the last tonic syllable. Example (7) shows an EP type 4 topic ending with a falling F0 movement.

(6) *BAL: porque / &he / de certa forma /=TOP= a bancada evangélica / eles tão / muito contra / essa coisa / né //

Table 1: Example (5) measurements for topic acoustic parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>syllabic duration (ms)</td>
<td>70 149 200 181 305 258</td>
</tr>
<tr>
<td>vowel duration (ms)</td>
<td>70 60 101 75 156 170</td>
</tr>
<tr>
<td>F0 peak (Hz)</td>
<td>260 279 251 217 238 265</td>
</tr>
<tr>
<td>F0 min (Hz)</td>
<td>234 260 182 178 210 211</td>
</tr>
</tbody>
</table>

Figure 5: Brazilian Portuguese type 4 of topic
Manipulations have shown that when type 4 topics end with rising movements, the F0 movement tends to be more relevant than F0 movements, duration tends to be more relevant than F0 movements.

Type 4 form of topic seems to be the most commonly used in EP, while it is marginally used in BP.

Table 2: Example (6) measurements for topic acoustic parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>(de)</td>
<td>cer</td>
</tr>
<tr>
<td>syllabic duration (ms)</td>
<td>-</td>
</tr>
<tr>
<td>vowel duration (ms)</td>
<td>-</td>
</tr>
<tr>
<td>F0 peak (Hz)</td>
<td>-</td>
</tr>
<tr>
<td>F0 min (Hz)</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 3: Example (7) measurements for topic acoustic parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>syllabic duration (ms)</td>
<td>348</td>
</tr>
<tr>
<td>vowel duration (ms)</td>
<td>121</td>
</tr>
<tr>
<td>F0 peak (Hz)</td>
<td>328</td>
</tr>
<tr>
<td>F0 min (Hz)</td>
<td>218</td>
</tr>
</tbody>
</table>

Table 4: distribution of topic intonational forms

<table>
<thead>
<tr>
<th>Languages</th>
<th>Intonational forms of topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOP1</td>
<td>58</td>
</tr>
<tr>
<td>TOP2</td>
<td>39</td>
</tr>
<tr>
<td>TOP3</td>
<td>1</td>
</tr>
</tbody>
</table>

6. Conclusions

In summary, BP and EP present the three intonational forms found in Italian and also a fourth one that is not possible in this language. Prosodic parameters of types 1 and 2 are highly similar in all three languages, although it was found only one type 1 topic in EP. Type 3 is the least common form in BP and EP and further data are needed in order to provide a more accurate description. Type 4 topics seem to differ in BP and EP regarding the possibility of a preparation between the semi-nuclei and the F0 values of the first tonic syllable onsets (with the Brazilian variety presenting higher values than the European).

Finally, this work raises some interesting questions. Are there any functional differences between the four intonational forms of topic or are they just “intonational allomorphs” (Signorini, 2005)? Why does the distribution of BP intonational forms of topic regarding types 1, 2 and 3 resembles the Italian distribution and not the EP one?

7. Acknowledgements

This research was found by CNPq and Fapemig.

8. References

PERSPEZIONI. Firenze: FUP.


Information patterning strategies in spontaneous speech: a cross-linguistic study

Maryualê M. MITTMANN¹, Alessandro PANUNZI², Emanuela CRESTI², Massimo MONEGLIA², Heliana MELLO¹, Tommaso RASO¹

¹Universidade Federal de Minas Gerais, Faculdade de Letras, Belo Horizonte, Brazil; ²Università di Firenze, Dipartimento di Italianistica, Florence, Italy

maryuale@gmail.com, alessandro.panunzi@unifi.it, elicresti@unifi.it, moneglia@unifi.it, heliana.mello@gmail.com, tommaso.raso@gmail.com

Abstract

We present a cross-linguistic study on information patterning strategies in two romance languages: Italian and Brazilian Portuguese. The language sample comes from two comparable corpora of spontaneous speech: C-ORAL-ROM (Italian section) and C-ORAL-BRASIL. We investigate the occurrence of information units in Italian (IT) and Brazilian Portuguese (BP), thus identifying differences and similarities in the way each language organizes information. Both speech samples are annotated at the informational level according to the textual and dialogic information units established by Language into Act Theory. Results show a prevalence of compound utterances in Italian in comparison with Brazilian, and also an overall tendency in Italian to pattern information at the textual level, while Brazilian presents a more frequent use of dialogic units. These differences could be a result of cultural influences in language use.

Keywords: spontaneous speech; C-ORAL-BRASIL; C-ORAL-ROM; Language into Act Theory.

1. Introduction

In this paper we develop a cross-linguistic study on information patterning strategies in two romance languages: Italian and Brazilian Portuguese. The language sample comes from two comparable corpora of spontaneous speech: Italian section of C-ORAL-ROM (Cresti & Moneglia, 2005) and C-ORAL-BRASIL (Raso & Mello, 2012).

Our aim is to investigate the frequency of occurrence of information units in Italian (IT) and Brazilian Portuguese (BP) and to determine the most frequent information patterns in both languages. We carry out a comparison in the use and distribution of information units according to the type of interaction – monologues, dialogues and conversations (multi-dialogues) – in order to identify the differences and similarities in the way each language organizes information.

2. Theoretical framework

Language into Act Theory (Cresti, 2000) was developed for the analysis of spontaneous speech data. It states a link between prosody, the accomplishment of speech acts (Austin, 1962) the organization of information. The referring unit for the analysis of the spoken language is the utterance, defined as the linguistic counterpart of a speech act. The utterance is the shortest linguistic unit that can be pragmatically interpreted and is delimited in the speech flow by prosodic breaks that bear a conclusive value. Mostly, a prosodically terminated sequence corresponds to the performing of a single speech act. Prosody plays an essential role in the identification of utterances, since through prosody the hearer can perceive the linguistic sequences these pragmatically and prosodically autonomous sequences: the utterances.

The utterance may be prosodically parsed into two or more units, creating a prosodic pattern. The units of the prosodic pattern are associated with informational functions, through which information is patterned in the utterance. Informational Patterning Hypothesis proposes that there is a systematic correspondence between the prosodic pattern and the information pattern of an utterance (Scarnano, 2009; Cresti & Moneglia, 2010).

The relation between the prosodic pattern (Hart; Cohen; Collier, 1990) and the information pattern is established by the expression of different information functions with different prosodic profiles. Each prosodic unit corresponds, in principle, to an information unit (IU). The core of the utterance corresponds to the unit that bears the utterance’s illocutionary force. It corresponds to the Comment IU.

The comment is the necessary and sufficient unit to form an utterance. Other prosodic units correlate with different information functions, that can be either textual or dialogic. Textual IU participate to the construction of the semantic content of the utterance. Dialogic IU are devoted to the successful pragmatic performance of the utterance (e.g. to regulate the relationship between speakers).

The set of textual information units (and its correspondent tags) is the follow:

a) Comment – COM: accomplishes the utterance's illocutionary force;
b) Topic – TOP: identifies the domain of application for the illocution;
c) Appendix of comment – APC: integrates the text of the comment;
d) Appendix of topic – APT: integration of the information given in the topic;
e) Parenthesis – PAR: adds information with metalinguistic value;
f) Locutive introducer – INT: signals a change of point of view on the subsequent locution.
The dialogic functions are:

- Incipit – INP: opens the communicative channel while signals a contrastive value with the previous utterance;
- Conative – CNT: pushes the listener to take part in an adequate way in the dialogue;
- Phatic – PHA: ensures the maintenance of the communicative channel;
- Allocutive – ALL: specifies to whom the message is directed, also signaling social cohesion;
- Expressive – EXP: emotional support of the utterance;
- Discourse Connector – DCT: signals the continuity of the discourse while establishes a relation between the previous and following units.

There are two cases when one terminated sequence does not correspond to a single illocutionary value: Multiple Comments and “Stanzas”.

Multiple Comments – CMM – are a chain of Comments forming an illocutionary pattern. It is an actional model that patterns two or more illocutionary acts for the performance of one conventional rhetoric effect.

A “Stanza” (Cresti, 2009) is a terminated sequence that does not correspond to only one speech act, but to a global linguistic activity, as a result of the intention of performing an oral text, such as narratives and argumentations. It corresponds to a sequence of Bound Comments – COB – with homogeneous illocutionary forces. A “Stanza” may contain other information units forming sub-patterns.

In Language into Act Theory, the information patterning is not explained in terms of given and new information, but rather as the patterning between what conveys illocution and what carries different functions.

3. C-ORAL: spontaneous speech corpora

The main goal of both the C-ORAL-ROM and the C-ORAL-BRASIL corpora is the documentation of the diaphasic variation, which is needed to represent spontaneous speech. Therefore, besides the variation between private/familiar and public contexts and among the three interactional typologies (monologues, dialogues and conversations), the corpora belonging to the C-ORAL Projects try to document the largest variation in terms of different interaction situations, so allowing a great variation of activity and, as a consequence, of different speech acts and information structures.

As in C-ORAL-ROM corpora, C-ORAL-BRASIL transcriptions incorporate the annotation of prosodic boundaries proposed by Moneglia and Cresti (1997). The annotation scheme segments the speech flow in two distinct levels. The first level deals with the demarcation of the fundamental entity in spontaneous spoken communication (utterances). The second level refers to the internal structure of the utterance, that can be built by one single tone unit (simple utterance) or by several tone units (compound utterance) (Moneglia & Cresti, 1997; 2006).

In order to study the information structure, the corpus should be tagged regarding information functions. Unlike the tagging of part-of-speech, for which there are already many automatic tools, the tagging of information units is done manually. The samples from IT and BP analysed in this study received informational tagging, using the set of informational units proposed by the Language into Act Theory and the Informational Patterning Hypothesis.

4. Methods

The samples come from the informal sections of C-ORAL-ROM Italian and C-ORAL-BRASIL corpora, selected for a strict comparison with each other. Each sample is detailed in the sections below.

Data were extracted through IPIC, a theoretically-bound XML Database designed for the study of linear relation among Informative Units in spoken language corpora (Panunzi & Gregori, 2012 and in this volume). The database is available for online research and can be accessed at http://lablita.dit.unifi.it/ipic/.

4.1 Brazilian Sample

The selection of texts for the Brazilian Portuguese sample followed a set of criteria adopted to ensure a high quality database to perform information structure studies. At the same time, the same basic structure of the entire C-ORAL-BRASIL informal corpus was preserved (Raso & Mello, 2009).

The BP sample, presented in Table 1 below, has 31318 words, 5483 terminated sequences and 9825 prosodic/information units.

<table>
<thead>
<tr>
<th>Type</th>
<th>Male/Female</th>
<th>Words</th>
<th>Utterances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conversations</td>
<td>15/9</td>
<td>9774</td>
<td>2039</td>
</tr>
<tr>
<td>Dialogues</td>
<td>6/8</td>
<td>11331</td>
<td>2450</td>
</tr>
<tr>
<td>Monologues</td>
<td>7/10</td>
<td>10213</td>
<td>994</td>
</tr>
<tr>
<td>Total</td>
<td>28/27</td>
<td>31318</td>
<td>5483</td>
</tr>
</tbody>
</table>

Table 1: Features of BP sample

4.2 Italian Sample

In order to be as much comparable with the BP sample as possible, the Italian sample maintains the same proportion between dialogic and monologic typologies. The texts chosen present a large variety of activities performed by the speakers during the recording sessions.

The Italian sample contains 29414 words, 5276 terminated sequences and 11517 prosodic/information units, as showed in Table 2 below.
Table 2: Features of IT sample

<table>
<thead>
<tr>
<th>Type</th>
<th>Male/Female</th>
<th>Words</th>
<th>Utterances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conversations</td>
<td>9/11</td>
<td>10141</td>
<td>1986</td>
</tr>
<tr>
<td>Dialogues</td>
<td>5/13</td>
<td>12435</td>
<td>1939</td>
</tr>
<tr>
<td>Monologues</td>
<td>9/7</td>
<td>11632</td>
<td>1351</td>
</tr>
<tr>
<td>Total</td>
<td>23/31</td>
<td>34208</td>
<td>5276</td>
</tr>
</tbody>
</table>

For more detailed information about the construction of the Brazilian mini-corpus and its comparable Italian counterpart see Mittmann; Raso (2012).

5. Results

The first important difference to point out regards the distribution of simples versus compound utterances in both samples. In Brazilian shows 71.4% of simple utterance in conversation, 73.6% in dialogue and 55.5% monologue, in Italian these measurements are, respectively, 66.6%, 68.2% and 39.1%.

The prevalence of compound utterances in Italian in comparison with Brazilian is statistically significant (chi-square=52.848 – p<0.0001). Furthermore, in Italian information is more likely to be patterned at the textual level, with high occurrence of compound Utterances with only textual IU (44% of all compound Utterances).

This hypothesis is strengthened by the fact that the number of textual compound utterances is also higher in Italian. While Brazilian shows a percentage of 11.00%, 9.2% and 31.8% of textual compound utterances respectively for conversations, dialogues and monologues, Italian presents 20.0%, 16.00% and 58.9%.

The distribution of illocutionary units shows that the greatest part of illocutionary units for conversations and, specially, dialogues is the Comment unit. In monologues, Bound Comments have a more important role, which is expected, since monologues give rise to more complex and more “textual” discourse, while conversations and dialogues are interactions more action grounded, and therefore present more a greater number and variety of speech acts and dialogic units.

Brazilian shows a relevant use of illocutionary patterns, represented in the Graphic by the Multiple Comments (CMM). Graphic 1 shows the distribution of illocutionary units in both samples.

In Italian there is a strong tendency to organize information in Topic-Comment structures, much more often than in Brazilian Portuguese. Distribution of textual units is showed in Graphic 2.

The only textual unit more frequent in BP is the Locutive Introducer (INT). In Italian, the distribution of INT does not present much variation between conversations, dialogues and monologues, while in the Brazilian the number of INT in monologues is much higher than in the other typologies. This indicates the higher use of reported speech in BP monologues, since reported speech is almost always introduced by an INT unit.

Brazilian shows a relevant use of illocutionary patterns, represented in the Graphic by the Multiple Comments (CMM). Graphic 1 shows the distribution of illocutionary units in both samples.

In Italian there is a strong tendency to organize information in Topic-Comment structures, much more often than in Brazilian Portuguese. Distribution of textual units is showed in Graphic 2.

The only textual unit more frequent in BP is the Locutive Introducer (INT). In Italian, the distribution of INT does not present much variation between conversations, dialogues and monologues, while in the Brazilian the number of INT in monologues is much higher than in the other typologies. This indicates the higher use of reported speech in BP monologues, since reported speech is almost always introduced by an INT unit.

Graphic 1: Distribution of illocutionary units in IT and BP

Graphic 2: Distribution of textual units in IT and BP

Comparing Brazilian and Italian with respect to all the dialogic units, we note that Brazilian uses much more Expressives and Allocutives, while Italian uses much more Incipits and Conatives. When we look at the distribution of dialogic units regarding its position inside the utterance, we notice that the Expressives are very often employed to open the utterance and/or to take the turn. In Italian, those functions are mostly performed by Incipits.

Allocutives and Expressives are signs of social cohesion in discourse, while Incipits signal the speaker's opposition with respect to the previous utterance. It is likely that in Brazilian culture the Incipit is perceived as
an aggressive way to take the turn or begin the utterance. For this reason, Brazilian tends to prefer Expressives to play this role.

It is important to emphasize that dialogic information units function governing the interaction. Dialogic information units are strongly linked to the interaction (and not the semantic content of the utterance). Therefore, they are sensitive to cultural nuances and, for this reason, they are a good way to investigate how linguistic features can be affected by cultural idiosyncrasies.

6. Conclusion

The differences observed in the data suggest cultural influences in language use, especially if we consider the distribution of dialogic units. These differences could be interpreted as a result of cultural influences in language use, since dialogic IU like Allocutives and Expressives are signs of social cohesion in discourse. However, a more qualitative look into the data is needed, in order to assure that such differences do not derive from sampling incompatibilities or problems in the information annotation.

Cross-linguistic studies are very valuable, in the sense that through the analysis of different languages we can observe which features are intrinsic to speech as a universal communicative medium and which are specific of each language. Individualizing what is specific to each language is necessary to develop and implement appropriate teaching strategies. The presence of comparable corpora and the study of the information structure in a contrastive perspective provide many useful elements for L2 teaching. The pragmatic perspective, often invoked in education, still lacks appropriate tools of research. Corpora such as C-ORAL-ROM and the C-ORAL-BRASIL and a theoretical perspective as Language into Act Theory can provide excellent tools to repair this deficiency.

7. Acknowledgements

This work was developed as part of the collaboration agreement between the LABLITA Laboratory of the Italianistic Department of University of Florence (UNIFI) and LEEL Laboratory of the Federal University of Minas Gerais (UFMG).

C-ORAL-BRASIL project was funded by the National Council for Scientific and Technological Development (CNPq), the Foundation for Research Support of Minas Gerais (Fapemig), the Faculty of Letters of the Federal University of Minas Gerais (Fale/UFMG) and by Santander Bank.

We thank Ida Tucci for the annotation of the Italian corpus and her help in the annotation revision of the Brazilian mini-corpus. We also thank Lorenzo Gregori for his work on the IPIC database.

8. References


Del Cerro, pp. 57–90.


O Apêndice de Comentário no Português do Brasil: uma análise baseada em corpus

Cássia Jacqueline Fernandes OLIVEIRA
Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6.627 - 31270-901 - Belo Horizonte - Brasil
cassiajfo@gmail.com

Resumo

Esse estudo consiste em uma análise baseada em corpora da unidade informacional de Apêndice de Comentário (APC) e teve como objetivo estabelecer uma análise contrastiva dessa unidade no PB e no italiano. Essa pesquisa é sustentada pela Teoria da Língua em Ato (Cresti, 2000), segundo a qual um enunciado é definido como sendo a menor unidade possível de interpretabilidade pragmática. As fronteiras entre enunciados e suas unidades internas são delimitadas pela entonação. Uma das possíveis unidades internas é o APC, que estabelece uma relação de integração com a unidade ilocucionária. Essa investigação foi conduzida em um subcorpus de PB com 20 textos do C-ORAL-BRASIL e em um subcorpus de Italiano com 20 textos do C-ORAL-ROM. A pesquisa demonstrou que no italiano há uma presença maior de unidades terminadas complexas, enquanto no PB apresenta mais enunciados simples. A porcentagem de APC em Italiano nos enunciados simples é quase 50% superior a do PB. Nas estrofes dos monólogos, os APC do PB superam os do italiano. Do ponto de vista entonacional, não aparecem diferenças entre as duas línguas. Informacionalmente, as proporções entre as várias funções nas duas línguas são perfeitamente comparáveis.

Keywords: estrutura informacional; atos de fala; Apêndice de Comentário.

1. Introdução

Ao longo da história, vários pesquisadores se propuseram a estudar a linguagem, principalmente a escrita, deixando em segundo plano a fala. Ainda que semelhante em alguns aspectos, sabe-se que cada uma delas possui as suas especificidades e que analisar a fala pela lente da escrita é um equívoco.

Sem dúvida, uma questão muito discutida pela linguística atual é compreender como o falante organiza a informação na fala, isto é, como se organiza a sua estrutura informacional. A Teoria da Língua em Ato (Cresti, 2000), que serviu de arcabouço teórico para esse estudo, foi desenvolvida para lidar com essas questões, inserindo o estudo da estrutura informacional dentro daquele dos atos de fala (Austin, 1962).

2. A Teoria da Língua em Ato

A Teoria da Língua em Ato (Cresti, 2000) fundamenta-se em um estudo empírico da fala espontânea realizado pelo LABLITA (Laboratorio Linguistico del Dipartimento di Italianistica dell’Università di Firenze).

A fala espontânea, de acordo com essa perspectiva teórica, é considerada como toda a produção linguística sonora dialogada ou monologada em situação natural, realizada livremente, em contextos e situações comunicativas naturais, formais ou informais.

“À imposição de um molde de segmentação do texto escrito sobre o discurso falado leva o pesquisador a tratar os dados de fala de forma problemática, enviesando especialmente a análise das relações sintáticas no discurso falado. Apesar disso, poucos pesquisadores atentam a este fato e se dão conta da relevância de preservar os aspectos entonacionais da fala em suas transcrições”

(Cresti apud Mittmann, 2012).

Na escrita, todavia, segundo Moneglia (2011) é clara a identificação de unidades linguísticas maiores do que a palavra (unidades da estrutura argumental, sentenças, orações, termos nucleares e dependentes), pois a língua escrita pode ser tranquilamente segmentada de acordo com critérios sintático. Na fala, ao contrário, não é possível utilizar estes mesmos critérios para identificar unidades de referência. Evidências de corpora orais têm mostrado que aproximadamente 30% dos enunciados não apresentam um verbo e não podem ser analisadas conforme parâmetros sintáticos empregados facilmente na escrita.

Em princípio, a unidade linguística que se percebe de maneira mais natural é o turno dialógico. Entretanto, segundo Cresti (2000), o turno dialógico não pode ser considerado como unidade fundamental de referência do discurso falado, porque os turnos apresentam uma ampla variação, podendo ser compostos de apenas uma palavra ou interjeição, ou mesmo de uma longa exposição. O conceito de turno é resultado de uma interpretação antes cognitiva do que linguística.

A Teoria da Língua em Ato parte do princípio de que a unidade linguística da fala deve corresponder à unidade fundamental da atividade comunicativa, já que é esta atividade que ‘sustenta’ a fala. A unidade linguística fundamental da fala deve corresponder à unidade fundamental da atividade comunicativa: o ato de fala (Austin, 1962). Partindo do princípio de que a fala espontânea consiste na execução de ações, delimitar a unidade de referência da fala deve corresponder a identificar, no fluxo da fala, as sequências linguísticas que se apresentam como suficientes e autônomas do ponto de vista pragmático, isto é, as entidades linguísticas que veiculam ações. Estas unidades são identificadas com o componente linguístico, o ato locutório, do ato de fala, conforme a

Esta afirmação fundamenta-se na hipótese de que seja possível estabelecer uma equivalência entre unidades do domínio das ações humanos (atos) e unidades linguísticas (enunciados). Assim, o enunciado é tido como a “contraparte linguística da ação”; isto é, o ato locutório é a contraparte linguística do ato ilocutório, e é interpretável pragmaticamente em autonomia. Isso significa, entre outras coisas, que um enunciado não precisa necessariamente possuir um verbo, e pode, inclusive, ser composto por uma única interjeição, desde que, entoado de maneira a cumprir uma ilocução. Dessa forma, a identificação dos enunciados se realiza através de uma quebra entonacional percebida como conclusiva. Isso significa que uma unidade de enunciado (ou a única se o enunciado for simples) deve ser uma unidade de raiz (o comentário) capaz de veicular autonomia pragmática. Esse princípio baseia-se na teoria perceptiva da entonação (‘t Hart, Collier & Cohen, 1990), acarretando uma relação biunívoca entre entonado e ilocução. A cada enunciado, ou seja, a cada unidade mínima de significado pragmático, corresponde-se uma única ilocução, uma intencionalidade do falante.

Para a Teoria da Língua em Ato, as unidades informacionais são identificadas no enunciado através de três critérios distintos: o critério funcional (função exercida pela unidade no enunciado), o critério entonacional (perfil entonacional característico de cada unidade) e o critério distribucional (posição da unidade no enunciado). Dessa forma, a junção desses três critérios possibilita a identificação das unidades informacionais da fala.

Segundo a Teoria da Língua em Ato há dois tipos de unidades informacionais: as unidades textuais e as unidades dialógicas. As unidades informacionais textuais são aquelas que compõem o texto do enunciado propriamente dito. Dentre elas encontramos as unidades de Comentário (COM), de Tópico (TOP), de Apêndice de Comentário (APC), de Parentético (PAR), de Introdutor Locutivo (INT), de Apêndice de Tópico (APT) e a Unidade de escansão (SCA). As unidades informacionais dialógicas ou não textuais, por sua vez, são aquelas que não contribuem para a constituição semântica de um enunciado, mas dedicam-se ao cumprimento pragmático desse enunciado sendo dirigidas ao interlocutor. São elas: Incipitário (INP), Conativo (CNT), Conector Dialógico (DCT), Fático (PHA), Alocutivo (ALL) e Expressivo (EXP).

A Unidade de Comentário é a mais importante de todas as unidades, pois é a única necessária e suficiente para execução de um enunciado. Sua função é a de realização da força ilocucionária, ou seja, a de cumprir um ato de fala. Entonacionalmente é tida como uma unidade prosódica de raiz que varia conforme o valor ilocucionário; isto é, é interpretável pragmaticamente em autonomia e possui sempre um núcleo, o qual carrega o valor funcional da ilocução. Distribucionalmente pode estar em qualquer posição no enunciado e é com relação a ela que é definida a posição das outras unidades.

A unidade de Tópico é a unidade textual cuja função é especificar no texto do enunciado o domínio de relevância ao qual a força ilocucionária se refere; isto é, o campo de aplicação da força ilocucionária do comentário. Ela tem caráter opcional e é subordinada melodicamente ao comentário, não sendo interpretável autonomamente.

2.1 A unidade de Apêndice de Comentário (APC)

A unidade de APC é por definição uma unidade de integração textual. A maior parte das expressões que são usadas funcionalmente como unidade de APC corresponde a um conteúdo “vazio” ou a um conteúdo genérico do ponto de vista semântico. Funcionalmente, o Apêndice integra textualmente as unidades de Comentário (COM), Comentário Ligado (COB), Comentários Múltiplos (CMM). Entonacionalmente é uma unidade tonal sem foco, com uma F0 sempre mais baixa do que a unidade da qual é apêndice, sempre com perfil nivelado ou descendente e intensidade baixa (Cresti 2000; Ulisses 2009, 2009b, 2010). Distribucionalmente deve suceder a unidade informacional de Comentário. É tida como uma unidade de sufixo.

Ex: *REG: omitir /=COM= só /=APC

2.1.1. A definição de APC e seus critérios de ocorrência

No subcorpus de PB foram localizadas 112 unidades informacionais de APC, enquanto que no italiano o número foi de 243 ocorrências. A análise prosódica dos APC encontrados em ambas as línguas revelou a existência de apenas uma forma entonacional (Cresti, 2000; Firenzuoli, 2003) como sendo a de um perfil nivelado e descendente para essa unidade textual.

Observou-se que embora seja uma unidade textual (e ocupe a quarta posição entre as demais unidades textuais no PB, com 10% de ocorrências, sendo mais frequente nos monólogos), a unidade de APC tem como função apenas integrar a unidade de COM, já que não serve de âmbito para aplicação da força ilocucionária como a unidade de TOP, não possui função metalinguística como o PAR, nem introduz uma metailocução como o INT, tampouco possui autonomia pragmática como a unidade de COM.

Funcionalmente deve estar posicionada após a unidade da qual faz a integração, a unidade de COM (CMM e COB) e pode exercer a função de informação tardia, repetição, retomada textual ou preenchimento (Tucci, 2006).
Na tipologia conversação, a informação tardia foi a classificação informacional que mais se destacou com 37,5% de ocorrências. Em seguida encontramos os preenchimentos (34,4%), as retomadas textuais (15,6%) e, por último, as repetições (12,5%). Nos diálogos, há a predominância dos preenchimentos (47,2%), as informações tardias (36,1%), repetições (11,1%) e retomadas (5,5%). Os monólogos procedem de maneira semelhante aos diálogos. Os preenchimentos destacam-se com 45,2% de ocorrências, as informações tardias com 35,7%, as repetições com 16,6% e as retomadas textuais com 2,4%.

Em âmbito geral, notou-se que as unidades de APC no PB desempenham mais a função de preenchimento, com 43% de ocorrências sobre o total de APC, seguido pela informação tardia com 36%, depois as repetições com 14% e, por último, as retomadas textuais incidindo em apenas 7% sobre o total de APC na amostra.

Entonacionalmente, são características da unidade de APC possuir as médias de F0 e intensidade inferiores às médias da unidade de COM e abaixamento do tom de voz. Nessa unidade, nenhum movimento funcional é encontrado, uma vez que ela visa a integração de novas estruturas linguísticas de modo a realizar uma expressão semântica, uma correção ou mesmo uma reestruturação de um enunciado. Do total de 112 APC localizados na amostra, 87% possuem curva analisável e apenas 13% não-analisável.

Distribucionalmente, a unidade de APC localiza-se após a unidade de COM, tendo como função informacional mais comum, nas conversações, a informação tardia, (37,5%), os preenchimentos (25%) e as repetições e retomadas textuais juntas (18,8%). Quanto aos padrões ilocucionários (CMM), a função mais recorrente é a de retomada textual (6%), repetições e preenchimentos totalizam (6%). Nas estrofes (COB), a única função informacional encontrada foi a de preenchimento, com (6%) de ocorrências. Quanto aos diálogos, após a unidade de COM, os preenchimentos são mais recorrentes (41,7%), depois aparecem as informações tardias (33,3%), as repetições (8,3%) e as retomadas textuais (2,7%). Nos padrões ilocucionários, os preenchimentos se destacam com (5,5%), em seguida com o mesmo valor estão as repetições (2,7%) e informações tardias (2,7%). Nas estrofes, a única função exercida pelo APC é de retomada textual (2,7%). Quanto aos monólogos, após a unidade de COM, as informações tardias e preenchimentos possuem o mesmo valor percentual (26,2%) cada uma. Em seguida estão as repetições com (16,7%) e as retomadas textuais com apenas (2,4%). Nos padrões ilocucionários (CMM), a maior função desempenhada pelo APC é de preenchimento (7,1%), depois de informação tardia (2,4%). Nas estrofes (COB), a categoria informacional mais saliente é a de preenchimento (11,9%), seguida apenas pela informação tardia (7,1%). Distribucionalmente, ainda nos atentamos para o fato de que entre a unidade de Comentário e a unidade de APC podem aparecer intercaladas as seguintes unidades: alocutivos, fáticos, conativos, expressivos e parentêtics; todavia, em nosso estudo, só foram verificadas as presenças das unidades de alocutivo, parentêtic e fático.

Observou-se, ainda, que há determinados contextos em que duas outras unidades podem ocupar a mesma posição do APC, com um perfil entonacional às vezes parecido, e serem confundidas com o APC. Tratam-se das unidades de PAR e COB. Para distinguir uma unidade de PAR de uma de APC, primeiramente, deve-se observar que se se retirar a unidade de APC percebe-se, frequentemente (do ponto de vista prosódico), a falta de algo para a realização completa do enunciado; o mesmo não ocorre quando da eliminação da unidade de PAR. Segundo, a unidade de PAR possui sempre valor modal ou, pelo menos, constitui uma intervenção metalinguística, cujo ponto de vista é externo àquele do resto do enunciado.

Quanto aos COB em posição de um possível APC, as principais pistas para decidir se tornam o valor cognitivo de ‘novo’ e as medidas de F0 e intensidade.

Outra situação que merece ser mencionada é a da coda. A coda ocorre quando há uma unidade tonal que parece ter todas as características prosódicas de um APC, mas com características informacionais distintas. Isso pode acontecer quando temos um COM cujo foco funcional é à esquerda e cujo conteúdo locutivo se estenda por várias sílabas de coda. Essa situação faz com que seja impossível ou pelo menos não natural a realização do COM, sem que se produza uma quebra entre o foco e o restante do conteúdo silábico; assim sendo, não há como afirmar que depois da quebra haja um APC, já que essa quebra é praticamente obrigatória; trata-se, portanto, de uma coda que produz uma unidade escansionada (SCA) à direita e não, como é frequente, à esquerda do foco funcional.

Morfossintaticamente, 67% do total de APC analisados nesse estudo são construções sintáticas e 33% são expressões, sendo os ADV os que mais se sobressaem com função de APC em qualquer uma das tipologias. Esse resultado já era esperado, porque é sabido que dentre as várias funções do ADV está a de determinar um fato, ampliando a informação nele contida, função essa desempenhada pelo APC. Outras categorias são bem mais raras, em PB.

Interessante destacar, ainda, a distinção estabelecida entre uma sequência de dois apêndices e um apêndice escansionado. Enquanto o primeiro apresenta um perfil prosódico concluído , o segundo, não.

3. Análise Contrastiva: APC no PB versus APC no Italiano

A fim de estudar a estrutura informacional em uma perspectiva interlinguística, buscou-se analisar o
comportamento da unidade de APC no PB e no italiano. Como o subcorpus brasileiro é altamente acional, para se estabelecer um parâmetro de comparação foi necessário manter a mesma proporção entre as tipologias diálogo e monólogo, e maximizar o número de atividades realizadas pelo falante no momento da interação.

Observamos que no PB há 82,7% de enunciados nas conversações, 85,6% nos diálogos e 66,6% nos monólogos. No italiano o percentual é parecido, à exceção dos monólogos. Nas conversações há 83,7% de enunciados, nos diálogos 83,4% e nos monólogos, 70,5%. Esses resultados nos permitem afirmar que no italiano os textos dialógicos se comportam de maneira similar aos textos do PB, havendo proporcionalidade de enunciados quando da comparação. Os textos monológicos, entretanto, apresentam medidas muito diferentes, permitindo-nos aventar a hipótese de que as pequenas diferenças encontradas entre as duas tipologias no subcorpus do PB são devidas à presença de duas conversações em que os falantes não realizam qualquer atividade levando, portanto, algumas medidas na direção dos monólogos.

Quanto aos monólogos no italiano, a diferença fundamenta-se no fato de haver nessa tipologia tanto menos estrofes e padrões ilocucionários, quanto mais enunciados.

É interessante notar que a diferença mais significativa entre uma língua e outra está no fato de que em termos percentuais o italiano tem menos enunciados simples que o PB. Enquanto no PB há 71,4% de enunciados simples nas conversações, 73,6% nos diálogos e 55% nos monólogos, no italiano essas medidas são, respectivamente, 66,6%, 68,2% e 39,1%, o que nos conduz ao fato de que no italiano há mais enunciados complexos do que no PB. Esta hipótese é reforçada pelo fato de que as unidades textuais também são superiores no italiano.

Quanto aos enunciados complexos com unidades textuais, enquanto na amostra brasileira o percentual é de 11%, 9,2% e 31% de enunciados complexos com unidades textuais para conversação, diálogo e monólogo, no italiano, respectivamente, encontramos 20%, 16% e 58,9%. O mesmo acontece com os padrões ilocucionários. No italiano, os padrões ilocucionários são mais comuns, enquanto os padrões ilocucionários simples são mais encontrados no PB. A diferença quanto as estrofes não parece muito significativa.

Outra diferença interessante entre os dois subcorpus diz respeito à inversão da distribuição dos APC nas três tipologias. Enquanto no PB há mais APC nos monólogos e menos nas conversações, no italiano os APC são mais comuns nas conversações e menos nos monólogos.

Enquanto no italiano há 243 ocorrências de APC sobre 1018 unidades terminadas com unidades textuais, no PB encontramos apenas 112 APC sobre 1012; isto é, a unidade de APC é muito mais recorrente no italiano, pois apresenta mais que o dobro (58%) dos APC encontrados no PB.

Em relação aos enunciados complexos com unidades textuais, o que se verifica é que há 3,1% a mais de ocorrência de APC, no italiano, na tipologia diálogo (conversação e diálogo) e nos monólogos apenas 0,8%. Nos padrões ilocucionários há mais 2,8% de ocorrências de APC no italiano do que no PB nos diálogos (conversações e diálogos), e 4%, nos monólogos. E em relação às estrofes, no italiano há mais 4,2% de ocorrências de APC do que o PB nos diálogos (conversações e diálogos), e 0,7%, nos monólogos.

Em síntese, o italiano mostra uma presença muito maior de unidades terminadas complexas, enquanto o PB apresenta muito mais enunciados simples. Esse é um aspecto muito interessante para a comparação entre as duas línguas. Mesmo considerando como baseline somente os enunciados complexos, a porcentagem de APC em italiano é quase 50% a mais do que em PB. O maior número de APC se deve principalmente ao padrões ilocucionários, mas também aos enunciados. Não há diferença quanto às estrofes. É interessante notar que nas estrofes dos monólogos, quando a fala necessariamente se complexifica, os APC do PB superam os do italiano nos monólogos. Do ponto de vista entonacional, não aparecem diferenças entre as duas línguas. Do ponto de vista informacional, as proporções entre as várias funções nas duas línguas são perfeitamente comparáveis. Do ponto de vista morfossintático, a comparabilidade entre as duas línguas é estrita.

Assim sendo, a pesquisa confirma a análise da fala espontânea do PB, com base em uma teoria elaborada a partir do italiano; comprova em detalhes as características informacionais, entonacionais e morfossintáticas da unidade de APC; observa que essa unidade é menos presente no PB do que em italiano, e que essa menor presença se justifica pelo fato de, em geral, a estrutura da fala em italiano aparecer mais rica de unidades textuais do que a fala em PB (à exceção do INT).

4. Referências


Subordinazione, coordinazione, giustapposizione. Firenze: Cesati.
Intonation and information packaging: evidence from Brazilian Portuguese

Lílian Teixeira de SOUSA
Unicamp
Campus Universitario Zeferino Vaz SN, Cidade Universitaria – 13083-970 – Campinas, Brasil
liliantsousa@gmail.com

Abstract

Intonation is generally understood to be related to information structure. Many scholars claim that new information is characterized by high tone (H), whereas given information is characterized by low tone (L). For Kohler (2004) and Baumann (2008), the NEW/GIVEN distinction do not explain entirely the relation between intonation and information packaging in German. According to them, it is also relevant to distinguish degrees of givenness. Taking into account the relevance of prominence relations for information packaging, my aim in this paper is to investigate which type of accentuation is used to indicate given/new information in Brazilian Portuguese (BP) as well as to determine whether degrees of givenness play a central role in intonation for the language in question. Since the contrast-emphasis paradigm is important to the Focus-Background dimension, an additional goal is to compare the interaction between intonation and contrast-emphasis. The analyses have shown relevant relations, on the one hand between falling contours (HL, >HL, LHL) and contrast and, on the other hand, between rising contours (LH, >LH, HLH) and emphasis. As for the degree of givenness, I have not observed a one-to-one relationship between degree of givenness and intonation.

Keywords: information structure; degree of givenness; intonation.

1. Introduction

Intonation is said to play a role in information packaging in many different languages. English and German, for instance, exhibit similar intonation patterns: high tones (H) indicate new referents in discourse whereas low tones (L) indicate given referents. Some scholars (Kohler, 2004; Baumann, 2008) have distinguished degrees of givenness in their experiments, gathering interesting data. In order to establish the relevance of prominence relations for information packaging, many experimental studies have been conducted in attempt to find similar patterns in other languages. Since studies on intonation and information packaging in Brazilian Portuguese (BP) are scarce, my aim in this paper is to investigate which type of accentuation is used to indicate given/new information in this language and also to determine whether degrees of givenness play a role in intonation in BP.

Since the contrast-emphasis paradigm is important to the Focus-Background dimension, an additional goal is to investigate the interaction between intonation and contrast-emphasis.

This paper is organized as follows: in Section 2, I summarize the literature on intonation and information packaging. Section 3 describes the methodology applied in this study. In Section 4, I examine the relationship between intonation and information packaging in BP. In section 5, I conclude the paper.

2. Intonation and Information Packaging

Understood as being pitch variation in the course of an utterance, intonation is therefore related to information structure. Many scholars agree that new information is marked by high tone (H) whereas given information is marked by low tone (L). However, Kohler (2004) and Baumann (2008) have argued that the NEW/GIVEN distinction do not explain entirely the relation between intonation and information packaging in German. In this language, it is also important to distinguish degrees of givenness. Baumann (2008) discusses different aspects of information structure which have been confined in the literature to the concept of givenness (New). According to him, between many issues and terminology found in the literature, three basic dimensions of information structure are often mentioned: 1) The division between what the utterance is about and what comments on it – Theme-Rheme; 2) the division of an utterance into an informative part and an uninformative (or newsworthy) part – Background-Focus; and 3) the cognitive representation of the referent or proposition in the interlocutor’s mind – Given-New. The first two dimensions are relational in nature and apply to the level of sentence or utterance, whereas the third dimension is non-relational and applies to the level of discourse (ex. (42) Baumann, 2008):

(1) A: What about John?
   B1: [My sister and me] [are going to visit] [him]
   Theme        Rheme
   Focus
   Background
   B2: [My sister and me are going to visit] [him]
   New          Given   New          Given

Considering the pragmatic partitioning of an utterance into a focus-background structure, Baumann argues that new information always occurs in the focus part, whereas given or accessible information can occur in the focus or background. Baumann also observes that accentuation does not depend only on the degree of activation. Baumann (2008: 99) says “if a speaker wishes to present a constituent as particularly newsworthy, s/he can highlight this constituent irrespective of its activation status”. This happens in contrastive utterances, in which given information may be focused using a particularly salient accent involving a pitch higher than the speaker’s
topline.

Authors like Pierrehumbert & Hirschberg (1990) and Kohler (2004) establish some contour patterns related to information packaging for English and for German, respectively. These patterns are summarized in the following tables:

<table>
<thead>
<tr>
<th>Pitch Accent</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>H*</td>
<td>New</td>
</tr>
<tr>
<td>L+H*</td>
<td>Addition of a new value</td>
</tr>
<tr>
<td>L*+H</td>
<td>Modification of Given</td>
</tr>
<tr>
<td>H+H*</td>
<td>Accessible</td>
</tr>
<tr>
<td>L*</td>
<td>Given</td>
</tr>
<tr>
<td>No accent</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Contour patterns related to information packaging according to Pierrehumbert & Hirschberg (1990)

<table>
<thead>
<tr>
<th>Pitch Accent</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>L+H*/L*+H</td>
<td>Emphasis (new information)</td>
</tr>
<tr>
<td>H* (Medial Peak)</td>
<td>New</td>
</tr>
<tr>
<td>H+L*/H+H*</td>
<td>Accessible or Given</td>
</tr>
</tbody>
</table>

Table 2: Contour patterns related to information packaging according to Kohler (2004)

For Pierrehumbert & Hirschberg (1990), an accent on a referring expression contributes to the perceived information status of the referent. The H* pitch accent is said to convey new information. The L+H* has contrast as its central meaning. The H+H* predicates what is mutually accessible to speaker and listener.

In the same line, Kohler (2004) establishes a relation between meaning and categorial change from early to medial peak and between meaning and a more gradual change from medial to late peak. Early peaks tend to denote established facts or end of an argument. Medial peaks usually indicate a newly introduced fact or the beginning of a new argument. Late peaks add a paralinguistic value to the information expressed, e.g. surprise or incredulity.

Intonational studies on BP have been not referring to degree of givenness, but rather to the focus/background relation. Fernandes (2007) claims that focused elements may have the same accent pitch which they generally receive in a neutral context (H*+L versus L*+H) or they may have the same tonal combination which they would receive in a neutral context (L*+H).

3. Methodology

3.1 On recording

For this study, I have recorded four native male speakers of BP, aged 18 to 30 years old, in an interactional context. To record the speakers, I used a game in which two speakers had to indicate people suspected of a crime, taking into account information available in a set of statements. Each speaker had a different set of statements containing distinct information, e.g., suspect 1 claiming to have been with suspect 2 in the library at the time the crime, and suspect 2 claiming to have been alone in the living room.

3.2 Degree of Givenness

I have considered three degrees of givenness, based on their cognitive status:

1) new or inactive: mentioned for the first time;
2) newsworthy or semi-active;
3) given or active.

The way the referents in the statements were restated by the speaker was used to classify the degree of givenness. If the referent was repeated, the information was considered given; if the speaker used a pronoun or synonym, the information was considered newsworthy. Referents not found in the statements were the only information considered new. Although the referents in this game were controlled, the speakers were able to produce spontaneous sentences.

At the end of the experiment, I selected 59 declarative sentences, which were later analysed with the Praat software (Boersma & Weenink, 2010, version 5.2.11). In those sentences, I distinguished 34 given referents, 44 newsworthy referents and 10 new referents. Since the number of utterances containing new information was significantly lower than both the ones containing given and newsworthy information, it was not possible to proceed to a detailed statistical analysis.

Since the concepts of newness and givenness of information are generally related to the Focus/Background terminology, I have analysed such interactions in the light of the concepts of contrast and emphasis. I considered contrastive those referents which were used (by the speaker) to correct something in the previous speech and emphasis all highlighted referents with no corrections.

3.3 Dato System

<table>
<thead>
<tr>
<th>Pitch Accent</th>
<th>Level Contours</th>
</tr>
</thead>
<tbody>
<tr>
<td>L*</td>
<td>L</td>
</tr>
<tr>
<td>H*</td>
<td>H</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dynamic Contours</th>
</tr>
</thead>
<tbody>
<tr>
<td>L+H*</td>
</tr>
<tr>
<td>L*+H</td>
</tr>
<tr>
<td>H+H*</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Boundary tones</th>
<th>Boundary tones</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>H</td>
<td>H</td>
</tr>
</tbody>
</table>

Table 3: ToBI contours and DaTo contours
The DaTo intonational annotation system (Dynamical Tones of Brazilian Portuguese) was used for my analysis. This system, devised by Lucente (2008), describes focus in intonation taking into account the notion of dynamical contour. The alignment in this approach was formulated according to a synchrony between phonation and articulation. The relation between ToBI contours and DaTo contours are in the Table 3.

4. Brazilian Portuguese (BP) information packaging

On the F0 analyse, I observed the contour type and pragmatic function of the aligned sentential elements. At the end of the data analysis, I distinguished the referents as either new, accessible or given. The number and percentage of referents in relation to degree of givenness are shown in Table 4.

<table>
<thead>
<tr>
<th>New</th>
<th>Accessible</th>
<th>Given</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>%</td>
<td>No.</td>
</tr>
<tr>
<td>LH</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>&gt;LH</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>HLH</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>&gt;HL</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>H</td>
<td>11</td>
<td>32</td>
</tr>
<tr>
<td>LHL</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>34</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 4: Number and percentage of new, accessible and given referents

I observed in the percentage data that LH is more frequently used to indicate given information, HL is more often used for accessible information and H for new information. The low frequency of new information in this corpus (only 10 utterances) does not allow for a statistical analysis. However, the number of occurrence indicates that falling contours (HL, >HL, LHL) are somewhat associated with information structure, because they tend to be connected to accessible information. HL and H contours are related to new information, in line with many other studies (Kohler, 2004; Yule, 1980; Pierrehumbert & Hirschberg, 1990).

Figure 1 shows the items ‘menino’ (boy) and ‘Rodrigo’ (proper name) as accessible and given information, respectively. Both items correspond to accessible/ given information updated in the utterance context. Since these items indicated no contrast or correction, I have considered them to be emphatic.

Figure 2 exhibits HL contour on the contrastive given referent “biblioteca” (library) and >LH on the proper names “Rodrigo” and “Alaide”, which were considered emphatic given referents. These data indicate the use of >LH contour to reintroduce a given referent in the discourse.
HL was the most frequent contour applied to express contrast. Moreover, there were some data which exhibit LH contour with contrastive meaning. Figure 4 is an example of LH contour with contrastive meaning.

As previously mentioned, the distinction between emphatic and contrastive referents was made with specific accent types. The number and percentage of the referents in relation to emphasis versus contrast are in Table 5.

<table>
<thead>
<tr>
<th></th>
<th>Emphasis</th>
<th>Contrast</th>
<th>Ambiguous</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
<td>No.</td>
</tr>
<tr>
<td>LH</td>
<td>14</td>
<td>26</td>
<td>-</td>
</tr>
<tr>
<td>&gt;LH</td>
<td>15</td>
<td>28</td>
<td>2</td>
</tr>
<tr>
<td>HLH</td>
<td>4</td>
<td>8</td>
<td>-</td>
</tr>
<tr>
<td>&gt;HL</td>
<td>5</td>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>H</td>
<td>12</td>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>LHL</td>
<td>-</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>100</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 5: Number and percentage of emphatic, contrastive and ambiguous referents

I observed in the percentage data that rising contours (LH, >LH, HLH) are strongly associated with emphasis, since LH and HLH were only used on emphatic referents. The data also show a more frequent percentage of falling contours on contrastive referents – HL (65%), >HL (16%), LHL (10%).

5. Final Remarks

In this paper, I analysed the interaction between degrees of givenness and intonation in Brazilian Portuguese. Despite the impossibility of a detailed statistical analysis, the results have revealed important relations, on the one hand between falling contours (HL, >HL, LHL) and contrast and, on the other hand, between rising contours (LH, >LH, HLH) and emphasis. Regarding the degree of givenness, the results have not indicated a strong relationship between degree of givenness and intonation. However, the results confirmed studies, which have related high tone to new information. Here, new information is also indicated with high tone. In sum, it is possible to state that LH contour is more frequently used to express given information, while HL is more frequently used to convey newsworthy information.

In closing, I would like to highlight that this study was realized using a spontaneous speech corpus, which is an important feature of a first analysis. Nonetheless, further research, including that using different methods, is crucial to confirm or refute the findings of this analysis.

6. References


PHONETIC STUDIES
The dynamics of perception and production of VOT patterns in English by Brazilian learners

Ubiratã Kickhöfel ALVES, Márcia Cristina ZIMMER
Universidade Federal do Rio Grande do Sul; Universidade Católica de Pelotas
Brazil
ukalves@gmail.com, marcia.zimmer@gmail.com

Abstract
In this paper, we present the results of an experimental study on the perception and production of English Voice Onset Time (VOT) patterns by Brazilian learners. Twenty-four participants from Southern Brazil took part in the study. All learners sat for both a discrimination test and an oral production task. The discrimination test was composed of an AxB task, in which we contrasted the three VOT patterns produced by native speakers of English: pre-voicing, short VOT and long VOT. For this test, productions of voiceless plosives were also manipulated on Praat, so that we could obtain artificial short VOT plosives. In the production test, learners were asked to read word-initial /b/, /d/, /g/ aloud. The preliminary results obtained from this experimental study suggest that the acquisition of voicing distinctions, both in terms of perception and production, may be characterized by a multitude of acoustic cues employed by learners, who, in their L2 developmental process, have to learn how to tune in to those cues which are most relevant in the language system to be acquired.

Keywords: VOT; L2 perception; L2 production.

1. Introduction
Learning L2 phonology can be characterized as a non-linear and dynamic process. Variables that are part of this complex system are fully interconnected, systems tend to stabilize for some time in attractor states and language development over time can grow or decline in a nonlinear fashion (Port & Van Gelder, 1995) Therefore, a multitude of variables, which operate at different levels, play a crucial role in second language learning (De Bot et al., 2007).

Departing from this dynamic conception of language acquisition, we present the results of an experimental study on the perception and production of Voice Onset Time (VOT) patterns by Southern Brazilian learners of English. The production of English word-initial stops tends to be difficult for Brazilian learners of English. In Brazilian Portuguese, voiced plosives are produced with pre-voicing (i.e., negative VOT), and voiceless plosives are produced with a short VOT pattern (also known as “Zero VOT”). This is different from what can be found in English, in which voiced stops are produced with either some pre-voicing or with Zero VOT, whereas voiceless initial plosives are produced with a long VOT pattern (aspirated). Given the fact that the Zero VOT pattern (short) is used in voiceless stops in BP but in voiced stops in English, Brazilian learners tend to show some problems in discriminating, identifying and producing the distinction between word-initial voiceless and voice plosive consonants in English.

Our main goals in this article are: (i) to assess whether learners in three different proficient levels are able to distinguish among the production of different VOT patterns of English stop consonants; (ii) to investigate if these students produce VOT values which become gradually similar, according to their proficiency level, to those patterns found in American English; (iii) to study the relation between perception and production in L2 learning. The preliminary results shown in this paper are discussed mainly regarding the dynamics and nonlinearity between the processes of discrimination and production of the L2 VOT patterns.

2. Method
Twenty-four Southern Brazilian learners of English took part in the study. All of them were taking their undergraduate majors in English in one of the institutions of the two authors. After having taken the Oxford Placement Test (Allan, 2004), learners were organized as belonging to three different proficiency groups: proficient (6 participants), intermediate (7 participants) and basic (11 participants). All learners took part in both a perception (discrimination) and a production test.

The discrimination test consisted of an AxB task. In this task, the stimulus presented to learners consisted of triads. In a test booklet, participants were provided with multiple choice questions and were asked to indicate if the initial consonant of the second word was similar either to the 1st word (e.g. beer – beer – peer), or to the 3rd word (e.g. beer – peer – peer), or whether the three words began with the same consonant (e.g. peer – peer – peer). In order to build the stimuli, we invited two speakers of North American English, who had been living in Southern Brazil for less than 6 months, to record their production of the stimuli in a professional studio. These two speakers read a set of pre-selected words, all of which starting with a high vowel (cf. Yavas & Wildermuth, 2006; Yavas, 2008). In building the stimuli, we contrasted the three VOT patterns produced by these native speakers of English: pre-voicing (found in some productions of initial voiced consonants), short VOT (also found in their productions of voiced stops) and long VOT (found in their production of voiceless plosives). For this test, productions of voiceless plosives were also manipulated on Praat (Boersma & Hayes, 2011), so that we could obtain artificial short VOT plosives: as the VOT of the plosives was reduced, the resulting manipulated consonant would have the same VOT duration as a voiced
segment. These artificial voiced stops were contrasted with the three natural VOT patterns in the AxB task. Therefore, four kinds of contrasts were tested in the AxB task: natural zero VOT vs. negative VOT (6 questions), Artificial zero VOT vs. Negative VOT (6), Natural zero VOT vs. artificial zero VOT (6) and positive VOT vs. negative VOT (6). In Figure 1, the overall design of the AxB experiment is presented:

![Figure 1: AxB Task Design](Image)

Our purpose in testing learners on a manipulated VOT pattern is to assess whether VOT was the only acoustic cue used in their distinction between voiced or voiceless plosives. Should VOT be the only acoustic at play, learners would not be able to discriminate between those plosives starting with a natural Zero VOT and those ones which had their VOT reduced.

In the production test, learners were asked to read words starting with the consonants /p/, /t/ and /k/ aloud. These target words, which were repeated twice, were presented isolated, in a powerpoint presentation shown on a laptop computer. In Figure 2, the overall design of the production test is presented:

![Figure 2: Production Task Design](Image)

The design described above allowed us to investigate our three hypotheses for this study. The first three hypotheses concern the results obtained from the AxB task, whereas the fourth hypothesis approaches the results of the production task.

1) There will be no significant differences among the three groups for the contrast between Negative and (natural) Zero VOT.
2) There will be no significant differences among the three proficiency groups in the discrimination of the contrast between Negative and (artificial) Zero VOT.
3) There will be no significant differences among the three groups for the contrast between Natural Zero VOT and Artificial Zero VOT.
4) There will be a significant difference among the three groups in the VOT values for each one of the consonants (/p/, /t/ and /k/), as the three groups are going to produce native-like VOT values.

As for the first hypothesis, we don’t expect Brazilian learners to discriminate between Negative and (natural) Zero VOT, as we hypothesize these learners consider these two patterns to be instances of the same category of voiced stops. With regard to (2), we hypothesized that all learners are able to perceive the difference between negative and manipulated VOT stimuli, so there will be no differences among groups. In (3), we hypothesized all learners are able to perceive the difference between natural and manipulated VOT stimuli, so there should be no differences among groups either. Finally, in (4), significant differences were predicted according to the participants’ proficiency level.

The experimental results and the discussion of these hypotheses are presented in what follows.

### 3. Results

Table 1 shows the results obtained from the AxB task.

<table>
<thead>
<tr>
<th>Contrasted VOT</th>
<th>Accuracy (%)</th>
<th>Similarity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural vs. Art. Zero</td>
<td>7.77</td>
<td>7.57</td>
</tr>
<tr>
<td>Negative vs. Artificial Zero</td>
<td>7.77</td>
<td>7.57</td>
</tr>
<tr>
<td>Natural vs. Zero</td>
<td>7.77</td>
<td>7.57</td>
</tr>
<tr>
<td>Negative vs. Positive</td>
<td>9.44</td>
<td>9.00</td>
</tr>
</tbody>
</table>

**Table 1: AxB Task Results**

Table 1 presents two main labels for its columns: “accuracy”, which indicates that learners were able to efficiently distinguish between the two patterns, and “similarity”, which presents the frequency rates with which learners chose the “all consonants equal” choice.

As we observe the data in Table 1, we notice that learners proved able to distinguish between negative and positive VOT patterns. These results are in accordance with a previous study carried out by Alves et al. (2011), which showed that these same participants reached ceiling effects in a task in which they were asked to identify the voicing of word-initial plosives. In other words, participants are already able to distinguish between voiceless and voiced stops in English. The results of this private study, corroborated by the discrimination findings between Negative and Positive VOT found in Table 3, motivated the present investigation on the discrimination of natural and manipulated stimuli, which gave rise to the three hypotheses guiding this study.

As for the first hypothesis, we concluded that, regardless of the participants’ proficiency level, they do not discriminate between Negative and (natural) Zero VOT patterns. Kruskal Wallis tests were run in order to check whether there was a significant difference among the three proficiency groups, but no significant differences were found (Accuracy: n.s. (X2(2) = 1.196, p = .550); (Similarity : n.s. (X2 (2) = 1.228, p = .541).
Hypothesis 1 was thus corroborated. In other words, learners in all proficiency groups tend to accurately judge both the negative and zero VOT patterns as corresponding to instances of voiced stops in English.

The second hypothesis investigated the discrimination between the Negative and the (artificial) Zero VOT patterns. Table 1 shows that learners in the three proficiency groups tended to discriminate these two patterns, as indicating them as referring to two consonants of different voicing categories. However, the results obtained from Kruskal-Wallis tests indicated a significant difference among the three groups with regard to their Accuracy rates (Accuracy: s. (X2(2) = 7,916, p = .019); Mann-Whitney (Prof and Basic) Similarity: n.s. (X2 (2) 2,353, p = .308). This is mainly explained by the lower rates found in the answers provided by learners in the Basic Proficiency group, which seemed to be more doubtful about discriminating these two patterns. Given these findings, hypothesis 2 was not corroborated.

The last hypothesis on the perception task investigated the discrimination between Natural and Artificial Zero VOT. This comparison is of great importance to the present study, as it may be indicative of whether VOT is the single acoustic cue Brazilian learners of English make use of when distinguishing between voiceless and voiced consonants. The results in Table 1 show high rates of discrimination between these two VOT patterns, regardless of the learners’ proficiency level. This was also confirmed by the results obtained from the Kruskal Wallis test, which showed there were no significant differences among the three groups (n.s. (X2(2) 1,968, p = .374); Similarity: n.s. (X2(2) = .392 p = .822), as the three of them tended to discriminate artificial and natural zero VOT patterns. Our third hypothesis was, therefore, corroborated.

Still regarding the results of the AxB task, as we pursued the perception of the artificial Zero VOT pattern further, we ran post-hoc pairwise comparison in which we contrasted the accuracy levels of all students in two different contrasts: Natural Zero vs. Negative and Artificial Zero vs. Negative. The results obtained from this Paired T-Test indicated a significant difference (t(23)=9,364, p = 0,000) between the rates given for each of these constraints. This result may be understood as we consider the fact that learners do not to discriminate between the Natural Zero and Negative VOT patterns, but do differentiate Artificial Zero and Negative VOT. Once again, this is indicative that learners tend to treat the Natural and Artificial Zero VOT patterns differently.

As for the production tests, the results, organized according to place of constriction, are shown in Table 2.

<table>
<thead>
<tr>
<th>Consonant</th>
<th>Proficient (6)</th>
<th>Intermediate (5)</th>
<th>Basic (11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/p/</td>
<td>43.94 (31.17)</td>
<td>21.75 (13.03)</td>
<td>26.28 (15.06)</td>
</tr>
<tr>
<td>/t/</td>
<td>61.97 (22.16)</td>
<td>61.18 (21.99)</td>
<td>57.49 (24.62)</td>
</tr>
<tr>
<td>/k/</td>
<td>87.00 (31.11)</td>
<td>75.79 (24.68)</td>
<td>86.05 (19.35)</td>
</tr>
</tbody>
</table>

Table 2: Production Task results (mean VOT in ms)

The results shown in Table 2 suggest that, regardless of the learners’ proficiency group, nativelike VOT values for /p/ and /t/ were not yet produced. With regard to the velar consonant, the three proficiency groups tended to present native-like values in their VOT production. This will be discussed further in the following section.

As VOT values tend to decrease the more fronted the place of constriction of the consonant is, we investigated our fourth hypothesis in each one of the consonants (/p/, /t/ and /k/) taken separately. As for the labial consonant, a One-Way Anova indicated a significant difference among groups (F (2) = 3,493, p = 0,049). This can be explained as the mean VOT value produced by the Proficient Group is much higher than those presented by the Intermediate and Basic Learners. Even though not even the proficient participants were able to produce near-native VOT values for /p/ (around 60ms, cf. Ladefoged and Cho, 2004), Hypothesis 4 was confirmed for this consonant.

The fourth hypothesis, however, was not confirmed for /t/ or /k/, due to two different reasons. Another One-Way Anova showed no significant differences – (F (2) = 0,102, p = 0,903) among groups in their mean VOT values for /t/, as the three groups of learners presented similar VOT values, which did not resemble the nativelike ones (about 75ms, cf. Cho & Ladefoged, 1999). Although significant differences could not be found in the mean VOT values found for /k/ (F (2) = 0.904 , p = 0,420) either, it is important to point out that, unlike what was shown in the values for the other two consonants, the main VOT values for this stop seem to be produced in a nativelike fashion by learners in the three proficiency groups. In other words, even though our fourth hypothesis was confirmed only for /p/, the mean VOT values found in each one of the three consonants tend to show a different behavior. This will be explained further in the section that follows.

4. Discussion

As already mentioned, the present investigation was motivated by a previous study developed by Alves et al. (2011), in which the same participants of this study presented high accuracy levels in an identification test. By considering the fact that these learners were already able to identify voicing patterns in English, but still seemed to show several problems concerning the production of aspirated (positive VOT) consonants, we inquired whether other factors, besides VOT, might have an influence in the discrimination and production of voiced and voiceless plosives in Brazilian Portuguese-English interlanguage.

Our discrimination task results confirmed our hypothesis that learners could not discriminate between productions of Negative and natural Zero VOT, since both patterns would be considered to be indicators of voiced stops, as can be found in the production of native speakers of English. Our hypothesis that participants in all proficiency levels discriminate Natural from Artificial VOT patterns was also confirmed, as simply reducing the VOT length of an aspirated stop was not enough to
Speech sounds are categorized by a multitude of acoustic cues that do not act in isolation. This considered, learning to perceive (and consequently produce) the sounds of a second language implies having learners tune in to those cues which play a more decisive role in this new sound system. This might imply giving importance to some cues whose role was not imperative in their first language system.

This may seem to be the case of the participants in this study. Although VOT patterns are regarded as the most important acoustic cue among native speakers of English (cf. Lisker & Abramson, 1964), this does not seem to be the single or most important aspect considered by our learners. Further studies need to investigate which other aspects might be playing a role, among which burst intensity, might have an effect on the perception of these voicing patterns.

The possible role of burst intensity should also be highlighted as we consider the production data. Significant differences were found only for the production of /p/, even though none of the three proficiency levels were able to achieve the target VOT values for this consonant. As for the velar consonant, no significant differences were found, as the three groups seemed to have achieved the target VOT values. Finally, no significant differences were found among groups for /t/, even though learners seem to be closer to achieve the target VOT values for this consonant than they are with regard to /p/. These results seem to be very interesting as the role of burst intensity is taken into consideration. If we consider that the cue of burst intensity is stronger for /p/, a possibility might be that, in other to distinguish between /p/ and /b/, learners might be making use of this cue more regularly than they attend to VOT values. In other words, it might be the case that acoustic cues vary not only in terms of the learners’ proficiency level, but also in terms of the place of constriction of the target consonant, as the acoustic correlates of VOT length and burst intensity may vary between /p/, /t/ and /k/. Additional statistical tests, which take each one of the consonants separately in the AxB task, may be indicative of a possible connection between perception, production and the use of different acoustic cues according to the place of constriction of the target consonant.

The possibilities discussed above deserve further investigation, as future studies should provide more detailed knowledge into what other acoustic cues are used not only by Brazilian learners of English, but also by learners of English from different first language systems. As to our future directions, it seems to us that data on the production of Brazilian Portuguese /pl/, /l/ and /k/ must also be measured, so that we can investigate whether different acoustic cues are also at play in the production of these learners’ L1 stop consonants. Furthermore, a control group with American participants also proves necessary, as it is imperative to confirm whether VOT is really the main cue which allows native speakers of English to distinguish between voiceless and voiced stops.

We believe that the results to be obtained from these future studies on the role of different acoustic cues, according to the learners’ L1 system, involved both in perception and production, can provide further insight into the view of language as a complex adaptive system (Beckner et al., 2009), according to which learning a second language is heavily attached to a complexity of variables in interaction (Herdia & Jessner, 2002, DeBot et al., 2007). This seems to be the case of learning an L2/L3 sound system, which cannot be confined to an “all or nothing” issue, since a variety of acoustic correlates might be playing a role in this complex process.

5. References


Acoustic description of the English and Brazilian Portuguese front vowel systems of Brazilian EFL teachers

Clerton Luiz Félix BARBOZA, Wilson Júnior de Araújo CARVALHO, Katiene Rozy Santos do NASCIMENTO
URN-UF, UECE, UECE
clerontluiz@gmail.com, wiljrac@gmail.com, katiene@hotmail.com

Abstract

The present research has as its main objective the description and comparison of the front-vowel systems of both Brazilian Portuguese (BP) and English as a foreign language (EFL) as realized by English teachers in western Rio Grande do Norte-Brazil. We focus on an usage-based analysis of the phonetic details, such as duration, Euclidian distance, F1 and F2. Our methodology made use of a set of four experiments used to elicit BP and EFL vowels in a CVC or CVCCV context. Two experiments were used to collect data from each language. The first made use of reading carrier sentences and the second used a street map as the main cue for eliciting data. Results regarding spectral data show overlap was found as regards the high-front vowel system of both languages. On the other hand, low-front vowels did not show acoustic overlap. As for duration, it seems to be used as the main acoustic cue to distinguish the exemplars of both languages, as EFL vowels are significantly longer than BP ones.

Keywords: front-vowels; EFL; BP.

1. Introduction

Traditional phonological theories assume the mental representation of the phonological level is simple, free of the details and redundancies found on the phonetic level. Much effort is made to try to find a set of rules, processes or restrictions capable of explaining satisfactorily the mapping from this simple mental representation to the phonetic level, which is complex. On the other hand, phonological theories based on use defend a mental representation capable of retrieving the phonetic details considered redundant by traditional theories. Once it assumes the mental representation to be complex, mapping from this representation becomes simple, as it is not necessary to use a set of rules, processes or restrictions which aim at simplifying or normalizing the phonetic realization. The view above is in consonance with the one defended by Bybee (2001) and Johnson and Mullenix (1997), respectively seminal texts as regards Phonology of Use and the Exemplar Model.

Having in mind the usage-based approaches commented above, the present research has as its main aim the description and comparison of the front-vowel systems of both Brazilian Portuguese (BP) and English as a foreign language (EFL) as realized by English teachers in western Rio Grande do Norte state, in north-eastern Brazil. Our specific focus lies on the analysis of spectral and duration cues of front-vowels as produced by EFL teachers in the aforementioned region. Our main hypothesis states EFL exemplars are markedly influenced by BP as regards their spectral and duration phonetic details.

On the following pages we present a brief overview of previous research on BP EFL vowel production and perception, our research methodology, our main results, as well as our conclusions.

2. Literature overview

Studies involving English vowels production and perception have been carried out for quite a long time. However, only quite recently usage-based approaches started to be used in this field of research. A glimpse of an exemplar approach to phonology is observed in the most seminal research on the field, presented by Peterson and Barney (1952). In this study, it is made clear the enormous amount of variation vowels are subject to. Such variation, however, is not enough to impede a correct perception by listeners most of the time. Figures presented by the authors are similar to the exemplar clouds reported on the present study, as a series of vowel realizations are more or less associated with a central, most robust exemplar.

Similar usage-based inferences are allowed in the research of a multitude of scholars and their work on the production of English as a second or foreign language (Baker & Trofimovich, 2005; Flege, Schirru, & MacKay, 2003; Cebrian, 2006), on perception only (Højen & Flege, 2006; Flege & MacKay, 2005), or on both skills (Jia et al., 2006).

Having in mind this study focus on BP EFL subjects, we discuss below only the results presented by Baptista (2000), Rauber et al. (2005), Bion et al. (2006), Rauber (2006), and Nobre-Oliveira (2007). Such studies focus on BP speakers of EFL and are therefore worth reviewing.

Baptista (2000) is a longitudinal production study which describes the acquisition of English vowels of BP speakers living in the US. Results indicate a holistic approach to vowel acquisition. For example, in acquiring [ɪ] some subjects lowered the production of the first sound of the diphthong [eɪ]. Other changes in the system are also mentioned, as the need to make the front-vowel space longer, once English has more vowels than BP.

Rauber et al. (2005) investigated the relationship between English vowel perception and production by advanced EFL learners in Brazil. Perception data indicate a good accuracy for distinguishing the [i, ɪ] pair, but a poor perception of the [ɛ, æ] pair. The same results were obtained in production, with the former pair being well produced, and the latter, poorly realized.

Bion et al. (2006) also involved production and perception of natural stimuli, but added synthesized vowels with fixed duration but variable spectral quality. Once again natural data indicated the pair [i, ɪ] was better
perceived and produced than the [ɛ, æ] pair. Synthesized stimuli revealed even with controlled duration, the former pair is easier to perceive than the latter. This results indicates duration is not a primary cue for distinguishing [ɪ, ɨ], whilst it is important for improving [ɛ, æ] perception.

Rauber (2006) again used synthesized vowels to study perception. Results indicated BP subjects use duration as their primary cue to distinguish both [ɪ, ɨ] and [ɛ, æ] vowel pairs. Once again, the former pair was easier to perceive than the former. As regards production, the same problem arises, with [ɛ, æ] showing grater vowel overlap than [ɪ, ɨ]. Duration was found to be important as well, once duration of the constituents in both pairs were significantly different.

Finally, Nobre-Oliveira (2007) carried out a study whose main focus was on perceptual training using both natural and synthesized stimuli. Production research was also carried out. The group which used synthesized stimuli had better production and perception results than the one which used natural stimuli.

Even though all the aforementioned studies are not grounded on a usage-based view of phonology, their results fit perfectly on the frame. Phonetic details associated with the realization of the BP front vowel system are seen influencing EFL throughout all the studies. BP influence seems to be concentrated on the spectral level, once most results indicate some degree of overlap especially of the [ɛ, æ] pair. Results involving duration, on the other hand, show a greater independence between the BP and EFL vowel systems, indicating training on this acoustic cue is important for our learners. It can be stated, thus, that BP vowel exemplars characteristics are more strongly linked to their EFL correlates, probably through a network, specially the low-front pair [ɛ, æ].

Next section deals with the research methodology used on this study.

3. Methodology

Our subjects were a group of 20 male English teachers. All but one had university level. None had ever been abroad. Four experiments were carried out. Two experiments involved the reading of CVC (EFL) and CVCV (BP) words in carrier sentences. Two involved role-playing location information over a small city map in both EFL and BP.

Exemplars of the BP front-vowels [ɪ, ɨ, ɛ, æ] were collected using the carrier sentence “X. Diga Y alto”. X and Y were words containing the same vowel, but only Y was acoustically analysed. Each sentence was repeated 3 times. 720 BP vowel exemplars were thus collected on this experiment, from now on called L1-1.

The second BP experiment involved the use of a small city map in which street names were used as cues to elicit the same vowel exemplars. Subjects were asked about how to go from one place to another. Each word was recorded 5 times. We analysed, thus, 400 vowel exemplars in this experiment, which we called L1-2.

First EFL experiment was similar to L1-1. [ɪ, ɨ, ɛ, æ] vowel exemplars were collected in the carrier-sentence “X. Say Y again.” Once again, X and Y were words with the same vowel exemplar, but only the one in Y position was acoustically analyzed. 1500 vowel exemplars were collected on this experiment, called L2-1.

The second EFL experiment also used a small city map. Procedures were identical to experiment L1-2, but given the bigger number of vowel exemplars analyzed in EFL data collection, the total number of tokens reached 500. This experiment was called L2-2.

Overall number of tokens reached the total of 3,120, with 8,680 values of FI, F2 and duration being analyzed. Statistical test used were mainly paired-samples t-test and repeated measures ANOVAS. SPSS was used to carry out all the statistical treatment of the data.

Acoustic analyses were carried out using Praat, version 4.6.21. Formant analyses were carried out in a point in the middle of the vowel, except for the diphthong [ɛɪ] in both languages, which had only the middle of the first vowel analyzed. Duration measures excluded VOT when applicable and included only the pressure peaks of the exemplar vowels with visible formants on the spectrogram. No duration was measured for the diphthongs.

Recordings were made in a quiet, but not acoustically treated, room. We used a Shure SM-58 unidirectional dynamic microphone and a digital Microtrack 24/96 recording WAVE 16-bit, 44Khz files.

Next section presents our data analyses results & discussion.

4. Results & Discussion

For the sake of brevity, we chose not to present a huge amount of tables with exact spectral and duration measurements. Instead, we will focus on presenting informative figures as much as we can. In case exact data is needed, we invite our readers to send us an email.

Paired-sample t-tests revealed significant (p < .001) differences between the [ɪ, ɨ] in all tests. We can observe in Figure 1 there is no overlap between the high-front EFL exemplars in experiment L2-1. Experiment L2-2 showed a very similar picture and is not reproduced.

![Figure 1: EFL exemplars [ɪ, ɨ] in experiment L2-1](image)

Figure 1: EFL exemplars [ɪ, ɨ] in experiment L2-1

Further evidence of the motor control our subjects have in dealing with the aforementioned pair is found
when Euclidian distance is analyzed, once paired-sample t-tests indicate non-significant differences (p = .693) between L2-1 and L2-2 values.

Comparisons of the English high-front exemplars [i] and [ɪ] with Brazilian Portuguese [i] and [e] in a spectral level were also carried out. Repeated-measures ANOVAs involving the L2-1 exemplars [i] and [ɪ] and L1-1 [i] found non-significant differences in both F1 and F2 between L2-1 [i] and L1-1 [i] (both p > .05). However, the same test found significant differences between L2-1 [ɪ] and L1-1 [i] (both p < .05). Figure 2 indicates graphically this high degree of exemplar overlap between L2-1 [i] and L1-1 [i].

![Figure 2: L2-1 exemplars (red) and L1-1 (blue)](image)

Results of repeated-measures ANOVAs involving L2-2 exemplars [i, ɪ] and L1-2 [i], found similar results, except for a significant difference in F2 for L2-2 exemplar [i] and L1-2 exemplar [i] (p < .05). The resulting figure, however, was very similar to the one presented above and was therefore not presented. We focus on the comparison between EFL [i] and BP [e].

![Figure 3: L2-1 exemplar (red) and L1-1 (blue)](image)

Figure 3 shows a high degree of exemplar overlap between EFL [i] and BP [e], indicating a degree of gesture influence as big as the one found between EFL [i] and BP [i]. Paired-sample t-tests involving L2-2 [i] and L1-2 [e] found significant differences only for F1 (p = .006). A comparison between L2-2 [i] and L2-2 [e] reached non-significant levels for both F1 (p = .06) and F2 (p = .469), indicating an even bigger degree of exemplar overlap.

As regards duration of the EFL and BP high-front vowel exemplars, results indicate all duration values were substantially different across experiments. Paired-sample t-tests found significant differences (p < .001) for all comparisons, with both L2-1 and L1-1 exemplars having a longer duration than the values found in L2-2 and L1-2 experiments. Focusing on the EFL results, Figure 4 indicates the exemplar [i] is significantly longer in both L2-1 (p = .003) and L2-2 (p < .001).

![Figure 4: duration for the exemplars [i, ɪ] in experiments L2-1 and L2-2](image)

Repeated-measure ANOVAs found a significant difference in duration between EFL L2-1 [i, ɪ] and BP [i] (p < .05). A paired-sample t-test also found a significant difference in duration between L2-1 [i] and L1-1 [e] (p = .026). Even though new ANOVAs involving EFL L2-2 exemplars [i, ɪ] and BP L1-2 [i] showed similar results (p < .05), a t-test for L2-2 exemplar [i] and L1-2 [e] achieved only non-significant levels (p = .824). This non-significant result reinforces the shared exemplar feature of EFL [i] and BP [e] on the duration level as well as the spectral.

We now turn to the spectral characteristics of the first element of the diphthong [eɪ] in both EFL and BP. We remind our reader no duration measurements were made for these two exemplars. A high degree of vowel overlap is again observed in Figure 5, involving L2-1/L1-1 [eɪ]. Paired-sample tests indicate, however, a significant difference for F2 (p < .001), but not for F1 (p = .232). These results are opposed to the ones found in the comparison of L2-2/L1-2 [eɪ] (F2 (p = .258); F1 (p < .001). We present only one figure owing to lack of room and to the high degree of similarity between them.

![Figure 5: L2-1 (red)/ L1-1 (blue) [eɪ]](image)
We now change our focus from high-front to low-front vowel exemplars in both BP and EFL. A superficial analysis of the spectral data reveals a more stable exemplar [ɛ] than [æ], as indicated by the standard deviation ellipses seen on Figure 6. We can also observe a huge amount of exemplar overlap of the two vowels. It indicates a good number of our subjects treat the EFL pair [ɛ, æ] as a single sound. Even though this overlap is easily observed, paired-sample t-tests revealed a significant difference between the exemplar pair in experiment L2-1 (F1 (p. < .001); F2 (p. = .238)).

The view that most of our subjects treat the EFL exemplar pair [ɛ, æ] as a single exemplar is reinforce by the data presented in Figure 7, regarding experiments L2-2. The exemplar overlap is even higher than the one found in Figure 6. However, this time a significant difference was found for F2 (p. = .014) but not for F1 (p. = .425) by the paired-sample t-tests. Final piece of evidence BP speakers treat EFL exemplar pair [ɛ, æ] mostly as a single exemplar was found by the analysis of the Euclidean distance between these vowels across experiments. A paired-sample t-test confirmed the non-significant difference (p. = .443) between experiments L2-1 and L2-2.

A comparison between EFL [ɛ, æ] and BP [ɛ], presented in Figure 8, shows the BP exemplar is significantly higher (p. < .05) than its EFL counterparts [ɛ, æ]. However, significant difference was found for F2 (p. = .475) when comparing L2-1 and L1-1 exemplars. The same results were found for L2-2 and L1-2.

Figure 8: L2-1 (red)/ L1-1 (blue) exemplars.

Such results indicate that in as regards spectral cues, advanced BP speakers of EFL tend to create a new exemplar which is associated with the English pair [ɛ, æ].

Finally, as regards duration measurements Figure 9 presents a boxplot of the data regarding L2-1 [ɛ, æ] and L1-1 [ɛ] vowel exemplars. Both EFL exemplars were realized with a longer duration than the BP one as revealed by a repeated-measures ANOVA which reached significant results (p. < .05). As regard duration differences between the EFL vowel exemplars [ɛ, æ], the same test failed to show a significant difference (p. > 0.05), indicating our subjects do not realize the exemplars differently.

Figure 9: L2-1 (red)/ L1-1 (blue) durations

L2-2 and L1-2 data is very similar to the figure presented above. We therefore do not present the boxplot regarding this set of data. Once again PB vowel exemplar [ɛ] was significant shorter in duration when compared to EFL [ɛ, æ] (p. < .05). And non-significant results were obtained between the EFL exemplar pair [ɛ, æ] (p. > .05).

5. Conclusion

A long tradition of interlanguage studies emphasizes the transfer of mother-tongue phonetic-phonological characteristics in the acquisition of a given foreign/second language. This tradition has created a perception that once
a positive or negative interference across languages has been noticed, all learners will face the same problems. This idea, however, is not completely true, once a huge amount of variables can influence positively or negatively a language acquisition construction course. The present research has concluded that BP dialectal variation is also responsible for EFL variation, once some of our results do not match previous research, made with subjects who speak other BP southern dialects, especially as regards duration as an important cue for producing the low-front vowel exemplars [ɛ, æ]. Other researches involving BP dialectal variation and its influence on EFL production are therefore necessary to achieve a more detailed view of EFL acquisition by BP learners.

As regards our own data, we are able to state our informants rely heavily on their BP vowel exemplars in order to produce EFL vowel categories. This could be observed mostly in the realization of the high-mid EFL vowels [i, ɛ, ɪ] which overlapped significantly with BP [i, ɛ, ɪ] in this study. This overlap was not found in the same degree with the EFL low-front exemplars [ɛ, æ], as these exemplars were realized significantly lower than BP [ɛ] exemplar. Duration results, by their turn, indicated EFL exemplars to be different than BP ones. Such acoustic cue seems to be important for the production of BP speakers of EFL, even though most fail to realize significant differences between the low-front [ɛ, æ] EFL pair.

Pedagogical implications for the teaching of EFL to BP speakers involve the early association of BP and EFL [i] exemplars, as well as BP [ɛ] and EFL [ɛ] vowel exemplars. Such early association would avoid the production of English [i, ɪ] as similar to BP vowel exemplar [i]. This was precisely what Baptista (2000) observed in her research. Another important implication is related to the EFL low-front exemplars [ɛ, æ]. Despite the fact they constituted a new vowel exemplar separate from BP [ɛ] exemplar, this creation of a single vowel category for two EFL categories indicates the high degree of training BP EFL speakers need in order to make their front-low vowel space longer so that to accommodate the English vowel space.

Finally, as a limitation of the present study we emphasize the lack of treatment frequency effects had in our research. Aware of the importance of frequency to the exemplar model as well as to usage-based phonology, a next logical step or our research is to include this variable in our future studies, alongside with the phonetic detail analyzed in the present research.

6. Acknowledgements

Clerto Barboza would like to thank CAPES and UERN by the financial and academic support to carry out the present research.

Wilson Carvalho would like to thank UECE for the academic and financial support.

Katiene Nascimento would like to thank CAPES for the financial support.

7. References


Flutuação do acento em palavras produzidas por falantes nativos do Português Brasileiro

Aline de Lima BENEVIDES
Universidade de São Paulo (USP)
aline.benevides@usp.br

Resumo


Palavras-Chaves: Fonologia; Acento; Português Brasileiro.

1. Objetivo

A finalidade desta pesquisa é verificar o condicionamento e/ou o estímulo da flutuação do acento em palavras produzidas por alguns falantes nativos do Português Brasileiro – doravante PB.

2. Composição do corpus

O corpus é constituído por um conjunto de palavras cuja pronúncia apresenta variação sendo considerada pelos gramáticos como fora da dita “norma culta”, por exemplo: [gra'tuji] > [gratu'i]u] e [no'bew] > [n'bew]. O processo estudado é tratado como silabada, que, segundo os gramáticos, é “o erro de prosódia que consiste na deslocação do acento tônico de uma palavra” (Bechara, 2005: 90).

As etapas de formação do corpus foram: i) a partir das gramáticas de Cunha & Cintra (2001), Lima (2002) e Bechara (1976; 2005), listamos as palavras consideradas “mais usuais”, totalizando 79 vocábulos; ii) apresentamos essa lista de verbetes para 12 colegas, os quais deveriam apontar em quais palavras já escutaram a pronúncia oscilante, resultando em 54 marcados e 2 palavras sugeridas para acréscimo; e, iii) realizamos um teste de produção com os 25 vocábulos não marcados, além deles, 11 verbetes que apresentaram poucas marcações também foram inseridos no teste, finalizando com 36 palavras.

A metodologia utilizada foi formar frases e solicitar para 14 falantes nativos as lerem. Ressalta-se que não houve rigor com os critérios sociolingüísticos, de modo que a faixa etária, o sexo, o grau de escolaridade e a naturalidade são diferentes, contudo, não há uma distribuição equivalente. Vale lembrar também que a marcação da tonicidade por meio do acento gráfico pode direcionar a leitura para uma ou outra forma, da mesma forma que a ausência dessa marcação também a direciona, uma vez que a neutralidade não é possível, optamos por seguir a ortografia oficial da língua portuguesa.

Das 36 palavras testadas, 25 foram produzidas com oscilação e, surpreendentemente, 2 palavras que não estavam sendo testadas também sofreram flutuação por 2 falantes. Assim, conclui-se a formação do corpus em 72 palavras.

3. O acento regular em Português Brasileiro

Entre as diversas teorias que tentam explicar a realização do acento regular em PB, que se contradizem em partes, fundamentalmente todos os teóricos admitem a realização nas três últimas sílabas a partir da margem direita da palavra. Diferente de outras línguas, como o francês, em que a tonicidade se dá sempre na última sílaba, o acento em português não é totalmente previsível.

Na literatura do PB, encontramos inúmeras teorias que buscam esclarecer o acento regular, diante da impossibilidade de explorar todas elas, resolvemos apresentar as três principais hipóteses para a atribuição do acento regular, segundo Ferreira-Netto (2007), que são:

- Hipótese do Acento Livre - previamente definido no léxico (Camara Jr., 2001);
- Hipótese do Molde Trocaico - definido pela característica rítmica padrão (Bisol, 1992 apud Collischon, 2010);

A primeira proposta prediz que o acento é livre, assim não há uma regra para a atribuição acentual, o que pode ocorrer é uma maior tendência a uma dada terminação. Seguindo esta hipótese, teríamos o acento sendo atribuído no léxico. A lacuna encontrada aqui diz respeito em como ocorre à organização desses vocábulos no léxico, que não é prevista pela teoria.
A segunda hipótese propõe que o peso silábico e o pé métrico são os mecanismos responsáveis pela atribuição acentual. As sílabas finais pesadas atraem o acento, caso não sejam pesadas, o acento cai na penúltima sílaba. Todos os casos que fujam a tais regras são inseridos na extrametricidade.

A terceira proposta se vale de regras diferentes para verbos e não verbos. Nesse, o acento cai na última vogal do radical derivacional. Assim, as paroxítonas com sílabas finais pesadas e as proparoxítonas, que não se enquadram nesse padrão, são consideradas casos marcados lexicalmente.

Como vemos, na segunda e na terceira propostas, centenas de palavras são englobadas na excepcionalidade, que incluem todas as proparoxítonas, algumas paroxítonas e oxítonas. Elas são denominadas pelos autores de casos extramétricos ou marcados, respectivamente. Ora, se um padrão acentual inteiro, o esdrúxulo, mais alguns casos dos outros padrões são considerados desvios às regras acentuais, podemos realmente considerá-los exceções?

O estudo realizado por Araújo et al. (2007) refuta os principais argumentos empregados pelos teóricos que insistem nas paroxítonas na excepcionalidade. O autor demonstra que o padrão esdrúxulo não deve ser considerado excepcional, posto que apresenta a mesma regularidade de entrada na língua que os demais padrões; os processos que reduziram as paroxítonas em paroxítonas, como a síncope e a apócope, não podem afetar todas as palavras por gerar palavras agramaticais, como *[medku] e *[bebdu]; e, por fim, que sua frequência de ocorrência está diretamente relacionada com o número de sílabas, assim as palavras trissílabas paroxítonas possuem uma frequência semelhante a dos demais padrões.

Tendo isso em vista, observamos que tais teorias parecem não esclarecer a regularidade do acento primário no PB, uma vez que nesse corpus, considerando apenas a pronúncia regular, aproximadamente 50% dos verbetes seriam compreendidos na extrametricidade ou na marcação lexical.

### 4. Os dados

Como vimos, o peso silábico é um dos fatores comumente tratados como influente para a atribuição acentual. As propostas afirmam que as sílabas com coda silábica, i. é., as sílabas pesadas, atraem o acento. Além disso, o acento paroxítono por ser o padrão mais produtivo é considerado o padrão acentual do PB, de modo que esta seria a tonicidade atribuída às novas entradas lexicais. Se essas características são realmente importantes para o acento, presume-se que a tonicidade oscilante ocorre em direção a elas.

Entretanto, tal fato não é verificado com uma percentagem significativa no corpus. Observa-se que dentre 34 vocábulos1 que poderiam ter a oscilação motivada pelo peso silábico, apenas 23,5% deles saem de uma sílaba leve em direção a uma sílaba pesada, contrapondo-se a 35,3% que saí de sílaba pesada em direção a uma sílaba leve. As demais percentagens referem-se a: 20,6% saindo de sílaba pesada para outra sílaba pesada e 20,6% saindo de sílaba leve em direção a uma sílaba leve, tendo como opção uma sílaba pesada.

Embora possamos considerar que oscilar de uma sílaba pesada para outra sílaba pesada não seja uma violação a sensibilidade da língua ao peso silábico. Além disso, que em casos como *interim* a flutuação vai de encontro à preferência por sílabas pesadas finais (Collischon, 2010), é possível encontrar casos, como condor, em que a oscilação se opõe a essa predileção.

Ao considerar o padrão acentual, observamos que as palavras proparoxítonas possuem, aproximadamente, 93% de oscilações a favor do padrão, sendo que os 7% restantes flutuam para as sílabas finais pesadas. Se tivéssemos apenas esses dados, pressuporíamos que as duas características apontadas são essenciais para a língua. Logo, a proposta de Bisol (1992 apud Collischon, 2010) seria a melhor hipótese para descrever o acento regular em PB. Contudo, o corpus também apresenta vocábulos paroxítonos, os quais oscilam para outros padrões.

Dentre 36 verbetes com acento na penúltima sílaba, 41,7% flutuam para a antepenúltima, 13,9% para a última sílaba e 44,4% mantêm a penúltima sílaba tónica. Essas são formadas por ditongos mediais que se tornam hiatos ou encontro vocálicos finais, que quando ditongos tornam-se hiatos e quando hiatos tornam-se ditongos. Esses casos ainda não serão considerados, pois receberão tratamento diferenciado ao longo da pesquisa, a saber, serão submetidos a testes.

Atente-se ao fato de que uma percentagem considerável de paroxítonas oscila em direção ao acento proparoxítono, que além de ser considerado um desvio, é apontado como um caso a ser evitado.

Com relação às palavras oxítonas, não há dados suficientes para qualquer afirmação, uma vez que todas elas são vocábulos dissílabos, desse modo não possuem outra opção para a flutuação.

Essa pequena apresentação dos dados foi apenas uma tentativa de esboçar reflexões que devem ser exploradas no desenvolvimento deste trabalho. Até o momento, não é possível tecer qualquer afirmação valendo-se apenas desses dados. Mas alguns questionamentos se instauram: ora, se há uma preferência pelo acento paroxítono, não esperaríamos que a oscilação partisse dele; se o PB evitasse o acento esdrúxulo, não encontraríamos uma percentagem alta de flutuações em sua direção; se há sensibilidade da língua ao peso silábico, elas não só segurariam o acento, mas também os atrairiam.

Com a finalidade de observar se essas podem ser características pertinentes para a colocação acentual, faremos um teste de produção com alguns falantes nativos do PB. Nele, iremos formar um texto com palavras inventadas, as quais possuirão os principais

---

1 Ressalta-se que não incluímos as sílabas que possuem ditongos finais ou mediais, posto que daremos um tratamento diferenciado a eles no desenvolvimento da pesquisa.
padrões silábicos permitidos na língua. Com isso, se o PB for sensível ao peso silábico ou tender ao acento paroxítono, haverá um grande número de atribuição de tonicidade a esse padrão; ou, se esses aspectos não forem relevantes para a língua, não verificaremos uma percentagem significativa de tonicidade nesse padrão acental.

Diferentemente do teste de composição do corpus, esse irá valor-se de algumas variáveis sociolinguísticas, a saber, faixa etária, sexo e grau de escolaridade. Para a faixa etária, faremos três divisões, que são: 1°) a partir de 20 anos até 34 anos; 2°) de 35 anos a 59 anos; e, 3°) mais de 60 anos. Para o grau de escolaridade, distinguiremos os indivíduos que possuem ensino superior (cursando ou completo) dos demais. Ao cruzar esses critérios, teremos 12 perfis distintos.

Para que possamos realizar um análise mais sólida, contaremos com 3 informantes para cada perfil, totalizando assim em 36 gravações. O resultado deste experimento será apresentado em trabalhos posteriores.

5. A Fonologia de Uso

A Fonologia de Uso está compreendida nos modelos de língua baseados no uso, os quais consideram o uso como o principal fator para a formação da gramática dos falantes. Dentro desta teoria, a frequência é um dos factores essenciais da teoria e como eles parecem explicar as questões levantadas e. os lexemas mais frequentes possuem ensino superior (cursando ou completo) dos demais. Ao cruzar esses critérios, teremos 12 perfis distintos.

Para que possamos realizar um análise mais sólida, contaremos com 3 informantes para cada perfil, totalizando assim em 36 gravações. O resultado deste experimento será apresentado em trabalhos posteriores.

5.1 Uma análise à luz da Fonologia de Uso

A Fonologia de Uso é um modelo de língua que tem como premissa o uso da língua e a frequência de uso. Nela, a gramática é formada fundamentalmente pelo uso da língua, de modo que as representações mentais estão em constantes modificações e reestruturações, ou seja, o uso estrutura e modifica as representações mentais. Tais estruturas são construídas por associações fonológicas ou semânticas, ou ainda, quando ambas compõem a estrutura, temos associações morfológicas.

Os itens lexicais não são apenas usos concretos, mas servem de gatilho para as novas entradas lexicais e para os usos menos frequentes. Os padrões produtivos por serem acessados mais rapidamente tornam-se o gatilho da língua, i.e., os lexemas mais frequentes fornecem seus padrões aos menos frequentes. É a alta frequência que é a responsável pela facilidade no acesso das palavras, pela produtividade e pela extensão dos padrões de língua. Além disso, os itens com alta frequência de ocorrência têm força lexical, por isso, possuem resistência morphológica e são menos suscetíveis a mudanças por analogia. Em contrapartida, são mais propícios a sofrer processos fonológicos, como reduções e aparagamentos.

Bybee ainda demonstra que a frequência influencia a aquisição de determinadas formas. Em um estudo do Antigo Inglês, a partir de Phillips (apud Bybee, 2001), mostra que a aquisição do ditongo <eo> se dá de modo diferenciado, conforme a frequência dos vocábulos. Os mais frequentes são adquiridos corretamente, enquanto que, os menos frequentes sofrem simplificações para /e:/ e /æ/. E, posteriormente, a vogal arredondada anterior perde o arredondamento, tornando-se /e/ e /æ/.

Se tal processo pode ocorrer na aquisição dos fonemas sendo influenciado pela frequência das palavras, acredita-se que processo semelhante ocorra com a tonicidade dos vocábulos, posto que os verbetes do corpus possuem baixa frequência de uso e o processo se daria por difusão lexical. Assim, hipotetiza-se que as palavras de baixa frequência se associariam com palavras de alta frequência por similaridades fonológicas, resultando na extensão da tonicidade de um vocábu no exemplo apresentado, útil, com frequência de

![Figura 1: Conexões fonológicas por [ut∫i下一篇] entre útil e sutil](image)
3.514, seria o atrador de *suitil*, com frequência de 11.125\(^2\), e fornerceria o padrão acentual a ele, resultando na forma com tonicidade oscilante.

Essas teias associativas organizam os vocábulos na representação mental do falante e a frequência reforça essas conexões. Com o tempo, o uso da forma oscilante em detrimento da forma regular conduz a modificações desse vocábulo na estrutura mental. A Linguística Cognitiva denomina essas estruturas frequentes, que são formadas por meio da experiência de uso, de estruturas arraigadas.

Possíveis evidências de que algumas das formas oscilantes estão sendo arraigadas pode ser verificada no teste de composição do corpus. Observamos que alguns verbetes que não foram apontados pelos colegas como sujeitos a flutuação acentual são produzidos pela maioria deles com o acento oscilante, por exemplo: *cuteter* e *duplex*.

Um indício de que esse seja um processo de mudança começa a ser encontrado em alguns dicionários (Aulete & Valente, 2012; Ferreira, 2004), os quais apresentam duas entradas lexicais para alguns vocábulos, sendo eles: a forma regular e a forma oscilante. Em geral, eles fazem referência um ao outro e há, em alguns deles, observações, no caso o Dicionário Aulete (2012), indicando a pronúncia do acento tônico, como se observar a seguir:

**Autópsia**  

**Autopsia**  
[Do gr. autopsia.] – s.f. 1. Autópsia (q.v.). [Cf. autopsia, do v. autopsiar.]  

**Observação:**  
[Nota: A 1ª ed. deste Dicionário marcou a pronúncia com o acento tônico no i, de acordo com o étimo. Porém, o uso português consagrou a forma esdrúxula autópsia, pelo que se adota esta acentuação.] (grifo nosso).

Com a finalidade de verificar em que medida as gramáticas também estão sendo afetadas, comparei duas edições da gramática de Bechara, uma editada em 1976 e a outra editada em 2005, que é uma edição revisada. Vale lembrar que os gramáticos costumam apontar dentro do processo de silabada, algumas palavras que admitem dupla prosódia, a forma regular e a oscilante são aceiteáveis pela norma culta. Ora, se as gramáticas não são afetadas pelo uso, não encontraremos mudanças entre as duas edições. Mas se houver modificações, pressupõe-se que o uso também pode afetá-las. Para tanto, examinamos o que é admitido como dupla prosódia em ambas as edições, como resultado tem-se o acréscimo de 6 palavras que passam a ser aceitas como verbetes com dupla prosódia (ver Figura 2). Dessa maneira, pressupõe-se que o uso já está afetando as gramáticas.

Tendo isso em vista, postula-se que, em um primeiro momento, o uso modifica as representações mentais, em um segundo, começa a afetar os dicionários, que já registram algumas das formas oscilantes, e, em um terceiro momento, afeta as gramáticas, as quais passariam a aceitar as duas formas como prosódias possíveis. Claro que, as modificações nas gramáticas são mais lentas devido à resistência normativa.

---

2 Segundo o índice de frequência do Projeto Aspa ( Avaliação Sonora do Português Atual).

**Figura 2: Palavras com dupla prosódia**

**6. Conclusões parciais**

Como vimos, encontramos na literatura do PB diversas teorias que objetivam explicar o acento regular, que vão desde teorias métricas até teorias que consideram aspectos morfológicos. De modo geral, elas se utilizam de um grande número de exceções, que na maioria dos casos incluem um padrão acentual íntero, o esdrúxulo. Em contrapartida, encontramos estudos que defendem o acento proparoxítono através de dados quantitativos.

O objetivo desta pesquisa é buscar um modo de análise que inclua todos os padrões acentuais, posto que admitimos que o acento antepenúltimo não é um caso excepcional. Ao longo das revisões, observou-se que a frequência é o principal mecanismo que inclui ou exclui determinados padrões acentuais nas análises. Ora, se ela é realmente um fator relevante por que não iniciarmos por ela? Foi essa pergunta que nos fizemos e é por ela que adentramos nossos estudos na Fonologia de Uso.

Até o momento, as pesquisas indicam que o uso da língua e a frequência de ocorrência são os principais motivadores das oscilações. Acredita-se que as representações mentais são modificadas a cada interação e que a cada nova ocorrência, da forma regular ou da oscilante, as estruturas se fortificam, conduzindo ao fortalecimento de uma das formas. No entanto, quando as duas formas são produzidas pelo mesmo indivíduo, o que parece ocorrer, em alguns casos, é uma especificação de uso.

Como viemos salientando, muito ainda se tem a dizer e explicar a respeito da flutuação acentual. Este é apenas um estudo piloto que apresenta questões que devem ser exploradas em novos trabalhos. Os próximos passos desta pesquisa serão baseados em testes de produção, os quais têm como finalidade investigar a sensibilidade da língua ao peso silábico, se há tendência
do acento cair no padrão acentual da língua (o paroxítono), o que ocorre nos encontros vocálicos [ia] e [uai] que os fazem oscilar em determinadas posições, se as oscilações têm algum correlato com as variantes sociolinguísticas, tais como grau de escolaridade e faixa etária? Em alguns verbetes como reorde, a tonicidade da língua inglesa parece influenciar na pronúncia desse vocábulo, assim, questiona-se: em que medida a origem etimológica influencia as flutuações? Em palavras como Nobel, o uso das formas regular e oscilante parece ser motivado pelo contexto. A hipótese é que quando o falante se refere à livraria produzirá [ənˈbɛɾ] e quando faz referência ao prémio utilizará [ˈnəbɛɾ]. Tendo isso em vista, será que há especificação de uso entre as duas formas nesse e em outros vocábulos?

Muitas respostas ainda devem ser dadas para se concluir as motivações da flutuação acentual no PB. Para tanto, buscaremos no decorrer desta pesquisa, além de realizar testes, novos apoios teóricos na tentativa de comprovar a motivação do deslocamento acentual. A princípio, as teorias que estão sendo estudadas são: a Fonologia de Uso, a Teoria dos Exemplares e a Linguística Cognitiva, contudo, se no decorrer da pesquisa novos aparelhos teóricos surgirem, eles também serão incorporados ao estudo.

7. Agradecimentos
Ao Prof° Dr. Paulo Chagas de Souza, meu orientador, pelos valiosos comentários e sugestões, a todos os participantes do GSCP pelas enriquecedoras contribuições e às amigas Fernanda Canever e Lívia Oushirom pela amizade, incentivo e sugestões. O presente trabalho foi realizado com o apoio da Pró-Reitoria de Pesquisa da Universidade de São Paulo.

8. Referências


Revisiting the acoustic and auditory approach to speech analysis

Luiz Carlos CAGLIARI, Mariane CARVALHO, Mariana MORETTO Gementi
Capes/CNPq (Proc. N. 303440/2009-0); Capes; Fapesp; UNESP Universidade Estadual Paulista “Júlio de Mesquita Filho”
lccagliari@gmail.com, mazicarvalho@bol.com.br, mariana_moretto@hotmail.com

Abstract
The difficulty to combine articulatory interpretation of speech with acoustic analysis has produced an epistemological conflict recently. To overcome this uneasy situation it is necessary to improve the phoneticians' ear-training and performance skills. A good acoustic analysis can be interpreted in audible based frameworks in the same way as an auditory analysis can be analyzed in acoustic based frameworks. Today it is important to ask how phoneticians carry on their scientific work at phonetics labs and in fieldwork. A short review of the history of phonetics shows that the conflict between acoustic and auditory approaches to phonetics in recent years is new and it has particularities not found in old times. The target is not to make a criticism on what is being produced recently in phonetics. However, the engineering tendency with all modern and technical facilities might not lose the main target of phonetics that is to produce a linguistic relevant analysis of speech.

Keywords: auditory analysis; acoustic analysis; ear-training; phonetic skills.

1. Revisiting the question
In this congress the linguistic corpora is the focus. This is an important and today topic aiming the development of the linguistic science, in particular for the description of languages, in search of universal phonetic and phonological principles and for the definition of particular languages parameters. However to achieve such goals with real scientific results there are required good theories with appropriated approach to the investigating object.

2. A historical conflict
It is important to ask how phoneticians carry on their scientific work at phonetic labs and in fieldwork. This question is not irrelevant, since it is expected that all scientist masters the approaches to the science they do. When technology is used in human science, the question often rises conflicts. Ladefoged (1973) and Yi Xu (2010), in different span of time put the question to phoneticians. Particularly, they showed the conflict between auditory and acoustic approaches to describe phonetic entities.

This conflict has created an epistemological situation in which the areas of phonetics and phonology took different ways, not rarely, producing contradictory results. In this way, the acoustic approach to speech restrict itself to the physics of speech, declaring the real science product. On the other hand, the auditory work describing the sounds of speech were treated as unscientific, idiosyncratic, highly individualized and without scientific value. Only with the support of an acoustic evaluation, the speech could be analyzed and describe properly. With all the recent facilities to carry on acoustic analyses of speech (PRAAT, WinPitch, SFS, ProTools, etc.), more people found convenient to produce acoustic works. If a paper has no acoustic printouts, statistic tables, graphics most certainly the paper will not be accept for publication and even for presentation. This is an awkward situation inside phonetics. The auditory description of speech has been used for centuries, has sophisticated the methodology and produced very nice, original and consistent pieces of linguistic description.

3. A false conflict
The scenario presented above is typical of some groups of researchers and cannot be extended to the phoneticians in general. Congresses and periodical still accept papers based entirely on auditory researches. A short review of the history of phonetics shows that the conflict between acoustic and auditory approaches to phonetics in recent years is new and it has particularities not found in old times. As a matter of fact, since the time when the technology to study the acoustics of sounds were presented to phoneticians (beginning of XX century), they started to look at the speech differently. The introduction of such technology and the set up of phonetics laboratories obliged the researchers, trained to do auditory analysis, to sophisticate their work, introducing in parallel acoustic analysis. The good company were welcomed because it helped linguistics to be seen as a science in modern terms. Besides, phonology were the linguistic area that brought more significant contributions to this idea at that time. So it seemed obvious that speech should be treated acoustically to be more scientific and audibly to be able to produce good phonological analysis.

A good example of the marriage between acoustic and auditory data to produce linguistic analyses is the MIT Report Preliminaries to Speech Analysis: the distinctive features and their correlates by Jakobson, Fant and Halle (1951). The reason by which the old phoneticians work with the two approaches is the fact that they used to do good ear-training and performance courses when students (Cagliari, 2007: 51-65, 130-131). It was unthinkable to work in phonetics without such training. On the other hand, the phoneticians found in the acoustic analysis an indispensable tool to check their auditory analyses. The two approaches were complementary.

Fant (1960) set up a definitive acoustic theory of speech, but he acknowledged the importance of auditory based analyses to achieve good acoustic based data. He said:

Heliana Mello, Massimo Pettorino, Tommaso Raso (edited by), Proceedings of the VIIth GSCP International Conference : Speech and Corpora
“The rules relating speech waves to speech production are in general complex since one articulatory parameter, e.g., tongue height, affects several of the parameters of the spectrogram. Conversely, each of the parameters of the spectrogram is generally influenced by several articulatory variables. However, to establish and learn these analytical ties is by no means a hopeless undertaking. Some elementary knowledge in acoustics is valuable, but the main requirement is a sound knowledge of articulatory phonetics” (Fant, 1967: 95).

Gordon Peterson recognized the difficulty in doing phonetics: “... it is clear that phonetics is a discipline of substantial complexity requiring much further experimental and theoretical research” (Peterson, 1968: 171; see also Fry, 1973, 1979: 4). Ladefoged was conscious of the necessity to work with an auditory and an acoustic approach to describe adequately the sounds of speech. He said: “Understanding speech is, in essence, a process of obtaining information from an auditory stimulus. This process involves discriminating between some sounds and considering other sounds to be similar” (Ladefoged, 1967: 143).

And, in another place, he comments: “Furthermore, although we could (with difficulty) characterize all possible systematic phonetic contrasts entirely in physiological terms, it would be ridiculous to overlook the fact that some phonological rules obviously work in terms of acoustic properties of sounds” (Ladefoged, 1971: 4).

Many other phoneticians share the same scientific point of view. As a matter of fact it should be obvious to think in that way. But things have never been smoothly in scientific agreement. The conflict started when people stopped receiving good ear-training and performance courses when students, mostly because these phoneticians came from engineering areas, like telephony and communication, or even from linguistics, but getting a different phonetics education. Obviously, when an engineering or a phonetician look at a spectrogram or other kind of printout they need to listen to the sound recorded to proceed any type of analyses. It means, in other words, that they do use auditory analysis to carry on any kind of interpretation for any kind of acoustic parameters. So, the point we make in this paper needs to be better understood.

As we know, the speaker's intuition is an essential tool to check linguistic value of data and language rules, following the generative (Chomsky, 1965) and the functional (Halliday, 1970a) approach to linguistic analysis. The speaker's consciousness of language works differently in different levels of linguistic analysis. A person recognizes that the word horse does not apply to the object pencil, and so on. An English speaker knows that it is wrong to say: ball the kicked boy backyard the in. The correct is: the boy kicked the ball in the backyard. The intuition about the language structure works better in the area of semantics and syntax. It works rather well in relation to the phonological system of the language, but the same cannot be said when the intuition assess phonetic data. Without a good training in recognizing and producing the speech sounds according to the phonetic categories linguistically determined (cf. Catford, 1968: 309-333; IPA phonetic transcription symbols), the naive speaker may fall in many strange and erroneous conclusions about the sound he is inquired to explain. For instance, it is difficult for a person without specific training to categorize the vowels of a language, even when it is his native language. Ladefoged (1973) carried out a famous experiment in this respect and showed that phoneticians trained in the cardinal vowels system could agree in the identification of vowels quality. But phoneticians without such training committed many inaccuracies and mistakes. It is hard to convince at first sight a Portuguese speaker that he pronounces differently the "a" in words like mais and maus (a front and a back low vowel), because the language treats them as belonging to the same phoneme /a/. A phonetician without the appropriated training may describe these vowels acoustically as being unique. With this kind of analysis it is impossible to interpret the language sounds in appropriated terms. The criticism must be extended to all phonetic parameters. This is the reason by which some acoustic analysis does not reflect the linguistic rules of the language. Statistics cannot save a wrong basic phonetic interpretation.

Another aspect of the question (conflict) is crash between traditional phonetic theories based on linguistic approaches to language sounds and new acoustic theories proliferated recently. The discrepancy between the old and the new has been being as a motif to introduce a new theory if when the results of the analysis and interpretation of the data are in clear contradiction with the linguistic analysis. In this respect, for instance, some papers show an interpretation of pitch variations that mischaracterizes the stress system and the rhythm of the language, since the oscillation between peaks and valley are interpreted differently from the way the speakers of the language do. Obviously any linguistic analysis must always convince the native speaker that the analysis refers to his own language. The most notorious example, however, is the acoustic interpretation of the typology of rhythm for languages. What sounds reasonable to the speakers ears that the rhythm may not change when the cadence varies has being interpreted acoustically as a chaos. It is hard to believe how some phoneticians look only to statistic data and not to the music structure of speech. As a consequence of such awkward interpretation of the rhythm, other levels of phonetic and phonological analysis has generated awkward categories of data and rules for the language. It is absolutely naive to believe that an acoustic analysis is performed without an auditory analysis, based on specific training. On the other hand, if we can account for a good acoustic analysis of speech, why not use them?
4. Doing articulatory and acoustic analysis

It is perfectly possible to transform an acoustic interpretation into an auditory analysis. The job has been done adequately this kind of interchanging approaches permuting acoustic and auditory analysis might be carried out easily. If not, it means that there is something wrong. However there is cases in which one approach does not match exactly to another, for the specification lack of some essential parameter or for unacceptable procedures producing unacceptable results. In spite of that, a good acoustic analysis can be converted into an auditory interpretation and vice-versa. A good example is the analysis of intonation produced according to Pierrehumbert (1980) and Halliday (1970) theoretical models. The first is inserted into the generative grammar and the second into the functional approach to grammar. Pierrehumbert's model is essentially acoustical and Halliday's is an audible based model. In both cases we have a record of the utterance that can be interpreted in one or in another model. We converted Pierrehumbert examples into Halliday's analysis, taking as point of departure the location of focus. What comes before it constitutes the pretonic component in Halliday terms, and from the focus up to the end of the utterance there occurs the tonic component, and the definition of the tones. In the other way round we took Halliday's analysis and converted them into a sequence of High and Low pitch tones following Pierrehumbert theoretical framework. Some results of this job are presented as follow.

Figure 1: An utterance analyzed according to Halliday's theory and compared with the interpretation in Pierrehumbert's framework.

1 In the example, the tone values in Hz are: Mid-High: 160.55; 162.54; 150.94. Mid:112.60; 117.56; Mid-Low: 110.42; 101.69; 91.84.

Figure 2: PRAAT printout showing the intonational analysis of the utterance Você acha que vai dar certo?

Figure 3: Examples from Pierrehumbert (1980) interpreted according to Halliday’s (1970) framework

5. Conclusion

The constitution of linguistic corpora is as important as the theories which motivate and give them a scientific support. However it is useless to have a good corpus if there is no well trained phonetician to study it. Moreover, it is not enough to gather the required data in a good statistic program. A solid phonetic theory, compromised with the linguistic description of a language, is fundamental to produce nice pieces of work. It is generally admitted that phonetics science needs to take into account either auditory description and acoustic interpretation of speech. Other instrumental techniques are also complementary. Behind the action of viewing, hearing and interpreting speech data, there must always be the phoneticians’ mind and the phonetic skill, acquired through specific ear-training and performance training with somebody who knows how to conciliate auditory and acoustic analysis. This kind of training cannot be achieved exclusively by reading textbooks or practicing individually. In this regard, doing phonetics is very similar of doing music.

6. References


Resumo

O presente estudo objetiva analisar a localização do pico da F0 na sílaba nuclear das questões totais das capitais do sudeste brasileiro, utilizando o corpus do projeto ALiB. Os resultados apontam para uma diferenciação regional que opõe Rio de Janeiro e São Paulo a Vitória e Belo Horizonte.

Palavras-chave: entoação; alinhamento; prosódia regional.

1. Objetivo

O objetivo do presente trabalho é descrever o fenômeno entoacional do alinhamento em enunciados do tipo questão total, produzidos por informantes cultos dos quatro estados do sudeste brasileiro - Belo Horizonte, Vitória, Rio de Janeiro e São Paulo. Essas questões foram recolhidas do corpus do projeto ALiB. Considerando análise feita com amostras de fala de informantes não-cultos retirados da mesma matriz, postula-se que o referido fenômeno pode ser objeto de uma diferenciação regional.

Essa descrição ajuda a enriquecer o conhecimento a respeito da diversidade de traços que caracteriza a questão total do português brasileiro encontrada por Silva (2011).

2. Pressupostos teóricos

O alinhamento do pico localizado na última sílaba tônica do enunciado interrogativo está sendo estudado sob perspectivas fonológicas e fonéticas. Do ponto de vista fonológico, esse comportamento prosódico mostra-se como peça-chave para distinguir a pergunta do pedido no PB (Moraes & Colamarco, 2007). A oposição fonológica entre esses dois atos ilocutórios se realiza através de um movimento ascendente na curva da F0, quando se produz uma questão total neutra, e de um movimento descendente na curva da F0, quando se produz um pedido. Silva, Couto e Pinto constatam que os falantes nativos do PB transferem essa marca quando falam uma língua estrangeira. No espanhol, língua investigada pelas autoras, a pergunta é realizada por meio de um movimento ascendente. O autor cita o fenômeno do alinhamento do pico como sendo uma variante linguística encontrada em dialetos do Sueco e do dinamarquês. A respeito do Português do Brasil, Antunes (2011) faz um estudo comparativo preliminar entre a entoação de enunciados interrogativos e assertivos neutros de duas cidades de Minas Gerais: Belo Horizonte e Mariana, com base no corpus do projeto AMPER. A autora constata que, enquanto no falar de Belo Horizonte o alinhamento do pico ocorre à esquerda da sílaba tônica em enunciados interrogativos, em Mariana essa localização é simetricamente oposta, ou seja, à direita da sílaba tônica final.

Silva (2011), ao comparar as variedades faladas nas capitais brasileiras, descreve para região sudeste uma representativa quantidade de enunciados interrogativos cujo pico está alinhado à esquerda da última sílaba tônica. Observa-se no gráfico abaixo a proporção em que ocorre o alinhamento do pico à direita (vermelho), padrão mais comum, em comparação à realização do pico à esquerda da sílaba (azul). Observa-se em relação a este tipo de comportamento o seguinte contraste: 37% no sudeste e menos de 20% nas demais regiões. Chama-se a atenção para os dados de duas capitais: Vitória, onde o movimento descendente na tônica é o comportamento predominante; e Belo Horizonte, onde esse mesmo contorno se realiza em cerca de 50% dos dados.

Gráfico 1: Valores percentuais do alinhamento do pico nuclear na fala dos não-cultos

Alinhamento do pico

3. Metodologia

3.1 Dados

A amostragem é composta por 19 dados de fala semi-espontânea retirados do questionário de prosódia do ALiB, cujos enunciados são apresentados a seguir. Em cada tópico, aparece primeiro a questão total que se espera como resposta do informante e, em seguida, a diretiva que o inquiridor formula para obtê-la.

- Você vai sair hoje?
  Se você / o (a) senhor (a) quer saber se alguém vai sair hoje, como é que você / o (a) senhor(a) pergunta?

- Eu vou sair hoje, doutor?
  Uma pessoa está internada em um hospital e quer saber do médico se vai sair naquele dia. Como é que pergunta?

3.2 Perfil sociolingüístico do informante

Quatro informantes cultos e naturais de cada localidade equitativamente divididos em duas faixas etárias, 18 a 30 anos e 50 a 65 anos, e entre os dois gêneros.

4. Análise

Dividiu-se a duração da última vogal acentuada em três partes iguais, denominadas de esquerda, meio e direita. Aferiram-se os valores da F0 nesses três pontos a fim de verificar o comportamento da entoação neles. Entende-se que, ao localizar o ponto máximo, pode-se descrever, de forma mais detalhada, os movimentos ascendente e descendentes nessas sílabas, isto é, conhecer se o seu pico está alinhado no início, no meio ou no final da vogal.

5. Resultados

Os resultados da presente pesquisa confirmam o falso de que o padrão fonológico L+<H*L%, descrito por Moraes, não ser o mais expressivo em termos percentuais na região sudeste. A realização do pico ocorreu com mais frequência no meio da última vogal tônica, 47% dos dados, ficando o restante dos dados divididos entre o alinhamento à direita da vogal, 42% dos dados, e o alinhamento à esquerda da vogal, 11% dos dados.

Em termos relativos, Belo Horizonte foi a capital em que o pico alinhado no meio da tônica esteve mais presente. Ao contrário do que foi encontrado para os falantes não-cultos, não foi verificado nessa capital nenhum enunciado em que o pico tenha ocorrido mais próximo à fronteira esquerda do constituinte. Já em Vitória, 10% dos dados apresentaram o pico alinhado à esquerda e 15% dos dados, o pico alinhado à direita. O alinhamento do pico ao meio da sílaba também se mostrou predominante em Vitória, somando 75% dos dados. No Rio de Janeiro, o padrão ascendente da F0 descrito por Moraes foi realizado na maioria dos dados, totalizando 70% dos dados, embora tenha-se encontrado também 15% de dados com pico alinhado ao meio e 15% de dados com pico alinhado à esquerda da última vogal tônica. Em São Paulo, por fim, os dados estão equitativamente divididos entre alinhamento do pico no meio e alinhamento do pico à direita, não sendo encontrado para essa capital um movimento descendente na vogal tônica.

<table>
<thead>
<tr>
<th></th>
<th>Medidas na última vogal tônica nuclear</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Duração [s]</td>
</tr>
<tr>
<td></td>
<td>E</td>
</tr>
<tr>
<td>Belo Horizonte</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>0.19</td>
</tr>
<tr>
<td>Vitória</td>
<td>0.22</td>
</tr>
<tr>
<td>Rio de Janeiro</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>0.19</td>
</tr>
<tr>
<td>São Paulo</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
</tr>
</tbody>
</table>

Tabela 2: Valores de duração e F0 dos picos na última vogal tônica nuclear

Gráfico 3: Valores percentuais do alinhamento do pico nuclear na fala dos cultos
5.1 Algumas ilustrações

a. Alinhamento no meio da última tônica

Figura 1: Frase Cê vai sair hoje?, dita pela informante M1 de Belo Horizonte

b. Alinhamento à esquerda da última tônica

Figura 2: Frase Cê vai sair hoje?, dita pela informante M1 de Vitória

c. Alinhamento à direita da última tônica

Figura 3: Frase Você vai sair hoje?, dita pela informante H1 do Rio de Janeiro

6. Considerações Finais

O presente estudo corrobora o fato de os falares do sudeste brasileiro apresentarem particularidades prosódicas no domínio intrassílabico da questão total neutra. Nas demais regiões, Silva (2011) constata que, na fala dos não-cultos, o padrão descendente ocorre em menos de 20% dos dados, ao passo que na região sudeste esse quantitativo cresce para quase 40%. Para fala dos cultos, os resultados supracitados mostram que o alinhamento do pico no meio da vogal é predominante nas capitais de Belo Horizonte e Vitória, capitais essas que também apresentaram comportamentos semelhantes na fala dos não-cultos. Observou-se ainda que, no falar carioca, ocorre o predomínio do alinhamento à direita e que, no falar paulistano, os movimentos intrassílabicos ascendente e descendente dividem o nº de ocorrências.

7. Referências


Rhythm variation in spontaneous and induced speech

Amedeo DE DOMINICIS
University of Tuscia
Via del Paradiso, 47 - 01100 - Viterbo - Province of Viterbo - Italy
dedomini@unitus.it

Abstract

This paper investigates to what extent the metrics of speech can be induced by the pragmatic conditions of communication. We elicited two Italian corpora: the first was elicited by means of an experimental collaborative task; the other is a natural polemical interaction with overlapping turn-taking. The analysis of the former shows that the trend towards a syllable-timed or a stress-timed rhythm can be experimentally induced and it is an effect of the communicative interaction. The analysis of the polemical corpus demonstrates that the goal-relatedness of the speakers. The qualitative and statistic results confirm that no stable rhythmic pattern exists. Furthermore, the metric trend of each turn changes according to the conversational purposes: in particular, the speaker may borrow his interlocutor rhythm – or the opposite one – in order to collaborate with or to dominate him by cutting or easing the antagonist rhythm.

Keywords: rhythm; conversation; Italian.

1. Introduction

Two main approaches to linguistic rhythm exist in the literature: the hypothesis of rhythmic discrete types and the assumption of rhythm as a variable property which does not belong to the linguistic system, but to conversational interaction. In this latter approach the rhythm function is to handle cooperation and conflict among the speakers. Therefore it is not stable, but varies according to its conversational functions.

1.1 Rhythm-property of the system

The hypothesis of rhythmic types goes back to the forties (Lloyd James, 1940; Pike, 1945; Abercrombie, 1967; Faure, Hirst & Chafcouloff, 1980; Dauer, 1983). It mainly consists of a binary classification (syllable-timed/stress-timed languages). But it has not yet been clearly experimentally validated (e.g. Shen & Peterson, 1962; Bolinger, 1965; O’Connor, 1965; Uldall, 1971; Lea, 1974; Lehiste, 1977; Donovan & Darwin, 1979; Roach, 1982; Wenk & Wiolland, 1982; Borzone de Manrique & Signorini, 1983; Dauer, 1983; Drake & Palmer, 1993). According to a weaker hypothesis, rhythm is a perceptual impression arising from the convergence of some clusters of phonological properties typical of a given language (e.g. Dasher & Bolinger, 1982; Nespor & Vogel, 1986; Dauer, 1987; Bertinetto, 1981, 1989; Nespor, 1990; Ramus, Nespor & Mehler, 1999). The linguistic typology (syllable/stress-timed) is not discrete and different systems are spread out over a continuum1.

1.2 Rhythm-variable property of conversation

This hypothesis derives from conversational analysis studies, and represents rhythmic features in Gestalt terms. Recently, a new impetus has been given by the so-called phonetic-details studies (cf. Sacks, Schegloff & Jefferson, 1974; Erickson, 1982; Erickson & Shultz, 1982; Cutler, 1991; Couper-Kulhen, 1989, 1990, 1993, 2001; Buder 1986, 1991, 1996; Auer et al., 1999; Buder & Eriksson, 1997, 1999; Local, 2003; Fon, 2006; House, 2007; Russo & Barry, 2008; Arvaniti, 2009; Reed, 2010). In this paradigm, during interaction, rhythm may vary due to the conversational tasks, it is not a property of the system, but a tactical resource of the speaker2.

2. Experimental analysis

We ran an experimental test in order to verify to what extent the speech metrics can be induced by some pragmatic conditions. Two corpora have been elicited. The first corpus was obtained by an experimental collaborative task in which the subjects were asked to synchronize their speech with a recorded one. The second is a natural corpus in which two speakers are engaged in a polemical interaction (the so-called quarrel between Vittorio Sgarbi and Mike Bongiorno during the TV show Telemike in 1991). In each of the two corpora we took two measurements: the interstress intervals (henceforth “Acc” = the temporal distance between the stressed syllables); the syllabic intervals (henceforth “Syl” = the duration of stressed and unstressed syllables). These measurements were used to check the metrical typology (stress/syllable-timing) and its variation along the corpus.  

2.1 Experiment on the collaborative corpus

We recorded the sentence Il capostazione ha spento la luce (“The station master has turned the lights out”). Then we manipulated the signal in order to build new ones with constant Syl or Acc. On these signals we built a “Listen & Repeat” test. The working hypothesis was that listening to these signals will induce the listener to a syllable- or a stress-timing rhythm, according to the manipulated signals. To the same purpose, before the original signal,  

---

1 Also according to the PVI hypothesis (Low & Grabe, 1995; Low, Grabe & Nolan, 2000; Grabe & Low, 2002; Patel & Daniele, 2003), rhythm is an intrinsic property of the system.

2 To this paradigm three more branches belong: the studies on the metrical feet variability (heterometry) (Brown & Weishaar, 2010); the studies on rhythm as an entrainment phenomenon (Cummins & Port, 1998; Port, 2003; Cummins, 2009); the studies of rhythm as an Adaptive Oscillator (Port, Cummins & Gasser, 1996).
we inserted three beeps\textsuperscript{3}, 150 ms apart (equal to the mean Syl in the original signal) and 496 ms apart (equal to the mean Acc in the original signal). The signals for this listening test (\textit{passive corpus}) are listed in tab. 1. The \textit{active corpus} is composed of the sentences that the subjects recorded after listening to the passive corpus.

Five university students took part in the experiment: S1 (man, age 52, born in Rome where he lives); S2 (woman, age 24, born in Terri where she lives); S3 (woman, age 51, born in Orune-Nu, but living in Tuscania-Vt); S4 (woman, age 19, born in Alatri-Fr where she lives), S5 (woman, age 19, born in Civitavecchia where she lives). They listen to signal A, and are asked to utter the same sentence into the microphone (signal 1 of the active corpus). Then they listen to the further 4 signals and repeat them into the microphone, trying to imitate them and keep as close to the same timing of the signal they listened with headphones. Thus, the \textit{active corpus} contains 25 signals (5 per subject), as shown in tab. 1.

2.2 Experimental expectations and results

Compared to signal 1 (recorded at the beginning of the session), signals 4 and 3 should show equalized Acc durations and close to 496 ms (but the absolute value depends on the speaking rate). Signals 5 and 2, should show equalized Syl durations and close to 150 ms (but the absolute value depends on the speaking rate). If these expectations are confirmed, then the syllable-timed or stress-timed rhythm is an effect of the communicative interaction. As we are able to induce it, it is not a property of the linguistic system. The results validate these expectations. Tab. 2 shows the Syl and Acc durations in the active corpus, and their standard deviation ($\sigma$). Signals 2 and 5 systematically approach the reference value (150 ms) as compared to signal 1; likewise, 3 and 4 systematically approach the reference value (496 ms). Therefore the $\sigma$ decreases.

2.3 Experiment on the polemical corpus

As for the polemical corpus it is the quarrel between two Italian TV showmen: Mike Bongiorno and Vittorio Sgarbi \textit{(Telemike in 1991)} - \textit{\url{YouTube}}. It was downloaded from \textit{YouTube}. Its low audio quality creates no problem with the duration measurements. It is a communicative situation where the speakers do not collaborate, but manage to hinder and sabotage each other.

2.4 Experimental expectations and results

In the polemical corpus we expect a minimal degree of rhythmic integration: i.e. anisochrony. The results confirm these expectations. Indeed, no stable rhythmic pattern exists. Furthermore, the metric of each turn changes according to the conversational purposes; in particular, the speaker may borrow his interlocutor rhythm – or the opposite one – in order to dominate by cutting or easing the antagonist rhythm. The chaining and the syntagmatic succession of metrical types in each speech turn is a function of the conversational strategy of the speaker to create dominance. As is seen in tab. 3, at the beginning both speakers alternate different metrics, in a sort of “skirmish” (opposed rhythm speech turns). A “truce” follows: Sgarbi tries to disrupt the metrical strategy of Mike, using an asynchronous rhythm. Then, both speakers resume their confrontation, but change their tactics: there is the first instance of speech turns overlapping (“mimetic”) where their rhythm is common: stress-timed and synchronized. Next is a second overlapping (“rolling”) where, again, the metrics of the quarrellers are completely asynchronously, but dynamically tuned. Then, there are two “interval” speech chains: a stress-timed trend chain by Mike and a following one by Sgarbi, showing an opposed syllable-timed trend. Then, the “mimetic” tactic resumes, but the rhythmic features are reversed, as compared to the previous one: there is a third turns overlapping with a common syllable-timed and synchronized metric. In the “end”, three chains by Mike show an alternate rhythmical trend: stress- and syllable-timed. Four examples of these turns are given below\textsuperscript{4}.

1. \textbf{Truce-rhythm} (tab. 3: chains 14-15; fig. 1-2). Sgarbi shows no rhythmic isochrony: an extreme case of polemical strategy (maybe in order to sabotage Mike’s rhythm). His rhythm is anisochronous: the $\sigma$ values are very high both for the Acc and the Syl mean durations.

2. \textbf{Mimetic-rhythm} (tab. 3: chain 16; fig. 3). It is a first overlapping where both speakers tend to have a common stress-timed rhythm, and synchronized interstress boundaries: the difference between their mean Acc duration is not significantly different, as assessed by Student’s and Anova tests.

3. \textbf{Rolling-rhythm} (tab. 3: chain 17; fig. 4), a second overlapping. The turns are anisochronous (the difference between their mean Acc duration is significantly different, as assessed by Student’s and Anova tests), but with a peculiarity: both speakers undertake a sort of rhythmic “rolling relay”, where the metrics of the both quarrellers is \textit{dynamically tuned}: i.e. each turn takes up the Acc durational trend towards increasing or decreasing of the previous one, uttered by the interlocutor. As you see in fig. 5, Sgarbi produces a sequence of three increasing intervals (62-190-388 ms), followed by a reply by Mike with three equally increasing intervals (336-322-417 ms); then, Sgarbi reverses the trend, realizing a 170 ms interval and Mike pursues the decreasing trend with a 291 ms interval. Finally, Sgarbi reverses again the trend and produces a 236

\textsuperscript{4} The signals are annotated by means of 6 \textit{Praat Tiers}, as follows: (1) orthographic transcription: Mike, (2) orthographic transcription: Sgarbi, (3) Mike’s Syl: IPA and boundaries, (4) Sgarbi’s Syl: IPA and boundaries, (5) Mike’s Acc: IPA and boundaries, (6) Sgarbi’s Acc: IPA and boundaries.

\textsuperscript{3} At least three evenly spaced beats are required in order to establish an isochronous chain (Couper-Kuhlen, 1990: 16).
3. References


4. Appendix

Table 1: Passive and active corpora

<table>
<thead>
<tr>
<th>Signal</th>
<th>Passive corpus</th>
<th>Active corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>the natural original signal (writer’s voice)</td>
<td>the natural original signal (subject’s voice)</td>
</tr>
<tr>
<td>B</td>
<td>three beeps (150 ms apart) + signal A</td>
<td>signal recorded after listening to the signal B</td>
</tr>
<tr>
<td>C</td>
<td>three beeps (496 ms apart) + signal A</td>
<td>signal recorded after listening to the signal C</td>
</tr>
<tr>
<td>D</td>
<td>signal A: equalized Acc durations: 496 ms</td>
<td>signal recorded after listening to the signal D</td>
</tr>
<tr>
<td>E</td>
<td>signal A: equalized Syl durations: 150 ms</td>
<td>signal recorded after listening to the signal E</td>
</tr>
</tbody>
</table>

Table 2: Mean Syl/Acc durations (ms) & their std. dv. (σ)

<table>
<thead>
<tr>
<th>Signal</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ms</td>
<td>σ</td>
<td>ms</td>
<td>σ</td>
<td>ms</td>
</tr>
<tr>
<td>Signal 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>141.2</td>
<td>43.78</td>
<td>159.8</td>
<td>47.44</td>
<td>149.0</td>
</tr>
<tr>
<td>B</td>
<td>201.7</td>
<td>82.06</td>
<td>197.5</td>
<td>47.07</td>
<td>190.3</td>
</tr>
<tr>
<td>C</td>
<td>176.3</td>
<td>74.82</td>
<td>156.4</td>
<td>57.47</td>
<td>156.5</td>
</tr>
<tr>
<td>D</td>
<td>177.3</td>
<td>55.12</td>
<td>163.9</td>
<td>35.70</td>
<td>152.1</td>
</tr>
<tr>
<td>E</td>
<td>144.8</td>
<td>44.82</td>
<td>152.7</td>
<td>36.56</td>
<td>149.2</td>
</tr>
</tbody>
</table>

| Signal 2 |    |    |    |    |    |
| Signal 3 |    |    |    |    |    |
| Signal 4 |    |    |    |    |    |

Figure 1: Sgarbi: no non puoi dirle perché dici delle cazz ['no, you can’t tattle as you talk bullshit']

Figure 2: Sgarbi: questo è il concetto ['this is the idea']

Figure 3: Mike: Dici tu. Va bene. Io non dico nessuna cazz ['You say. Ok. I don’t talk bullshit’]; Sgarbi: Si d’accordo, le diciamo insieme, stiamo insieme a dirlo..io ['Yes, ok, we talk at once. We talk at once, ...']

Figure 4: Mike: Adesso parlo io. Si adesso mettila giù ['No it’s my turn. Yes now put it down’]; Sgarbi: Ora litighiamo. Vuoi fare a pugni con me? No puoi parlare. No ['Now we are going to have a quarrel. Do you want to box with me? No you can talk. No']

Figure 5: Rolling-rhythm. Variation of interstress intervals duration by Mike and Sgarbi, arranged in sequence

Figure 6: Mike: Io non ho ancora.. Io non ho ['I have not yet .. I have not.. ‘]; Sgarbi: Quelli che parlano a vanvera. È andato ['Those who blether. It’s gone’]
<table>
<thead>
<tr>
<th>Strategy</th>
<th>Chain</th>
<th>Mike</th>
<th>Sgarbi</th>
<th>Sylld</th>
<th>$\sigma$ (Sylld)</th>
<th>Accd</th>
<th>$\sigma$ (Accd)</th>
<th>T</th>
<th>ANO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skirmish</td>
<td>1</td>
<td>A</td>
<td></td>
<td>151.5</td>
<td>52.15</td>
<td>266.1</td>
<td>33.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>S</td>
<td></td>
<td>110.6</td>
<td>22.03</td>
<td>272</td>
<td>107.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>S</td>
<td></td>
<td>121.5</td>
<td>17.67</td>
<td>440.5</td>
<td>228.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>A</td>
<td></td>
<td>100.8</td>
<td>44.25</td>
<td>254.8</td>
<td>20.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>A</td>
<td></td>
<td>108.9</td>
<td>52.79</td>
<td>225.4</td>
<td>15.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>S</td>
<td></td>
<td>164</td>
<td>4.24</td>
<td>244</td>
<td>74.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>S</td>
<td></td>
<td>319</td>
<td>2.82</td>
<td>478.5</td>
<td>144.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>S</td>
<td></td>
<td>392</td>
<td>11.31</td>
<td>467</td>
<td>224.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>A</td>
<td></td>
<td>150.1</td>
<td>85.41</td>
<td>538.5</td>
<td>10.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>S</td>
<td></td>
<td>117.3</td>
<td>6.02</td>
<td>1 stress</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Truce</td>
<td>11</td>
<td>A</td>
<td></td>
<td>187.8</td>
<td>106.39</td>
<td>540</td>
<td>14.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>A</td>
<td></td>
<td>77</td>
<td>29.17</td>
<td>172.2</td>
<td>16.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>S</td>
<td></td>
<td>120.6</td>
<td>8.96</td>
<td>297.3</td>
<td>35.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>*</td>
<td></td>
<td>114.6</td>
<td>53.01</td>
<td>190.6</td>
<td>94.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>*</td>
<td></td>
<td>156.2</td>
<td>47.78</td>
<td>249.5</td>
<td>149.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mimetic</td>
<td>16</td>
<td>Ac</td>
<td>Ac</td>
<td>Mike: 180.4 Sgarbi: 144.3</td>
<td>Mike: 43.8 Sgarbi: 45.7</td>
<td>Mike: 376 Sgarbi: 345.4</td>
<td>Mike: 14.4 Sgarbi: 27.2</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td>Rolling</td>
<td>17</td>
<td>*</td>
<td>*</td>
<td>Mike: 126.8 Sgarbi: 128.3</td>
<td>Mike: 56.5 Sgarbi: 80.7</td>
<td>Mike: 346.2 Sgarbi: 217.6</td>
<td>Mike: 47.7 Sgarbi: 102.5</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>Interval</td>
<td>18</td>
<td>A</td>
<td></td>
<td>168.2</td>
<td>63.38</td>
<td>348.2</td>
<td>11.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>S</td>
<td></td>
<td>144</td>
<td>13.1</td>
<td>302.7</td>
<td>166.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mimetic</td>
<td>20</td>
<td>Sc</td>
<td>Sc</td>
<td>Mike: 181.7 Sgarbi: 166.5</td>
<td>Mike: 22.09 Sgarbi: 36.33</td>
<td>Mike: 566 Sgarbi: 494.3</td>
<td>Mike: 90.50 Sgarbi: 217.19</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>End</td>
<td>21</td>
<td>A</td>
<td></td>
<td>168.1</td>
<td>73.82</td>
<td>401</td>
<td>59.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>A</td>
<td></td>
<td>195.4</td>
<td>62.51</td>
<td>359</td>
<td>41.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>S</td>
<td></td>
<td>263.6</td>
<td>27.5</td>
<td>362</td>
<td>76.36</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Sequence of turns and rhythm trend. A = stress-timed; S = syllable-timed; * = asynchronous rhythm; c = synchronized durations; Sylld = syllables mean duration; Accd = interstress intervals mean duration; $\sigma$ = Std Dv; T = two sample Student’s t-test, assuming an unequal variance & a level of confidence $\alpha = 0.05$; ANO = one-way, or single-factor ANOVA ($\alpha = 0.05$). Non-significant values are on a grey background.
Some more applications of the tonal grid annotation

Amedeo DE DOMINICIS
Università degli Studi della Tuscia
Via del Paradiso, 47 - 01100 - Viterbo - Province of Viterbo - Italy
dedomini@unitus.it

Abstract

In this paper we give some more detailed evidence concerning the operating mechanism of the tonal grid annotation. Using some chunks of speech in different languages, we will show how a grid works to find the unexpected perturbations, their tonal shape, and the relations the grid establishes among the macroIP constituents which are their functors or bearing units. Then we will argue that the tonal grid is not only a mere annotation technique, but also (and above all) a new theoretical approach to understand the constituency of the Intonational Phrase (IP). Particularly, the architecture of the grid helps to find the relation between the F0 prominences (material prominences) and the prominences that result from the metrical or syntactic hierarchies (metalinguistic prominences) within the same IP or across IPs (i.e. within what we name a macroIP). The relation between two or more material and metalinguistic prominences identifies what we call a Nucleus.

Keywords: Tonal Grid; Nucleus; Intonation; Annotation.

1. Introduction

This paper aims at providing a theoretical setting for the tonal grid annotation system (De Dominicis, 2010a, 2010b). The tonal grid (henceforth TG) is a graphic device which represents linguistic phenomena such as (grammatically) unexpected and recurrent tonal and segmental perturbations, syntactic and lexical discontinuities, and pragmatic functions (e.g. the focalizations). Moreover, the TG reconstructs the relations among these phenomena: i.e. the possible tonal ‘rhyme(s)’ between two (or more) tonal perturbations at close/remote range, or the phoric relationship between constituents within the clause.

The TG is an upgraded version of the syntactic grid (Blanche-Benveniste, 1990, 1997; Blanche-Benveniste et al., 1979; Blanche-Benveniste et al., 1990; Bilger, 1982; Bilger et al., 1997), which is specially suited to highlight the disfluencies and the fragmentary nature of speech (false starts, hesitations, repetitions), and how they contribute to build meaning and grammatical functions. A syntactic grid consists of two main dimensions: the horizontal axis represents the sequence of the syntagmatic positions (or constituents); the vertical axis shows the possible different paradigmatic occurrences lying on the same position. By adding a syntagmatic construction to its paradigmatic fragments one gets a discursive configuration. It may recur at regular intervals, like a rhyme, and so give the discourse a peculiar architecture.

The TG supplements the syntactic one by adding tonal (or intonational) features; it highlights the recurrence of the same tonal pattern on different syntagmatic positions, or on the whole paradigmatic set of constituents belonging to the same syntactic position. In both cases, if a given tonal perturbation recurs, then each instance is an occurrence of a tonal pattern rhyme. Moreover, the TG brings up the recurrence of some special segmental perturbations: so these ones establish another kind of rhyme (or phoric relation). Inside a TG some places are provided, where several kinds of rhyme interface. These relations – specially if they link two or more F0 prominent positions (due to a focalization or to the metrical hierarchy) – identify what we call an intonational Nucleus: a new (relational) definition of the Nucleus in a macro-Intonational Phrase (macroIP).

2. Interfaces

The theoretical starting point of our approach relies on a multidimensional conict of the intonation theory (and of the Nucleus). Firstly, the reference units should be defined at the interface among F0 contour, metrical hierarchy, syntactic and pragmatic functions. Some special syntactic functions of the oral production (short and juxtaposed clauses, often without nominal constituents) should also be added to those interfacing functors.

Secondly, as for the intonation analysis, strictly speaking, we will consider tonal rhymes (i.e. non automatic, intentional tonal perturbations), and some (likewise intentional) segmental perturbations.

On the whole, the intonation is considered as the main mean for achieving the textual cohesion (Couper-Kuhlen & Selting, 1996; Ladd, 1996; Selkirk, 2000; Truckenbrodt, 2007; De Dominicis, 2009, 2010a, 2010b).

3. Case studies

The data concern Italian and English. They come from CLIPS corpus (dialogue “DGmtA01T”-Turin archive) and Human Communication Research Centre (HCRC) corpus.

The data have been analyzed by Praat. For each speaker the F0 has been extracted. Then both corpora have been syntactically annotated. As for the intonational annotation, the INTSINT system (Hirst & Di Cristo, 1998; Hirst, Di Cristo & Espesser, 2000) has been adopted, according to its automatic version (mono_momel-intsint.praat15). Using the typical Praat Tiers & TextGrid annotation system, the two labelling levels (syntactic & intonational) have been aligned to the audio signal. On this basis, the TGs have been built. They are shown in § 6.

The INTSINT system encodes the intonation curve as a sequence of tonal targets whose succession represents
the F0 contour in a stylised format. The symbols are as follows: “H” (Higher) and “L” (Lower); “S” (Same), “D” (Downstepping) and “U” (Upstepping). Three more symbols refer to the speaker tonal range: “T” (Top), “B” (Bottom), “M” (Mid).


The data are organized into the TGs as follows. The first line is the orthographic transcription. Each grid represents a clause (C) or an Intonational Phrase (IP). PROM means a point of tonal prominence. The speakers are labelled “P1” and “P2” (in Italian), or “G” and “F” (in English). The constituents that are marked by an intonational prominence are underlined.

The first case study is an Italian conversation and consists of three speech turns:

P2: passando sopra gli sci ? (‘passing over the ski?’) P1: no gli sci io non ce li ho (‘no the ski, I don’t have ski’) P2: io passo sopra gli sci o no ? (‘do I pass over the ski or not?’) 

The figures 1-2 show their F0 contours and three annotations tiers (from the top to the bottom: orthographic transcription, syntactic segmentation and INTSINT tonal annotation). In the figures 1 and 2 the system has not detected the final F0 rising (ranging respectively 25 and 38 Hz).

The tables 1-3 represent the three TGs that correspond to the three speech turns. Each speech turn corresponds to an intonational phrase (IP1-3). For each IP the grid shows the speaker (P1 and P2). In the grids the syntactic functions (for instance V, Prep, NP) and the tones on each constituent are annotated.

In these three IPs (and speech turns) we note two parallel phenomena referring to the promotion of some constituents along the hierarchical prosodic structure as a consequence of increasing their tonal prominence.

IP1 (by P2) contains the entry “sci” (‘ski’) in a PrepP syntactic position and with a B tone. In IP2 (by P1) it has a different syntactic function (in a VP), but it has a higher tone (H) and a PROM function. Finally, in IP3 (by P2) “sci” (‘ski’) is again in a PrepP syntactic position (like in IP1) and it is also PROM (like in IP2) with a U tone. Similarly, the same destiny applies to another entry. It is “io” (‘I’), the personal pronoun that occurs in IP2 (with a B tone) and it is repeated in IP3 (with a M tone): in IP2 it is not PROM, but it becomes so in IP3 where it also shows an increasing of tonal prominence. Therefore, two entries (“sci”, “io”) have the same parallel destiny: passing from one IP to another, and from one speaker to another, they increase both their degree of tonal prominence and their hierarchical rank in the prosodic structure. In the last IP (IP3 by P2) both attain the PROM position. So, in the same IP two PROMs cohabit.

According to the theoretical hypothesis of the present paper, only these two last PROMs (“io” and “sci” in IP3) are the real intonational Nuclei. The construction of their hierarchical position is the result of two cooperative strategies of the interlocutors: both a syntactic-lexical cross-reference architecture and a growing enhancement of the prominences. This macrostructure may be interpreted in a functionalist way, as a discursive cohesion mechanism, which results from an intra-speaker/inter-speaker involvement. A further remark concerning the linguistic theory of the intonational Nucleus may be added to this functional interpretation. The data show that the two Nuclei are established not on the basis of the simple physical prominence of F0 curve, but by means of a linguistic mechanism relying on the tonal relationship among constituents. This outcome is interesting and complex because the nuclear splitting (or doubling) is not supported by the predictions of the intonational phrasing theory. The two Nuclei govern a single macroIP which is composed of three IPs.

The second case study consists of two pieces of an English conversation:

G: (erm) have you got a collapsed shelter F: yes I do G: right G: you’ve to go up north and then round the collapsed shelter

The figures 4 and 5 show their F0 contours and three annotation tiers (from the top to the bottom: orthographic transcription of G and F, syntactic segmentation and INTSINT tonal annotation).

In the first four IPs (Tables 4-7) we remark a promotion of some constituents along the hierarchical prosodic structure as a consequence of increasing their tonal prominence. IP1 and IP4 (by G) contain the phrase «collapsed shelter» in the same syntactic position (NP). The repetition by the same speaker is complemented by an increasing tonal prominence: from the initial M to the final S (= H). Therefore, both are good candidates to be the functors of the relation constituting the intonational Nucleus of the macroIP, even if they are not the most prominent part of the F0 contour, in a physical meaning.

In the following four IPs (Tables 8-11) we remark two intonational prominences: the first on «plane crash» (NP in IP1 by G) and the second on «that» (PrepP in IP4 by F). Their tonal complements are M and H.

The increase of their tonal prominence results from the cooperation between the two speakers.

Therefore, their relation constitutes the real Nucleus of the macroIP.
4. Effects of the manipulation

In order to experimentally verify our hypothesis concerning the intonational Nucleus of the macroIP, we have manipulated the corresponding signals by erasing the tonal prominences that were originally associated with the nuclear constituents (figures 6-7). The tables 12-15 and 16-19 represent the corresponding four TGs.

The first HCRC conversation is an information exchange (“G: have you got a collapsed shelter? - F: yes, I do - g: right - G: you’ve to go up - G: north - G: and then round - G: the collapsed shelter.”). On the contrary, after the manipulation, it becomes something different: every relation between the “collapsed shelter” and “to go up north and then round” has disappeared. The conversation may be interpreted: “G: have you got a collapsed shelter? - F: yes, I do - G: right - G: you’ve to go up - G: north - G: and then round. The collapsed shelter...” - [8].

The second HCRC conversation is an information exchange too (“G: There a site of the plane crash - F: uh-huh - F: I’ve got that, I’ve got a site of plane crash - G: well, it’s just - G: below there - F: just below that”). On the contrary, after the manipulation, it becomes something different: F seems to be in need to know, whereas G does not collaborate. The conversation may be interpreted: “G: There a site of the plane crash... - F: uh-huh - F: I’ve got that? I’ve got a site of plane crash? - G: well, it’s just - G: below there! - F: just below that” - [9].

5. Conclusions

In this paper we gave some more detailed pieces of evidence concerning the operating mechanism of the TG annotation. Using some chunks of speech, in different types of language, we have demonstrated how a grid works to find the unexpected perturbations, their tonal shape, and the relations the grid allows to establish among the macroIP constituents that are their functors or bearing units.

Then we argued that the TG is not only a mere annotation technique, but also (and above all) a new theoretical approach to understand the constituency of the IP. Particularly, the architecture of the grid helps to find the relation between the F0 prominences (material prominences) and the prominences that result from the metrical or syntactic hierarchies (metalinguistic prominences) within the same IP or across IPs (that is, within what we name a macroIP). The relation between two or more material and metalinguistic prominences identifies what we call Nucleus of the macroIP.

The theory claims that the Nucleus must be one (per IP) and obligatory. So, in order to verify this outcome, we simply predict that by removing a single prominence (no matter what kind, material or metalinguistic, it is) there should not be a change of the phonological type of the IP (or macroIP), whereas by erasing all the prominences that enter into a relation to form a Nucleus, a categorical shift of the IP or macroIP (e.g. a change in sentence modality, or syntactic interpretation) would be triggered.

6. References

7. Appendix

Figure 1: First CLIPS conversation turn (P2). F0 contour, transcription, syntactic annotation, INTSINT annotation

Figure 2: Second CLIPS conversation turn (P1). F0 contour, transcription, syntactic, and INTSINT annotations

Figure 3: Third CLIPS conversation turn (P2). F0 contour, transcription, syntactic, and INTSINT annotations

Figure 4: First HCRC conversation. F0 contour, ‘g’ and ‘f’ transcription, syntactic, and INTSINT annotations

Figure 5: Second HCRC conversation. F0 contour, ‘g’ and ‘f’ transcription, syntactic, and INTSINT annotations

Figure 6: First HCRC conversation after manipulation

Figure 7: Second HCRC conversation after manipulation

Table 1: TG of the IP1: ‘passing over the ski?’

<table>
<thead>
<tr>
<th></th>
<th>passando sopra gli sci?</th>
</tr>
</thead>
</table>
P2 | passando sopra gli sci? |

Table 2: TG of the IP2: ‘no the ski, I don’t have ski’

<table>
<thead>
<tr>
<th></th>
<th>no gli sci io non ce li ho</th>
</tr>
</thead>
</table>
P1 | no gli sci io non ce li ho |
P2 | no gli sci io non ce li ho |
Table 3: TG of the IP3: ‘do I pass over the ski or not?’

Table 4: TG of the IP1.

Table 5: TG of the IP2

Table 6: TG of the IP3

Table 7: TG of the IP4

Table 8: TG of the IP1

Table 9: TG of the IP2

Table 10: TG of the IP3
A prosódia das interrogativas absolutas na fala carioca: leitura versus fala espontânea

Vivian PAIXÃO, Dinah CALLOU
UFRJ/Capes/CNPq
vivian_bp@live.com, dcallou@gmail.com

Abstract
This paper analyzes absolute/total questions in Rio de Janeiro dialect, taking into consideration the prosodic differences between isolated read sentences and the ones produced in spontaneous colloquial speech. Another topic brought into discussion is the kind of corpora for the study of spontaneous speech prosody, since, in general, non-controlled speech corpora have insufficient recording quality. Besides, it is difficult to find comparable samples, due to the variety of attitudes and emotions which determine speech prosody.

In the first stage of the study, interrogative sentences obtained in the UFRJ Acoustics Phonetics Laboratory were analyzed, as well as affirmative sentences were manipulated in order to make its melodic contour similar to the interrogatives. These manipulated sentences were submitted to perception tests with native listeners, who confirmed its authenticity as interrogatives in Portuguese, without significant differences in relation to the original questions. In a second moment, we studied spontaneous speech data, taken from the corpus of NURC / RJ project, and we compared the results of both corpora: the same patterns were found, however some differences were detected at the micro-melodic level.

Keywords: prosody; total questions; spontaneous speech versus reading.

1. Introdução

1.1 Objetivos
O objetivo desta pesquisa é analisar as chamadas interrogativas absolutas − segundo a terminologia sugerida por Font-Rotchés & Mateo-Ruiz (2011) − na fala carioca, observando as diferenças prosódicas entre as sentenças lidas de forma isolada e aquelas produzidas espontaneamente dentro do contexto de fala. Utilizou-se, para tanto, um corpus de fala lida, obtido em laboratório, e uma amostra de fala espontânea.

O trabalho visa também a discutir a questão dos corpora no estudo da prosódia da fala espontânea, uma vez que, nesta área de pesquisa, esbarra-se frequentemente com uma dificuldade técnica de gravação muito inferior, o que dificulta e chega a impedir, em determinados casos, a análise acústica de alguns trechos. Além disso, ao lidar com um corpus de fala espontânea, o pesquisador depara-se ainda com outro problema: a dificuldade de encontrar amostras comparáveis, devido à grande variedade de atitudes e emoções que influenciam na prosódia da fala. Trabalhos como o de Moraes (2006) mostram que, mesmo entre as interrogativas globais (yes/no questions), há uma variedade de contornos melódicos. A dependente do tipo de pergunta que representam (neutra, com desconfiança ou confirmativa, por exemplo), sentenças estruturalmente idênticas têm entonações distintas.

1.2 Trabalhos anteriores
O português brasileiro, e, mais especificamente, o português falado no Rio de Janeiro carece de estudos relativos à prosódia que utilizem dados de fala espontânea. Os trabalhos de Moraes (1996, 2006, 2008), elaborados a partir de dados obtidos em laboratório (fala controlada) são algumas das maiores referências para o estudo da entonação modal a partir da análise acústica.

O recente trabalho de Silva (2011) traz um estudo comparativo entre todas as capitais do Brasil, mas utiliza um corpus de fala semiespontânea do projeto Atlas Linguístico Brasileiro (ALiB).

No que se refere especificamente às interrogativas na fala espontânea do Rio de Janeiro, além dos trabalhos aos quais este dá continuidade (Paixão, 2011a, 2011b), encontramos apenas o trabalho de Souza (1995), que utiliza o corpus do projeto NURC para analisar três tipos de perguntas (totais, disjuntivas e parciais). Quanto às perguntas totais, os resultados obtidos apontam diferenças em relação àquelas produzidas em contexto de leitura, analisadas por Moraes (1984). A pesquisa de Souza, assim como o presente trabalho, mostra que, no contexto de fala espontânea, as interrogativas não seguem um padrão melódico tão regular quanto as produzidas em contexto de leitura, e dá margem a uma investigação mais aprofundada acerca de quais seriam os fatores determinantes dessas diferenças.

2. Corpora
Os dados utilizados para este trabalho foram retirados de duas amostras distintas: uma de leitura de sentenças isoladas e outra de fala espontânea.

2.1 Corpus de fala controlada (leitura)
A primeira amostra, de fala controlada, é composta de treze sentenças interrogativas e outras treze afirmativas, que se diferenciam das perguntas apenas pela entonação. Todas as sentenças, gravadas no Laboratório de Fonética Acústica da UFRJ por uma informante do sexo feminino, tinham estrutura sintática similar, seguindo a ordem prototípica do português brasileiro sujeito-verbo. Procurou-se ainda controlar a tonicidade da última palavra das sentenças: cinco delas eram terminadas em
palavras proparoxítonas, quatro em paroxítonas e quatro em oxítonas – houve uma sentença a mais com proparoxítona para garantir que houvesse dados com esse tipo de palavra, uma vez que elas sofrem frequentemente processos de sincope e consequente ressilabação. O tamanho das frases também é similar: todas têm entre sete e dez sílabas.

A lista foi lida três vezes pela informante em diferentes ordenações, tendo sido a primeira e a última gravações descartadas.

2.2 Corpus de fala espontânea

O corpus de fala espontânea é composto de outras dez sentenças retiradas de inquéritos do tipo diálogo entre dois informantes (D2) do projeto NURC-RJ.

O número de sentenças analisadas é restrito devido a várias dificuldades encontradas na busca pelos dados e sua análise. As interrogativas globais são escassas nos corpora obtidos através de entrevistas; raramente ocorrem perguntas nos diálogos entre informante e documentador (DID) ou nas elocuções formais (EF: aulas e conferências). Quando isso acontece, em geral, consistem nas chamadas tag questions – esse tipo de pergunta, muito frequente no português brasileiro, tem uma conformação prosódica diferenciada, conforme observa Serra (2009) em capítulo dedicado exclusivamente a essas sentenças.

Nos inquéritos do tipo D2 (diálogos entre dois informantes), as interrogativas são um pouco mais frequentes, mas ainda escassas. Além disso, por se tratar de fala espontânea, muitas vezes não se respeitam os turnos e as falas são sobrepostas, inviabilizando a análise dos enunciados.

3. Metodologia

A metodologia utilizada na análise e na ressíntese dos dados foi a do Método de Análise Melódica da Fala de Cantero&Font-Rotchés (2009). Esse método permite comparar vozes de diferentes indivíduos (inclusive homens e mulheres), uma vez que os gráficos são plotados com base em números estandardizados, e não nos valores brutos de frequência da voz.

3.1 Análise acústica

A análise dos dados, feita através do programa Praat, deu-se da seguinte maneira: primeiramente, os enunciados foram segmentados em sílabas. Aferiu-se, então, a medida da frequência fundamental (F0) de um ponto central da vogal de cada sílaba. Nos casos em que havia uma sílaba mais prolongada, que apresentasse oscilação de mais de 10% na medida de F0 de uma mesma vogal (o que equivale a um semitom musical, aproximadamente), consideraram-se dois pontos na mesma sílaba.

Uma vez obtidos os valores absolutos em Hertz, fez-se a estandardização desses valores, isto é, mediu-se a distância tonal, em porcentagens, entre uma vogal e a vogal subsequente, para poder criar a curva melódica representada em um gráfico gerado pelo Microsoft Excel, como se vê no exemplo a seguir (Gráfico 1).

Gráfico 1: Curva estandardizada da sentença “A Priscila usa óculos?” (fala controlada)

3.2 Manipulação acústica (ressíntese)

As sentenças afirmativas gravadas pela informante foram submetidas à manipulação acústica, também no programa Praat, conforme descrito a seguir: segmentaram-se as sentenças em sílabas da mesma forma como havia sido feito com as interrogativas, e marcou-se um ponto na curva de F0 na vogal de cada uma das sílabas – com exceção daquelas em que havia uma diferença de mais de 10%, em que foram marcados dois pontos. Em seguida, cada um dos pontos foi deslocado para o mesmo valor em Hertz da sílaba correspondente na gravação original (interrogativa).

Dessa forma, a diferença na curva de F0 das perguntas originais e manipuladas foi, principalmente, a regularidade: nas interrogativas “verdadeiras”, temos uma curva mais sinuosa, com mais oscilações da frequência entre os pontos marcados para medição, enquanto nas ressintetizadas a curva mostra-se mais regular (cf. figura 1).

Figura 1: Comparação entre as curvas de F0 da sentença “A Luísa estuda música?” original (acima) - - - e ressintetizada (abaixo) - - -.

4. Resultados

4.1 Resultados com dados de fala controlada

Nesta primeira etapa, verificou-se um padrão melódico para as interrogativas condizente com aquele descrito pela literatura: há um primeiro pico seguido de uma descida e uma inflexão final ascendente, no caso de sentenças terminadas em palavras oxítonas; ou circunflexa, quando
a sentença é terminada em palavras paroxítonas ou proparoxítonas. Tanto no primeiro quanto no segundo picos, a subida representa um aumento de, em média, 30% em relação ao tom inicial da sentença.

Mesmo em se tratando de dados fornecidos por uma mesma informante, a similaridade entre os gráficos das diferentes sentenças é impressionante. Ao sobreporem-se os desenhos das curvas estandardizadas de duas sentenças com tamanho similar, percebe-se que elas são praticamente iguais. As diferenças, aparentemente, devem-se apenas à localização das sílabas acentuadas em cada uma delas (cf. Gráfico 2).

4.1. Manipulação acústica e testes de percepção

Os enunciados foram submetidos a testes de percepção com ouvintes nativos cariocas. Foi solicitado aos voluntários que ouvissem duas gravações: a primeira delas continha as perguntas originais. Em seguida, escutaram as sentenças ressintetizadas (afirmativas transformadas em interrogativas).

Os voluntários reconheceram todas as sentenças como sendo, indubitavelmente, interrogativas autênticas da língua portuguesa. Quando indagados quanto a possíveis diferenças entre as sentenças, os participantes da pesquisa disseram apenas ter achado aquelas pertencentes à segunda gravação (ressintetizadas) “menos enfáticas” ou “ditas sem muita vontade”.

Curosamente, uma das informantes declarou que as frases do “grupo 1” e do “grupo 2” haviam sido gravadas por pessoas de sexos diferentes, como se pode inferir a partir da resposta transcrita a seguir:

O grupo 2 tem uma voz mais intensa, transmitindo mais firmeza. Aparentemente, na frase 6, trocaram os grupos, sendo assim, quem fala a frase do grupo 1 é o integrante do grupo 2, e no grupo 2 quem fala a frase é a integrante do grupo 1.

Esta ouvinte, em particular, atribuiu à diferença entre as sentenças originais e ressintetizadas a uma diferença de qualidade vocal, e, mais do que isso, ao sexo do falante – a voz ressintetizada foi identificada como voz masculina, e a voz natural, como feminina. Embora essa opinião tenha sido emitida por apenas um dos voluntários, levanta um interessante questionamento sobre uma possível tendência da fala masculina a uma maior regularidade no nível micromelódico, isto é, a apresentar menos oscilações na frequência fundamental nos pontos da frase em que esta variação não é um parâmetro decisivo para a caracterização modal da sentença.

4.2 Resultados com dados de fala espontânea

Os resultados obtidos com os dados de fala espontânea, mesmo com todos os empecilhos que envolvem o trabalho com esse tipo de corpus, confirmam, em certa medida, aqueles da fala controlada.

Os gráficos de alguns enunciados não apresentam os picos tão destacados quanto nos dados de fala lida, mas ainda assim pode-se perceber uma primeira e uma segunda subida, de acordo com o esperado. O Gráfico 3, a seguir, ilustra a configuração um pouco mais planificada dos dados de fala espontânea, com um primeiro pico muito sutil, enquanto o Gráfico 4 assemelha-se mais aos de leitura, uma vez que tem os dois picos bem destacados.

5. Conclusões

5.1 Fala controlada (leitura)

Os resultados obtidos com os dados de fala controlada confirmam aquilo que a literatura afirma a respeito das interrogativas absolutas no português: a presença de um pico de F0 na tônica final é, certamente, a característica
mais marcante desse tipo de sentença. O pico inicial, no entanto, também se faz presente, embora bem menos destacado do que o final em alguns dados.

5.2 Fala ressintetizada

Com relação às sentenças afirmativas ressintetizadas, foi possível perceber que, mesmo realizando-se a alteração da medida de F0 em cada silaba, as diferenças no nível micromelódico (regularidade na oscilação de F0 entre os pontos demarcados na curva) foram suficientes para que ouvintes nativos as interpretassem como tendo expressões de atitudes distintas relacionadas à ênfase.

5.3 Fala espontânea

Os resultados com dados de fala espontânea, apesar de se assemelharem em certa medida aos de leitura, mostram menor oscilação da F0, o que pode se dever à velocidade de fala e à tendência a uma demarcação mais nítida, na leitura, dos contornos prosódicos que determinam a modalidade da sentença.

5.4 Questões levantadas

A resposta de uma das informantes do teste de percepção realizado com as sentenças ressintetizadas permitiu levantar a hipótese de que a fala masculina apresente, no dialeto carioca, menor oscilação de F0.

Além disso, a observação dos dados, em especial daqueles de fala espontânea, nos leva a pensar na questão da neutralidade de atitudes e supor uma possível influência de fatores de ordem morfossintática na melodia da sentença. É sabido que a ênfase em um dos elementos da frase é fator determinante de alterações em diferentes parâmetros acústicos (frequência, intensidade, duração). A hipótese que se pode aventar é a de que determinados itens lexicais ou classes de palavras possam conter uma ênfase inerente. Em outras palavras, levanta-se a possibilidade de que, ainda que o falante procure dizer a sentença de forma neutra, sem focalizar nenhum elemento, determinados itens tenham a tendência a receber maior ênfase, seja por sua carga semântica ou papel sintático, e o que interferiria no contorno melódico da sentença.

6. Ideias para a expansão da pesquisa

Conforme já se viu, a necessidade de realização de trabalhos referentes à prosódia da fala espontânea no português brasileiro esbarra na dificuldade de se trabalhar com os grandes corpora de fala disponíveis, como o do projeto NURC. Em se tratando de palavras, conhecendo-se a dificuldade é ainda maior, uma vez que esse tipo de sentença não é recorrente nas entrevistas.

Apesar disso, é possível elaborar um corpus de fala que concilie alto grau de qualidade técnica e espontaneidade de fala Pretende-se dar continuidade ao estudo das interrogativas com um corpus elaborado especificamente para isso, que se utilize de estratégias para induzir os informantes a produzir perguntas. Tem-se, por exemplo, o HCRC Map Task (Anderson et alii, 1991), utilizado por pesquisadores de diversas línguas, que consiste em uma tarefa a ser realizada em duplas e que, para tanto, exige que um participante faça perguntas ao outro. O corpus utilizado por Pinto (2009) na análise de transferências prosódicas também é uma ideia interessante: consiste em um “jogo da verdade” entre pessoas conhecidas.

A realização de um trabalho sobre as interrogativas no português falado no Rio de Janeiro a partir de um corpus específico é, portanto, necessária para se confirmar resultados aqui apresentados e expandi-los, na medida em que se podem controlar melhor a possível interferência de fatores de ordem morfossintática e lexical.

7. Referências


The perception of foreign accented speech. Segmental and suprasegmental features affecting the degree of foreign accent in L2 Italian

Elisa PELLEGRINO
University of Naples “L’Orientale”
Via Duomo, 219. 80138 Napoli
epellegrino@unior.it

Abstract

The factors affecting the degree of foreign accent have been a matter of debate for years. This study is intended to investigate the role of suprasegmental and segmental features that make L2 Italian speech perceptively deviating from Italian native speech. Fifty-six Italian listeners listened to the excerpts of read speech produced by 8 Chinese learners and 2 Italian native speakers and rated them for degree of accentedness. The L1 and L2 corpora were spectro-acoustically analyzed. At suprasegmental level, we calculated the following rhythmic and prosodic parameters: articulation and speech rate, fluency, tonal range, percentage of silence and mean duration of empty pauses. At segmental level, we measured the duration of stressed and unstressed vowels and syllables and the length of stressed open and closed syllable. We also considered the syllable composition and pronunciation mistakes. The comparison between the results of the perception test and the data from the spectro-acoustic analyses have shown that both some suprasegmental parameters (i.e. duration of stressed and unstressed vowels and syllables, percentage of mispronunciations mistakes) are relevant features differentiating the strength of perceived foreign accent of Chinese speakers.

Keywords: foreign accent; segmental and suprasegmental cues; Chinese-accented Italian.

1. Introduction

Researchers have generally come to a consensus that the age at which the acquisition of a second language begins may greatly affect the outcomes of the process itself. Late second language acquisition indeed has often been considered one of the primary factors preventing the attainment of a native-like proficiency especially at the level of L2 speech perception and production (Birdsong 2006; Matsuoka & Smith 2008). Nevertheless, there is no widespread agreement among researchers on the role played by segmental and suprasegmental cues in foreign accent detection.

Over the years, the bulk of studies on the perception and production of non-native speech have focused mainly on the segmental features deviating from the native speakers’ pronunciation. (Flege, Bohn & Jang, 1997; Flege, Munro & MacKay, 1995; Walley & Flege, 1998). Similarly, most recent theoretical models accounting for L2 speech production and perception (Flege, 2003; Best, 1995; Major, 2001) have examined above all the production and perception of segments and have investigated the phonetic transfers from L1 to L2.

As a consequence, for years the role played by the suprasegmental features of speech in the perception of foreign accent has been relegated to a subordinate position (Piske, MacKay & Flege, 2001). Nevertheless, in recent decades the trend has changed. Recent studies on second language acquisition, (De Meo & Pettorino, 2011, 2012; Horgues, 2010), on the perception of foreign accent (Boula de Mareüil et al., 2004; Boula de Mareüil & Vieru-Dimulescu, 2006; Marotta & Boula de Mareüil, 2010), research undertaken on speech synthesis (Magen, 1998; Munro, 1995; Ramus & Mehler, 1999), on automatic approaches for foreign accent identification (Piat, Fohr & Illina, 2008) argue for a major role of prosody in the perception of non-native speech and in the recognition of the foreign speaker’ L1. On the same wavelength there is the research on the relationship between foreign accent, communicative effectiveness, credibility, reliability and persuasiveness in L2 Italian (Pettorino et al., 2011; De Meo, Pettorino & Vitale, 2012). Such work based on a pragmatic and acquisitional approach assessed both qualitatively and quantitatively the role played by the single rhythmic and prosodic parameters in carrying out effective communication.

2. The study

2.1 Objectives

Since it was shown that the suprasegmental features of speech are as essential as segmental ones both in the perception of foreign accent and in the detection of non-native speakers’ mother tongue, in the present study we considered both levels of analysis.

In order to facilitate the reading of the study results, we have divided this article into two main sections. One is devoted to the identification of the suprasegmental correlates of foreign accentedness. The other is dedicated to figuring out the contribution of phonemic deviations to the strength of perceived foreign accent in Chinese-accented Italian.

2.2 Participants

In the study we recruited two groups of participants with distinct roles: 10 speakers and 56 native listeners.

2.2.1. The speakers

The group was composed of 8 non-native speakers (NNS) from China and 2 Italian native speakers (NS) from Campania Region.

The Chinese students ranged in age from 20 to 22 years of age. They had already studied Italian in China for three years and had attained an intermediate level of
linguistic competence (B1- CEFR). On their arrival in Naples, they were enrolled in a course of Italian specifically designed to help them improve Listening and Speaking skills. At the time of the test, they were following a study curriculum in Italian and Linguistic Studies at the University of Naples “L'Orientale”.

The Italian native speakers, aged between 23 and 26, were students of Foreign Languages and Literatures at the same University. They constituted the control group.

2.2.2. The listeners
The group of listeners was composed of 56 native speakers of Italian, from Campania region, ranging in age from 20 to 50 years old. Since the competence in the foreign speaker’ L1 and the familiarity with a specific foreign accent were proved to mitigate the strength of perceived foreign accent and to facilitate the recognition of the interlocutors’ provenience (Marotta & Boula de Mareüil, 2010), none of the listeners were competent in Chinese or familiar with Chinese-accented Italian.

2.3 Materials and Methods
In order to prevent the topic, the word length and the syntactic sentence structure from affecting the results of the study, the Italian native and non-native speakers were involved in a read speech task. The subjects were instructed to read a 50-word text on jet-lag symptom drawn from an Inflight Magazine but with simplified lexicon and syntax. The recordings were taken in single sessions with every speaker in an anechoic chamber, at 44.100 Hz sampling rate.

The Italian listeners listened to the single excerpts of read speech produced by the 10 speakers in a randomized order. Each speech sample was rated through an accent degree rating test, based on a four-point scale: 0= native speaker (N); 1= mild accent (M); 2= strong accent (S); 3= very strong accent (VS).

After the perception test, the corpora of L1 and L2 Italian were spectro-acoustically analyzed for single speech chains, that is the part of a spoken utterance comprised within two silent pauses (Pettorino & Giannini, 2010). For each speech chain, we measured the number of syllables really uttered, their duration, the lowest and highest f0 values, the occurrence of disfluencies, the length of silent pauses between the speech chains. On the basis of these measures, we calculated articulation rate and speech rate, fluency, tonal range, percentage of phonation time, silence time and disfluency time, mean duration of silent pauses.

At segmental level, we carried out analyses on the duration of stressed and unstressed vowels and syllables and the length of stressed open and closed syllables. We also considered syllable composition and pronunciation mistakes. We used the open-source softwares Wavesurfer 1.8.8 and Praat (v. 4.1) for speech analysis.

---

1 In Italian the stressed vowels and syllable are longer than unstressed ones. See a.o. Avesany et al (2006).

2.3.1. Brief description of the analyzed prosodic features

- Articulation Rate (AR) was calculated as the ratio between the number of syllables really uttered and phonation time (syl/s). It is considered as a qualitative index because it measures the level of accuracy of the articulatory gesture and gives indications on the spatiotemporal organization of speech. With high values of AR the perfect achievement of the articulatory targets is compromised. Lower values of AR, on the contrary, allow the articulators to perfectly reach the expected acoustic targets. It is a rather stable parameter because its variations are limited by the anatomical and physiological constraints of the phonatory organs.

- Speech Rate (SR) was calculated as the ratio between the number of syllables and total time of the utterance, including both silent pauses and disfluencies (syl/s). Unlike AR, SR is considered as a quantitative parameter that measures the productivity level of speech. Its variations depend on the number and length of silent and non silent pauses. The more and longer empty and filled pauses, the lower the speech rate.

- Fluency (F) was calculated as the ratio between the number of syllables and the number of speech chains (syl/SC). It measures the frequency of silence, indeed, the higher the fluency, the fewer the empty pauses. Like SR, fluency is not a stable index. Its variations can be ascribed to many sender-dependent factors such as his/her socio-cultural background, emotional state, degree of control of the communicative event, ability to organize discourse on line and intention to give emphasis to his/her own speech.

- Tonal range. This corresponds to the interval between the f0 minum and maximum and it was measured in semitones (st). Low values of tonal range signal a monotone and flat speech. On the contrary, a wide interval between f0 minimum and maximum shows a more varied and dynamic speech.

- Percentage of phonation time, silence time and disfluency time. Phonation time is constituted by the syllables actually uttered in the speech. The silence time, instead, includes empty pauses, that is respiratory and emphatic pauses. The latter are commonly used to give the speech more emphasis in order to attract the listeners’ attention on a specific portion of the discourse.

---

Unlike silence time, disfluency time comprises filled pauses such as false start, vocalizations, nasalizations, lengthening, repetitions, and corrections. These occur above all in spontaneous speech together with the words that speakers plan and utter. They may signal to the listeners when the speaker is uncertain, or when he/she has to make choice and the speech planning process is delayed. They may also inform about the speaker’s confidence in what he/she is saying.

3. Results

3.1 Perception test

The results of perception test have showed that 96% of listeners rated the two Italian participants as NS. The remaining 4% did not answer. As for the rates given to foreign speakers (Figure 1), 7 Chinese students out of 8 were unanimously recognized as NNS. Only the speaker no. 7 was rated as NS by 4% of listeners.

Fig 1: Accent rate for foreign speaker

If we consider the rates for degree of accentedness:

- Speakers no. 5 and 7 were rated “mildly accented” respectively by 89% and 64% of listeners.

- Speaker no. 6 was perceived with a very strong foreign accent by 57% of listeners. Only 25% rated her accent “strong” and 18% rated her speech as “mildly accented”.

- Speakers no. 1-4 and 8 were rated with “strong foreign accent” by more than 50% of listeners, with very high difference in percentage from those who rated their accent as “very strong” (about 30%) and “mild” (30-40%).

In order to determine the acoustic correlates of degree of foreign accent, the results of the perception test were then compared to the acoustic data.

3.2 Suprasegmental features

The comparison between the results of perception test and the data from spectro-acoustic analysis have showed some evident differences between native and non-native speech at suprasegmental level.

First of all, the speech rated as “native” is the one with the highest values of articulation rate, speech rate, tonal range and fluency (Table 1).

<table>
<thead>
<tr>
<th></th>
<th>AR (syll/s)</th>
<th>SR (syll/s)</th>
<th>F (syll/SC)</th>
<th>TR (st)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>6.2</td>
<td>5.1</td>
<td>13.8</td>
<td>9.5</td>
</tr>
<tr>
<td>M</td>
<td>4.6</td>
<td>4.0</td>
<td>13.2</td>
<td>8.2</td>
</tr>
<tr>
<td>S</td>
<td>4.1</td>
<td>3.6</td>
<td>10.7</td>
<td>7.7</td>
</tr>
<tr>
<td>VS</td>
<td>4.0</td>
<td>3.3</td>
<td>8.7</td>
<td>5.8</td>
</tr>
</tbody>
</table>

Table 1: Mean values of suprasegmental features per groups of speakers

Secondly, if we shift our attention to the supra-segmental correlates of mild, strong and very strong foreign accent, from Table 1 it is possible to infer that there are both some stable parameters within the three groups of speakers (M, S, VS) and some suprasegmental features that instead differentiate the three degrees of foreign accent.

The most steady parameters are AR and SR; their values indeed do not change meaningfully among the three groups of participants (Table 2).

<table>
<thead>
<tr>
<th></th>
<th>AR (syll/s)</th>
<th>SR (syll/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>4.6</td>
<td>4.0</td>
</tr>
<tr>
<td>S</td>
<td>4.1</td>
<td>3.6</td>
</tr>
<tr>
<td>VS</td>
<td>4.0</td>
<td>3.3</td>
</tr>
</tbody>
</table>

Table 2: Mean values of AR and SR per degree of foreign accent

All foreign participants, indeed, speak at an articulation rate of about 4.2 syll/s and at a speech rate ranging from 3.3 syll/s of the VS speaker to the 4 syll/s of M speakers. The stable values of AR and SR can be ascribed to the particular kind of speech, that is read speech uttered by speakers with the same L1 and level of competence in L2 Italian. Generally speaking, indeed, read speech is more uniform among speakers in terms of rate than other kinds of speech, such as spontaneous speech, because it does not involve on-line speech planning processes.

The suprasegmental features that instead differentiate the three degrees of foreign accent are tonal range and fluency. Table 3, on the next page, summarizes the values for these two parameters for degree of foreign accent.
Table 3: Mean values of fluency and tonal range per degree of foreign accent

<table>
<thead>
<tr>
<th></th>
<th>F(syll/SC)</th>
<th>TR(st)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>13.2</td>
<td>8.2</td>
</tr>
<tr>
<td>S</td>
<td>10.7</td>
<td>7.7</td>
</tr>
<tr>
<td>VS</td>
<td>8.7</td>
<td>5.8</td>
</tr>
</tbody>
</table>

M speakers have the broadest tonal range and are the most fluent readers. As shown in Table 3, they produce speech chains of about 13 syllables, instead S and VS speakers utter speech chains that are respectively 10.7 and 8.7 syllables long. The diverse values of fluency are due a different amount of silent pauses made by the 3 groups of readers (Table 4).

Table 4: Number and mean duration of silent pauses

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
<th>Mean duration (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>8</td>
<td>0.3</td>
</tr>
<tr>
<td>S</td>
<td>10</td>
<td>0.3</td>
</tr>
<tr>
<td>VS</td>
<td>12</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Even though silences do not diverge in their mean duration, M speakers pause less than S and VS speakers do. The reason for the discrepancy in the frequency of silences seems to be imputable to speakers’ adoption of diverse pausing strategies while reading the target text.

For example, M speakers pause between complete sentences separated by full stops and coherently with the thematic organization of the sentence. The S and VS speakers, instead, silence when there are sentence boundaries marked by full stops and when there are boundaries of lower syntactical levels, usually marked by commas in the text. VS speaker also produces empty pauses within a sentence, that do not correspond to any syntactic boundaries. These latter rather occur when she produces disfluencies like word repetitions. The utterance of extra-words make the number of syllable to utter increase and consequently the speaker is lead to silence even before a syntactic boundary.

The divergences between the three groups extends also to the composition of speech time (Figure 2). Higher percentage of phonation time to detriment of silence time and disfluency time signals to native listeners a gradual reduction of foreign accent.

3.3 Segmental features

At segmental level, we carried out contrastive analyses on the length of stressed and unstressed vowels and syllables\(^3\). We measured the duration of stressed open and closed syllables uttered both by Italians and by Chinese participants. The syllable composition and segments mispronunciations were considered too. The data concerning mean duration of stressed and unstressed syllables (Figure 3), stressed and unstressed syllable (Figure 4), stressed open and closed syllables (Figure 5) mirror the difference in the mean values of articulation rate between native and non native speakers (Table 2).

The higher the articulation rate, the shorter the vowel and syllable length.

Figure 3: Mean duration of stressed and unstressed vowels (ms)

Figure 4: Mean duration of stressed and unstressed syllables (ms)

Figure 5: Mean duration of stressed open and closed syllables (ms)

As shown in figures 3-5, there seems to be a direct correlation between the rate of accentedness and vowel and syllabic duration. The speaker with very strong foreign accent always utters the longest vowels and syllables, regardless if they are stressed or not. On the contrary, the milder the rate of foreign accentedness, the lower the difference in vowels and syllabic duration from L1 Italian speakers.

\(^3\) In Italian stressed are longer than unstressed vowels. For a review of the acoustic and articulatory differences between stressed and unstressed vowels, see a.o. Avesani et al. (2007).
However, if we consider the duration ratios between vowels and syllable uttered by NSs and NNSs (Table 5), the major differences between L1 and L2 speech lie in the articulation of unstressed vowels and syllable. These latter are much longer than those uttered by the native speakers.

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>S</th>
<th>VS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstressed Vowel</td>
<td>1:1.52</td>
<td>1:1.72</td>
<td>1:1.75</td>
</tr>
<tr>
<td>Stressed Vowel</td>
<td>1:1.09</td>
<td>1:1.20</td>
<td>1:1.31</td>
</tr>
<tr>
<td>Unstressed Syllable</td>
<td>1:1.34</td>
<td>1:1.45</td>
<td>1:1.48</td>
</tr>
<tr>
<td>Stressed Open Syllable</td>
<td>1:1.03</td>
<td>1:1.16</td>
<td>1:1.25</td>
</tr>
<tr>
<td>Stressed Closed Syllable</td>
<td>1:1.05</td>
<td>1:1.05</td>
<td>1:1.13</td>
</tr>
</tbody>
</table>

Table 5: Duration ratios between vowels and syllable uttered by the native speakers and foreign speakers per degree of foreign accent

The different values of vowel and syllable length between NSs and NNSs lead to a different syllable composition. On the one hand, we find the native speakers in whose syllables the consonantal component occupy the largest portion (54.4%). On the other, there are S and VS speakers. In their syllables, indeed, the vowel represents the longest sounds (S 48.1%; VS 48.5%). Mildly accented speakers lie in the middle. In their syllables, the consonantal part tends to be equivalent to the vocalic portion (50.6%).

Other factors affecting the performance of the Chinese participants are pronunciation mistakes. Their speech, indeed, is characterized by a great deal of the typical deviations that mark the interlanguage phonology of Chinese L2 Italian learners (Costamagna, 2011; Dal Maso, 2005).

In the corpus, for example, there is evidence of:

- phoneme substitutions (i.e. [‘ta:pido]>[‘la:pido]), due to the students’ difficulty to articulate the opposition [l]/[r],
- phoneme alterations (i.e. [akkom’pan:a]>[*akkom’pan:a]) imputable to the Chinese learners’ tendency to replace unvoiced stops with their voiced versions and
- phoneme insertion or deletion [d’general’mente]>[‘*dgeneral’mente]; [‘bordo]>[‘*b:do], depending both on the speakers’ tendency to simplify the Italian syllabic structure from CVC to CV and on their difficulty to produce syllable with vibrant or liquid codas.

Additionally, other attested phonemic deviations were concerned with accent shift and geminate timing. Some paroxytone words like [‘sindrome], for instance, were treated as paroxytones [‘*sin’dro:me], while some geminates were uttered as singleton (i.e. [ap’e:ti:to]>[“ap’e:ti:to]).

Nevertheless, among the segmental errors mentioned above, those that most significantly affect the native listeners’ rate of accentuatedness were shifts of accents and misarticulation of long consonantal sounds. The lower the percentage of these two kinds mistakes, indeed, the milder the rate of foreign accent (Table 6).

Table 6: Percentage of pronunciation mistakes per degree of foreign accent

<table>
<thead>
<tr>
<th>Kind of mistake</th>
<th>M</th>
<th>S</th>
<th>VS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insertion</td>
<td>4.4%</td>
<td>1.6%</td>
<td>1.7%</td>
</tr>
<tr>
<td>Substitution</td>
<td>2.2%</td>
<td>3.3%</td>
<td>2.6%</td>
</tr>
<tr>
<td>Deletion</td>
<td>0.4%</td>
<td></td>
<td>0.9%</td>
</tr>
<tr>
<td>Alteration</td>
<td>1.7%</td>
<td></td>
<td>1.2%</td>
</tr>
</tbody>
</table>

Table 7: Percentage of pronunciation mistakes per degree of foreign accent and kind of deviation

However, the total number of pronunciation mistakes seem to affect the listeners’ rate. The speaker with very strong foreign accent is the one with the highest percentage of segmental errors (17.6%). The S and M speakers, instead, produce the 15.1% and 11.4% of pronunciation mistakes.

4. Conclusions

The role of segmental and suprasegmental features in the perception of foreign accent has been a very controversial issue. For years research interests have focused on segmental deviations from native pronunciation and on phonetic transfers from L1 to L2.

In the light of the recent increasing attention paid to the contribution of prosody to the perception of foreign accent, the present study was intended to determine both geminate consonant is shorter than the vowel preceding a singleton (Bertinetto, 1981; Zmarich & Gili Fivela, 2005).
the suprasegmental features and the segmental ones leading Italians to formulate the rate of foreign accent when listening to Chinese speakers of L2 Italian.

The comparison between the results of perception test and data from the spectro-acoustic analyses of L1-L2 Italian corpora has revealed that both levels play a role in influencing the rate of perceived foreign accent.

At suprasegmental level, fewer silences, fewer disfluencies, higher fluency and wider tonal range are perceived as signals of mild foreign accent.

At segmental level, instead, the speakers with native-like pronunciation are those whose speech was characterised by shorter duration of stressed and unstressed vowels and syllable, lower percentage of mispronunciation mistakes and errors relating to gemitine timing and shift of accent.

5. References


Pettorino, M., Giannini, A. (2010). Il parlato dei mass


Perceptual competence and persuasiveness: L1 and L2 compared

Luisa SALVATI¹, Anna DE MEO², Massimo PETTORINO³

¹University for Foreigners of Siena; ²University of Naples “L’Orientale”

Abstract

This research aims at analysing the perception of prosodic features in learners of L2 Italian, from a comparative perspective with L1 Italian. In particular, we chose spontaneous argumentative speech, which implies the perlocutionary act of convincing, in order to investigate the relationship between the degree of persuasiveness of a speaker and the prosodic features characterizing her/his speech, in relation to the perceptual competence of non-native learners. A corpus of argumentative speech in L1 and L2 Italian has been collected. For the corpus in L1, 8 Italians, divided into two groups, were asked to take part in a debate and argue for or against a specific topic. The aim was to convince an audience of 19 Italians, who evaluated the persuasiveness of each speaker, judging it as “positive” or “negative”. For the corpus in L2 Italian, we carried out the same procedure with 10 Chinese learners of Italian, who argued (5 pros, 5 cons) in front of an audience made up of 8 Chinese people. The data obtained are significant because they show not only that there is a relationship between persuasiveness and prosodic features, but that this relationship is strongly influenced by the perceptual competence of the listeners.

Keywords: prosody; perception; second language acquisition; persuasiveness.

1. Introduction

The prosodic competence in a second language is the result of a complex of variables, such as the quantity and the quality of exposure to the second language, the way of using L1 and L2, the language learning pathways, and the individual differences in terms of motivation, attitude, affective filter and age. The last factor is probably one of the influential: the period in which an individual can develop the same skills of a native speaker is limited to the first years of life. After this phase, it is very difficult that a non-native speaker is able to acquire an L2 prosodic proficiency comparable to that of a native speaker (Birdsong, 1999).

To these variables it must be added the influence of the L1 prosodic models on the L2 perception. Some studies have recently focused on the influence of the perceptual segmentation and the resulting phonetic and phonological identification of the acoustic elements on the rhythmic organization which characterize the speech production in the various languages (Flege, 1991; Best & Tyler, 2007). In the case of a foreign language acquisition, it seems that a high degree of typological similarity between the languages in contact may cause positive transfer for the learning of morphosyntax, vocabulary and pragmatics, while a negative impact of the L1 or of other known languages may occur with regards to L2 pronunciation. Flege (1987) notices that this influence is also active on the L2 perceptual competence, since learners have real difficulties in discriminating the L2 sounds, particularly if they are similar to those of the native language.

1.1 Perceptual competence of Chinese learners of L2 Italian

The Chinese is a tonal and isolating language, typologically distant from Italian. Therefore, when dealing with the study of the Italian language, Chinese learners spend a lot of time trying to understand a language which is completely different from their L1, unless they have previously learned another foreign language typologically close to Italian. In particular, from the point of view of oral comprehension, Costamagna (2011) states that Chinese learners access to speech understanding with great difficulty, because they are unable to perceive and segment the Italian speech chain. The development in comprehension is also influenced by the Italian morphological organization: in the early stages, Chinese learners try to grab the prominent elements that can facilitate the comprehension, as they perceive the linguistic message in L2 as an indistinct mass of sounds without distinguishing the discriminatory elements. In more advanced levels, they develop a greater awareness of the distance existing between the two languages, above all as regards the prosodic structure. The skill of using variations of intonation for pragmatic purposes can be seen only in advanced levels, since in the early stages, they generally recognize interrogative and exclamative sentences.

Therefore, what characterizes the perceptual competence in Chinese learners of L2 Italian is a little progression from one stage of interlanguage to the other one, as shown by De Meo & Pettorino (2011) in a study on the relationship between language proficiency and prosodic competence. A Chinese can achieve a C1 high level of language competence (C1 level of the Common European Framework of References - CEFR) and, at the same time, not adequately develop the ability to communicate effectively with Italian native speakers using the appropriate prosody and intonation. Oral comprehension is also delayed by the different Chinese and Italian pragmatic-communicative models and this often makes the oral interaction in L2 Italian difficult (Costamagna, 2011).
2. Material and method

The present research aims at analyzing the perception of rhythm and prosodic features in Chinese learners of L2 Italian, in a comparative perspective with the Italian native speakers. Using the task of the debating the relationship between the degree of persuasiveness achieved by the speaker and the related rhythm and prosodic features of her/his speech was investigated.

It should be clarified that the study was carried out with the awareness that the argumentation and a speaker’s persuasiveness is the result of a series of elements: the content of the text, the way the speaker expresses her/his opinions, the body language. Given these variables, the prosodic component was isolated to verify its influence on the ability to persuade the audience, not only because, through the voice, the speaker can arouse emotion and, therefore, persuade, but also because the voice may be spectro-acoustically analysed, allowing measurable and comparable results.

2.1 The debating structure

The debating is not simply a discussion where speakers argue about a topic, but it is rather an interactive exchange of ideas, with a strict protocol of rules which imply the alternation of arguments for and against a given topic, imposing a time limit to respect and finally involving the audience judgment called upon to evaluate individual speakers on linguistic, paralinguistic and extralinguistic parameters. When the debating involves also foreigners who argue in L2, it becomes an intercultural interaction between natives and non-natives, who are characterized by different linguistic behaviours and cultural backgrounds. In this perspective, the features normally defining the debating become even more complex because of cognitive factors related to language learning processes, sociolinguistic and sociopragmatic factors. This type of arguing involves intercultural communication skills as well, i.e. proper skills to interact by negotiating meanings, values, symbols, ideas, on a “common ground” (Fetzer & Fischer, 2006) between natives and non-natives.

For this research, a debating was held between a team of Italians and a team of Chinese learners of Italian. The debating took place in two phases. In the first one, chaired by a moderator, members of each team alternatively argued on the topic, having a time limit of two minutes. In the second phase, both groups had a time limit of six minutes to discuss freely, without any moderator, in order to convince the audience.

2.2 The corpora

The corpus in L1 and L2 Italian was audio-recorded using Goldwave 5.58 and videotaped by a Sony handycam HDR-SR8E and then orthographically annotated on the basis of the indications given by the CLIPS project “Lexicons and Corpora of Written and Spoken Italian” (Albano Leoni & Giordano, 2005). Here we will refer to the spectro-acoustic analysis conducted, using Wavesurfer 1.8.8, on the corpus recorded during the first phase of the debating, where each speaker talked without being interrupted.

For each speaker measures were performed in order to determine the number of speech chains, the number of syllables for each speech chain, the duration of each speech chain, the duration of silent pauses, the duration of non-silent pauses or disfluencies, the maximum and the minimum $f_0$ value for each speech chain. Furthermore, for each speaker the following calculation were carried out: articulation rate AR, i.e. the ratio between the number of syllables and the speech chain duration (syl/s), speech rate SR, i.e. the ratio between the number of syllables and the utterance time (syl/s), fluency (F), i.e. the ratio between the number of syllables and the number of speech chains (syl/SC), the percentage of silence duration, the mean duration of silent pauses (s), the percentage of disfluencies duration, the tonal range, i.e. the difference between the maximum and the minimum $f_0$ value in an utterance, measured in semitones (st) in order to compare data relating to different speakers.

2.3 The native and non-native participants

The Italian speakers were three female and one male university students, aged between 20 and 25, all coming from the Campania region (southern Italy). The Chinese participants, two male and two female students of Italian at the University of Tianjin, aged between 20 and 25, who had been living in Naples for four months, had a language competence of Italian corresponding to B2 level of CEFR. Before the debating, rules were explained to both groups separately and some tips on how to practice for the discussion, both individually and in groups, were given. Moreover, a large part of this introductory phase was dedicated to comment on the parameters the speakers would have been judged on: persuasiveness, voice volume, speech rate, pauses, intonation, posture and gaze, gestures, language use and competence. Afterwards, several debating simulations were held. For the Chinese learners, a textbook aiming at development of the argumentative skills in L2 Italian was used (Barki & Diadori, 1994).

3. The perception of persuasiveness in the L1 corpus

The corpus in L1 Italian consists of a debating between native speakers (NS) in front of an audience of native listeners (NL) about the following topic: “It is better to live in a big city”. The team in favor was made up of one man and three women, while the team against was formed by 4 women. The audience, consisting of 19 NLs, male and female, had to judge the persuasiveness of each speaker in terms of “positive” or “negative.”

To investigate the perceptual level, the persuasiveness degree of each speaker was related to the prosodic features of her/his speech, in order not only to verify the existence of a link between persuasiveness and prosody. The most significant relationships were found
between persuasiveness and AR (Figure 1), fluency (Figure 2), mean duration of silent pauses (Figure 3), disfluencies (Figure 4).

The graphs show that the Italian listeners tend to accept an argumentation pronounced with a greater articulatory accuracy and many medium-long silent pauses, which may give them time to think about what they have just listened: the more the Italian speaker produces long silences, the more the native listener perceives him/her as more persuasive. Instead, persuasiveness decreases if disfluencies increase, as if the native listener perceives those silent pauses, which are used to fill the spaces between sentences, as disturbing elements.

In conclusion, from the results it is possible to assume that a native listener tends to perceive an hyper-articulated speech with many long silent pauses and few disfluencies as more persuasive. Furthermore there are not significant relationships among persuasiveness, speech rate and tonal range.

4. The perception of persuasiveness in the L2 corpus

The corpus in L2 Italian consists of the debating between two groups of non-native speakers (NNS) in front of non-native listeners (NNL). The team in favour consisted of four female and one male Chinese students, while the team against was made up of three male and two female Chinese students. In order to eliminate, as far as possible, the text variable, the assigned topic was the same as the one used in the previous L1 debating.

In this section, the relationship between the prosodic features characterizing the speech of NNSs and their ability to persuade a non-native audience will be analysed. Data were used to evaluate if the L1 and the L2 debating share the same characteristics, and to determine how NNLs perceive their peers speaking in a foreign language. In literature there are very few studies which deal with these questions and they mainly relate to foreign languages others than Italian.

Results show that the 84% of NNLs judged in a very positive way all the speakers, regardless of prosodic features. However it is worth reflecting upon how the relationship between persuasion and prosody is related to the perceptual ability of the listener. Indeed, a comparison between NLs evaluation (in the L1 debating) and that of the NNLs (in the L2 debating) points out that, while the NSs perceive a clear relationship between persuasiveness and the related prosodic features, the Chinese learners competence does not seem enough to detect a significant connection between prosody and persuasiveness. With regards to this difference, it can be assumed that there are
two co-existing causes. On the one hand, the assignment of a judgment on persuasiveness involves a four-step process: listening to speech, understanding the acoustic message, comparing it with one own opinions, and finally giving the judgment. It seems that the non-native learner pays more attention to single words rather than to the argument as a whole, unlike the NL, who has the tools to reach the next phases of the comprehension process. The speech perception in L2, indeed, is strongly influenced by the mother tongue prosodic structure, which may affect the learner’s oral comprehension ability. Chinese learners, who have a native language characterized by rhythmic and intonation structures very different from Italian language, access to speech perception with great difficulty, because they are unable to perceive and segment the speech chain effectively.

On the other hand, there are idiosyncratic sociolinguistic and cultural mechanisms in the NNSs: from this perspective, the Chinese students positively evaluate their peers to reward the effort and the commitment they face dealing with another language. The development of the L2 perceptual competence, therefore, slows down because of the different pragmatic-communicative patterns of the learners.

The combination of these two elements - one cognitive, the other one socio-linguistic - leads the NNSs to identify with difficulty the suprasegmental components and their pragmatic value. This is even more interesting when we consider that the CEFR, with reference to the listening comprehension skills of B2 learners, indicates that s/he is able to understand the main ideas of a complex text on both concrete and abstract topics, including technical discussions in their field of specialization.

Considering the results obtained by this research, it can be added that an L2 Italian learner, with the so-called autonomy level of a language knowledge, is able to perceive and decode complex messages, but s/he is less able to evaluate them in terms of persuasiveness. The data shed new light on the studies regarding perceptual competence from an acquisitional point of view and on the ability of oral understanding. They also reveal a certain lack of attention to the prosodic dimension of L2 communication, both in acquisition and teaching, and finally, in the assessment field, because of the absence of any reference to language suprasegmental aspects in the CEFR descriptors.

5. Conclusion
The task of the present study was to analyze the perception of rhythmic-prosodic features in Chinese learners of L2 Italian in argumentative speech, from a comparative perspective with L1 Italian. To this purpose, a relationship between prosody and persuasiveness was outlined: it emerged that Italian listeners find most persuasive a well structured L1 speech, with many long silent pauses and few disfluencies. These data about spontaneous speech confirm the research carried out by De Meo et al. (2011) on read speech.

Instead, with regard to the non-native speakers, for spontaneous argumentative speech, there is no significant relationship between persuasiveness and prosody, since the Chinese students have always attributed highly positive evaluations, which do not allow detecting a trendline that can link the above variables. Regarding this issue, this study proposes two explanations, one cognitive and the other cultural-pragmatic.

Further research could have repercussions in the field of language teaching, since the oral texts administered to learners should be constructed, adapted and chosen not only on the basis of morphosyntactic structures and language functions, but also according to the various levels of perceptual competence that the L2 learners develop. Finally, it is interesting not only to extend the investigation to the relationship between persuasiveness and textual/kinesic variables, but also to study the link existing with the prosody by the technique of low-filtering, in order to eliminate other variables.

6. References

Medidas da variação prosódica diatópica no espaço românico
Antonio ROMANO, Lurdes de Castro MOUTINHO, Rosa Lídia COIMBRA, Albert RILLIARD
Laboratorio di Fonetica Sperimentale “A. Genre” - Dip. di Lingue e L.S. e C.M. - Università degli Studi di Torino (Itália); Centro de Línguas e Culturas - Universidade de Aveiro (Portugal); LIMSI-CNRS - Orsay (França) antonio.romano@unito.it, lmoutinho@ua.pt, rlcoimbra@ua.pt, albert.rillard@limsi.fr

Abstract
The aim of this article is to propose an experimental method for automatic assessment of prosodic similarities between dialects within a large linguistic domain (Romano et al., 2011; Moutinho et al., 2011). Data have been collected in the framework of the international AMPER project (Atlas Multimédia Prosodique de l’Espace Roman / Atlas Multimédia Prosódico do Espaço Românico) and measurements taken into account for this experiment refer to varieties of European Portuguese and regional Italian varieties. General indexes such as coherence and congruence have been tested and, between different varieties, prosodic similarity is measured on the basis of a weighted correlation formula providing elements for the definition of dialectometric distances. Italo-romance dialects were also considered in some case in order to enlarge the testing to the assessment of prosodic persistence between similar languages spoken by the same speakers (Romano, 1999). Since intonation within the Romance domain may show different patterns, this study is intended to provide useful elements explaining how these patterns could define homogeneous contiguous areas vs. discontinuous dialectal spaces or converging solutions between separate regions.

Keywords: Prosody, Dialectometry, Italo-Romance, European Portuguese.

1. Introdução

Na verdade, a entoação no espaço linguístico românico apresenta diferentes esquemas, marcados diatopicamente, que podem ser convergentes ou divergentes de uns falares para outros no mesmo domínio linguístico (Contini, 1992; Romano, 1999; Contini, 2008; Contini et al., 2008; Turculeţ, 2008) e o presente estudo propõe um método de avaliação objetivo destas dinâmicas.\(^1\)

2. O projeto AMPER
A participação no projeto AMPER permitiu-nos a obtenção de dados prosódicos comparáveis num conjunto de pontos bastante denso em algumas das regiões linguísticas que fazem parte deste Projeto.\(^2\) Depois de termos trabalhado sobre um conjunto limitado de frases, com comparações escolhidas entre perfis duma seleção de variedades italianas, portuguesas e brasileiras (Romano, 1999; Romano & Moutinho, 2004; Interlandi et al., 2007; Felloni, 2011), pretendemos, neste artigo, discutir resultados de análise relativos a semelhanças e diferenças manifestadas nas configurações entoacionais obtidas para numerosas estruturas frásicas em diversas variedades do Italiano (em AMPER-ITA) e do Português europeu continental (em AMPER-POR).

---

\(^1\) Ainda que baseado num trabalho comum dos 4 autores e numa versão preliminar coletiva, para fins académicos explicitamos que a versão final deste artigo foi redigida por A. Romano (§§2, 4.1 et 5), L. de Castro Moutinho (§§3), L. de Castro Moutinho e R.L. Coimbra (§1 et 6) et A. Rilliard (§4.2).

\(^2\) Uma parte destes dados foram recentemente publicados (DVD AMPER 2011). O tratamento em curso de umas duas dezenas de outros inquéritos deveria melhorar, nos próximos meses, a cobertura do AMPER-ITA, nomeadamente a secção do projeto consagrada aos falares italo-romanos e às variedades do italiano regional. Também no âmbito do AMPER-POR (português brasileiro e europeu), estão em curso novos inquéritos e análise de dados, que serão objeto de nova publicação.

---

Figura 1: Os dados do DVD AMPER 2011 (ed. por P. Mairano): 62 pontos de inquérito e 108 falantes
3. Dados

O corpus submetido a esta análise é constituído por um total de 28 frases declarativas e 28 interrogativas, com as mesmas estruturas sintáticas e obedecendo a restrições de tipo fonético e sintático. Esta recolha foi efetuada com a colaboração de 36 informantes (18 homens e 18 mulheres), provenientes de diferentes regiões. Da totalidade do corpus gravado, foram selecionadas, para cada informante, 3 repetições de cada uma das estruturas e modalidades, o que perfaz um total de 6048 enunciados analisados para o estudo que aqui apresentamos.

4. Metodologia

4.1 Medidas de correlação

Com o objetivo de serem estabelecidas semelhanças entre os dados de duas variedades, comparam-se as sequências de valores de frequência fundamental (f0), duração (D) e energia (I) com base numa variável determinada em Romano (1999, 2008):

\[ \rho_{x,y} = \frac{\text{Cov}(X,Y)}{\sigma_x \cdot \sigma_y} \]  

(1)

donde:

\[ -1 \leq \rho_{x,y} \leq 1 \]  

(2)

\[ \text{Cov}(X,Y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_x)(y_i - \mu_y) \]

X e Y representam sequências de n valores de f0, embora pudessem igualmente referir-se a séries de dados relativos à energia e à duração.

O resultado assim obtido precisa de ser validado por referência a um patamar previamente definido, estabelecendo-se o intervalo de oscilação da variável, quando se trata de repetições da mesma frase, na mesma variedade e produzida pelo mesmo falante (v. abaixo).

Apresentamos em seguida um exemplo da utilização do índice de correlação de Romano & Miotti (2008).

Na Figura 2 podemos observar os valores de correlação especialmente elevados (0,76÷0,84), quando comparamos enunciados declarativos, representados nos gráficos da esquerda e caraterizados por curvas bastante similares. Os valores positivos de correlação apresentam-se, somente em dois casos, para os locutores escolhidos para este caso específico (gráficos à direita, de baixo para cima): uma fraça correlação (0,26) diz respeito às duas questões nas variáveis 0905 e 0820, que são globalmente dastante similares, mas com diferenças localizadas bem visíveis, as quais são colocadas em evidência com setas a tracejado, no início dos enunciados. O índice baixa para 0,04, na comparação entre 0276 e 0905: as diferenças localizam-se, desta vez também, na parte terminal dos contornos, depois da realização do perfil correspondente ao acento de frase, com evoluções completamente inversas. Na comparação entre 0276 e 0820, as diferenças estendem-se sobre outras porções da curva (afetando também as vogais acentuadas) e a correlação torna-se negativa (−0,23).

![Figura 2: Comparações entre curvas (Romano & Miotti, 2008). Frases interrogativas de variedades de três espaços diferentes (0820, 0276, 0905). Medida de correlação em baixo à esquerda de cada gráfico](image)

No entanto, este index não tem em conta a importância percetiva de certas modificações da curva, em correspondência com as posições acentuais ou de fronteira prosódica. Como já foi referido, mostra-se muito sensível às variações individuais e precisa de uma avaliação prévia sobre as repetições de um mesmo locutor e as realizações de locutores de uma mesma localidade, definindo, respetivamente uma medida de “coerência” e de “congruência” (v. infra e §4).

4.2 Medidas de semelhança

Para evitar estas avaliações prévias, para melhor poder ter em consideração a energia e também para limitar a influência de valores absolutos de f0, uma nova medida foi proposta em Moutinho et al. (2011). Esta medida constitui uma avaliação objetiva da semelhança percetiva entre duas curvas entoacionais comparáveis. Permite também ignorar as diferenças de registo que poderiam existir entre dois locutores, para se concentrar na proximidade morfológica dos contornos. Esta medida mostrou-se pertinente na avaliação da proximidade percetiva de dois contornos prosódicos e, nesse aspeto, pareceu-nos adaptada a este tipo de situação. A referida medida baseia-se na seguinte fórmula:

\[ \text{Moutinho} = \sum_{i=1}^{n} |x_i - y_i| \]

![Equação 4.1: Medida de semelhança de Moutinho](image)

3 Estas restrições foram estabelecidas, desde o início, para o projeto AMPER (v. DVD AMPER 2011).

4 Esta medida de correlação é obtida graças aos escriptes Matlab™. Outros métodos de avaliação de distâncias são estudados por A. Rilliard (cf. Rilliard et al., 2011).
Expressos em semi-r do mesmo valor de energia-stante diferentes e 6

\[ R_{f_1,f_2} = \frac{\sum_i w(i)(f_1(i) - m_1)(f_2(i) - m_2)}{\sqrt{\sum_i w(i)(f_1(i) - m_1)^2 \sum_i w(i)(f_2(i) - m_2)^2}} \]

Aquí, \( f_1 \) e \( f_2 \) representam os valores de \( f_0 \) dos dois contornos entoacionais (expressos em semi-tons); \( m_1 \) e \( m_2 \) os valores médios destes contornos de \( f_0 \) para a totalidade da frase, e \( w \) a ponderação devida à energia do sinal, calculada como a média dos dois valores de energia medidos num ponto dado para as duas frases comparadas, expressos em dB.

O índice \( i \) varia entre 1 e o número de pontos de medida de \( f_0 \) para a frase considerada. Os valores de \( f_0 \) e de energia extraídos segundo o protocolo AMPER são utilizados para a seguinte medida: 3 pontos de \( f_0 \) por vogal, ponderados a partir do mesmo valor de energia média da vogal.

A distribuição das medidas de correlação não seguem uma lei normal, consideramos o valor da mediana como indicadora central, preferindo aquele ao valor da sua média.

5. Resultados

Comparando os valores da semelhança entre repetições para um mesmo falante e para falantes dum mesmo dialeto, obtivemos medidas de coerência e congruência.

Nos gráficos da Figura 3, propomos a avaliação da coerência de seis locutores de AMPER-ITA (pontos 061, 062, 06g e 06h, v. DVD AMPER 2011).

Os dados dos quatro primeiros locutores do primeiro diagrama acima representado (0616, 0621, 0625 e 06g5)\(^6\) evidenciam uma boa coerência (>90%), enquanto que para o locutor 06g6 a dispersão de valores assinala a presença de repetições com curvas bastante diferentes e para o locutor 06h7 uma coerência reduzida (mesmo assim >80% em média).

No segundo diagrama, avalia-se a congruência entre os dados de 0621 e 0625 (da localidade 062) e entre os dados de 06g5 e 06g6 (da localidade 06g): Os dados relativos ao ponto 062 estão associados a valores de congruência bastante elevados (à volta de 94%) e com uma dispersão bem concentrada (superior a 90%), enquanto que os dados de 06g apresentam uma congruência média inferior a 85% (no entanto ainda bastante elevada) e oscilações que poderiam ser consideradas localmente mais importantes\(^7\).

Se, pelo contrário, compararmos, os valores desta medida para falantes de dialetos próximos, obteremos uma estimativa da (dis)semelhança entre as amostras.

Figura 3: Medidas de coerência intra-falante e de congruência inter-falantes (dados AMPER-ITA: Romano, 1999 – 061-062; Felloni, 2011 – 06g-06h)

Figura 4: Dendrograma com o agrupamento de clustering hierarquizado e mapa dialectométrico da distância prosódica média dos dados das diferentes regiões em relação aos dados do ponto 016 (Trinta, Beira Alta) [Moutinho et al. (2011)]. A intensidade do cinzento é uma função linear da distância entre os pontos. Códigos: 006 = Alfândega da Fé (Trás-os-Montes), 00q = Monte Gordo (Algarve), 001 = Prado (Braga, Minho), 001 = Monforte (Alto Alentejo), 012 = Aradas (Beira Litoral)

Uma quantificação dos resultados obtidos encontra-se detalhada em Moutinho et al. (2011), com a apresentação de vários casos de congruência reduzida para certas localidades (20-60%), melhor para outras, como é o caso do ponto 016 (Trinta, Beira Alta) e que escolhemos como localidade de referência para uma primeira proposta de avaliação geoprosódica destes dados. Para dar conta das relações existentes entre os dados de outras localidades com os obtidos para estas (somente 25% de semelhança entre 016 e 012 e 75% entre 016 e


\(^6\) O último algarismo acrescentado ao código da localidade designa o código do locutor.

\(^7\) Isso não significa necessariamente que o inquérito realizado no ponto 062 reproduza uma estimativa da prosódia típica desta localidade melhor do que a obtida para o ponto 06g: uma congruência menos boa poderia ser o sintoma de uma prosódia mais variável localmente entre variedades diastáticas e/ou os idióteos de género.
061), foi adotado o método de análise dialectométrica (cf. Goebel, 1981, 1996) baseada numa avaliação cruzada da distância prosódica média dos dados das diferentes regiões em relação aos dados deste ponto, tendo sido proposto um agrupamento hierarquizado (v. Figura 4)\(^8\).

6. Conclusões
As medidas que nós aplicámos aos dados das variedades presentes na BD AMPER, mesmo que não nos permitam substituir a análise do dialectólogo foneticista tradicional, permitem, sem dúvida, colocar em evidência algumas divergências e convergências prosódicas de diferentes falares. Estas constatações proporcionam indicações sobre a distância percetiva que podemos esperar encontrar entre dialetos.

Nas nossas pesquisas, depois de termos discutido as possibilidades e modalidades de aplicação da distância propostas em trabalhos precedentes, fizemos a avaliação da variabilidade prosódica em dados referentes a uma primeira seleção de variedades. Deste modo, este estudo deve ser lido como um esboço de dialectometria prosódica.

É indispensável que estes resultados sejam confirmados através de análises que incidam sobre um número mais vasto de falantes de cada uma das variedades e sobre a base de um conjunto de pontos de inquirito mais densa e completa, para ambas as línguas, mas especialmente no que diz respeito ao AMPER-ITA.

7. Agradecimentos
Este trabalho foi possível graças ao financiamento dos programas H. Curien (Edige 697912/QASSID) / F. Pessoa (Dossier FCT, Proj. 6818) entre a Universidade de Aveiro e a Universidade de Grenoble. Agradecemos a Maria Chiara Felloni por nos ter disponibilizado, para o italiano, dados dos seus inquiritos.

8. Referências

AMPER (2011). Intonations Romance, Géolinguistique, hors série 4 (P. Mairano, ed.).


F2 transition as a cue to place of articulation in Brazilian Portuguese coronal fricatives

Rui ROTHE-NEVES, Fabiana Andrade PENIDO
Laboratório de Fonética da FALE/UFMG
Av. AntonioCarlos, 6627 – 31270-217, Belo Horizonte/MG - Brazil
rothe-neves@ufmg.br, fabianapenido@gmail.com

Abstract

In Portuguese, frication noise has been tested for consonant recognition, but to the best of our knowledge formant frequencies were not investigated yet. We tested whether the second formant (F2) transition said to be useful in English are also a cue to place of articulation for coronal fricatives in Brazilian Portuguese. Subject performed a rating task in which they had to listen to a syllable and quickly respond whether they heard [s(a)] or [ʃ(ə)a] (1st block) and [s(u)] or [ʃ(ə)u] (2nd block), and rate their confidence in their answers using a 3-pt scale. Hit and false alarm rates of all response alternatives to [(s)a]-[(ʃ)əa] and [(s)u]-[(ʃ)əu] were computed. Slope and A3 were estimated by maximum likelihood estimation of ROC. For [a] there was good evidence in our data that F2 transition is an important and sufficient cue to place of articulation in coronal fricatives. Also, F2 transition described for English, once adapted to the formant frequency values reported for BP, are useful to distinguish between [sa]-[ʃa]. However, listeners could not distinguish between [(s)u]-[(ʃ)əu] only on the basis of F2 transition. This result points to a possible role of F3 transition, which was said to become an important cue for rounded vowels.

Keywords: speech perception; phonetics; phonology; Brazilian Portuguese.

1. Introduction

Two kinds of cues are shown to be used in the distinction between coronal voiceless fricatives: the spectral shape differences in the frication noise (4-8kHz for [s], 2-4kHz for [ʃ]) and the spectral changes in formant transitions between the noise and the adjacent vowel (Harris, 1958; Heinz & Stevens, 1961; Hughes & Halle, 1956; Dorman, Raphael & Isenberg, 1980 for English; Guerlekian, 1981 for Spanish). In European Portuguese, frication noise has been described to have center frequencies around 5kHz for [s] and 3kHz for [ʃ] in the European variety (Lacerda, 1982; Jesus, 1999), and in Brazilian Portuguese (BP), Haupt (2008) and Santos (1987) described values around those of English: for [s] 4.5-7.4kHz, for [ʃ] 2-4.6kHz. Lacerda (1982) tested those frequencies for consonant recognition, but to the best of our knowledge formant frequencies were not investigated yet. We decided to test whether second formant (F2) transition said to be useful in English is also a cue to place of articulation for coronal fricatives in Brazilian Portuguese.

2. Methods

2.1 Subjects

Twelve female subjects with age varying from 14-28 years with no history of hearing problems participated in the study.

2.2 Stimuli

Four vowels were synthesized with HLSyn (Sensimetrics, Inc.), two tokens for [a] and two for [u]. One token of either vowel was manipulated in F2 transition in the first 50ms: one compatible with a transition from [s], the other from [ʃ] (Table 1). Formant values for the steady-state part were those presented by Escudero et al. (2009), and the F2 transition values, from Nitttrouer and Miller (1997).

<table>
<thead>
<tr>
<th></th>
<th>[(s)a]</th>
<th>[(ʃ)əa]</th>
<th>[(s)u]</th>
<th>[(ʃ)əu]</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 Initial</td>
<td>483</td>
<td>483</td>
<td>310</td>
<td>310</td>
</tr>
<tr>
<td>F1 Final</td>
<td>683</td>
<td>683</td>
<td>310</td>
<td>310</td>
</tr>
<tr>
<td>F2 Initial</td>
<td>1449</td>
<td>1769</td>
<td>1391</td>
<td>1711</td>
</tr>
<tr>
<td>F2 Final</td>
<td>1329</td>
<td>1329</td>
<td>761</td>
<td>761</td>
</tr>
<tr>
<td>F3 Initial</td>
<td>2624</td>
<td>2624</td>
<td>2609</td>
<td>2609</td>
</tr>
<tr>
<td>F3 Final</td>
<td>2324</td>
<td>2324</td>
<td>2309</td>
<td>2309</td>
</tr>
</tbody>
</table>

Table 1: Initial and Final Transition Formant values (Hz)

A 160ms raw frication noise with no filtering was synthesized using the Klatt cascade model implemented in Praat (Boersma & Weenink, 2011). Noise was subsequently single-pole filtered in different formant frequencies with a bandwidth of 230Hz. Noise filter frequencies were taken from a normally distributed, randomly generated 100 number sequence with mean = 4830 and sd = 50. With this procedure, each frication noise was not exactly the same, so that subjects would not get tired responding many times to one and the same stimulus, but at the same time the main effect would normally cluster around halfway between the center frequency of a [s] and that of a [ʃ]. The 100 different frication noises were then concatenated with the four vowel tokens (150ms) to produce 400 synthetic syllables of 310ms duration.

2.3 Procedure

The 200 tokens for [(s)a] and [(ʃ)əa] were presented in a block, the 200 tokens for [(s)u] and [(ʃ)əu] in another block within the same session. Subjects were allowed to have a break between blocks. A rating task was used (Macmillan & Creelman, 2005), in which subjects were required to listen to a syllable and respond whether they heard [(s)a]...
or [ʃ] in the first block, and [(s)u] or [(ʃ)u] in the second block, and to rate their confidence in their answers using a 3-point scale.

Data was collected in a quiet room. PercEval (LPL, CNRS/Université de Provence) in the BP version was used for sound presentation with a circumaural headphone and response entry.

3. Results

Hit and false alarm rates of all response alternatives to [sa]-[ʃa] and [su]-[ʃu] were computed. Instead of δ', a more common measure of sensitivity in classification task, we used the area under the ROC curve (Aδ) produced by the cumulative δ’ for each response level in the 3-point scale. Slope and Aδ were estimated with ROC-kit (Metz, Herman & Shen, 1998). So, the standard assumption of unit slope in δ’ is unnecessary. According to the results in the “slope” column in Tabs. 2 and 3, this would have been here a very strong assumption.

As Aδ values range from .5 (no sensitivity, or confusion) to 1 (complete sensitivity), for [a] there is good evidence in our data that F2 transition is an important and sufficient cue to place of articulation in coronal fricatives (Table 2). Also, F2 transitions described for English, once adapted to the formant frequency values reported for BP, are useful for listeners to distinguish [sa] from [ʃa].

As Aδ values range from .5 (no sensitivity, or confusion) to 1 (complete sensitivity), for [a] there is good evidence in our data that F2 transition is an important and sufficient cue to place of articulation in coronal fricatives (Table 2). Also, F2 transitions described for English, once adapted to the formant frequency values reported for BP, are useful for listeners to distinguish [sa] from [ʃa].

Table 2: Results for the [ʃa]-[sa] distinction

<table>
<thead>
<tr>
<th>Subject</th>
<th>n(ʃa)</th>
<th>Slope</th>
<th>Aδ</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1st run</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATR</td>
<td>100/99</td>
<td>2.227</td>
<td>0.976</td>
<td>0.011</td>
</tr>
<tr>
<td>MTR</td>
<td>100/99</td>
<td>2.888</td>
<td>0.811</td>
<td>0.035</td>
</tr>
<tr>
<td>PGF</td>
<td>100/99</td>
<td>0.679</td>
<td>0.965</td>
<td>0.013</td>
</tr>
<tr>
<td>RPA</td>
<td>98/99</td>
<td>0.299</td>
<td>0.839</td>
<td>0.043</td>
</tr>
<tr>
<td><strong>2nd run</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACA</td>
<td>100/99</td>
<td>0.274</td>
<td>0.977</td>
<td>0.016</td>
</tr>
<tr>
<td>CM</td>
<td>90/62</td>
<td>0.685</td>
<td>0.867</td>
<td>0.034</td>
</tr>
<tr>
<td>DOS</td>
<td>100/100</td>
<td>0.869</td>
<td>0.967</td>
<td>0.011</td>
</tr>
<tr>
<td>DSF</td>
<td>100/100</td>
<td>2.564</td>
<td>0.984</td>
<td>0.012</td>
</tr>
<tr>
<td>FAP</td>
<td>100/100</td>
<td>2.406</td>
<td>0.964</td>
<td>0.018</td>
</tr>
<tr>
<td>LM</td>
<td>100/100</td>
<td>1.131</td>
<td>0.644</td>
<td>0.041</td>
</tr>
<tr>
<td>SHA</td>
<td>100/99</td>
<td>1.46</td>
<td>0.95</td>
<td>0.016</td>
</tr>
<tr>
<td>TPA</td>
<td>100/100</td>
<td>1.04</td>
<td>0.846</td>
<td>0.029</td>
</tr>
</tbody>
</table>

Table 3: Results for the [ʃu]-[su] distinction

<table>
<thead>
<tr>
<th>Subject</th>
<th>n(ʃu)</th>
<th>Slope</th>
<th>Aδ</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1st run</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATR</td>
<td>100/100</td>
<td>1.392</td>
<td>0.58</td>
<td>0.041</td>
</tr>
<tr>
<td>MTR</td>
<td>100/98</td>
<td>2.512</td>
<td>0.422</td>
<td>0.043</td>
</tr>
<tr>
<td>PGF</td>
<td>100/99</td>
<td>0.92</td>
<td>0.387</td>
<td>0.043</td>
</tr>
<tr>
<td>RPA</td>
<td>99/99</td>
<td>1.341</td>
<td>0.414</td>
<td>0.053</td>
</tr>
<tr>
<td><strong>2nd run</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PGF</td>
<td>98/97</td>
<td>0.909</td>
<td>0.662</td>
<td>0.041</td>
</tr>
<tr>
<td>ACA</td>
<td>100/100</td>
<td>0.744</td>
<td>0.665</td>
<td>0.046</td>
</tr>
<tr>
<td>CM</td>
<td>69/66</td>
<td>1.11</td>
<td>0.529</td>
<td>0.054</td>
</tr>
<tr>
<td>DOS</td>
<td>100/100</td>
<td>0.72</td>
<td>0.559</td>
<td>0.044</td>
</tr>
<tr>
<td>DSF</td>
<td>100/100</td>
<td>1.168</td>
<td>0.544</td>
<td>0.044</td>
</tr>
<tr>
<td>FAP</td>
<td>99/100</td>
<td>1.04</td>
<td>0.943</td>
<td>0.017</td>
</tr>
<tr>
<td>LM</td>
<td>100/100</td>
<td>1.151</td>
<td>0.589</td>
<td>0.041</td>
</tr>
<tr>
<td>SHA</td>
<td>100/97</td>
<td>1.614</td>
<td>0.555</td>
<td>0.048</td>
</tr>
<tr>
<td>TPA</td>
<td>99/100</td>
<td>1.471</td>
<td>0.571</td>
<td>0.044</td>
</tr>
</tbody>
</table>

Table 3: Results for the [ʃu]-[su] distinction

4. Conclusion

For [a] there is good evidence in our data that F2 transition is an important and sufficient cue to place of articulation in coronal fricatives. However, listeners could not distinguish between [(s)u]-[(ʃ)u] only on the basis of F2 transition. This result points to a possible role of F3 transition, which were said to become important in rounded vowels. It will be a matter of future studies.

5. Acknowledgements

This research was funded by CNPq/PQ nº 311484/2009-3 and by the Fundo de Apoio à Pesquisa, FALE-UFMG.

6. References


Transplanting credibility into a foreign voice: an experiment on synthesized L2 Italian

Massimo PETTORINO, Anna DE MEO, Marilisa VITALE
University of Naples “L’Orientale”
Naples, Italy
mpettorino@unior.it, ademeo@unior.it, vitalem@unior.it

Abstract

This study intends to verify through perceptual tests conducted on original and artificially modified speech whether a relationship exists among the degree of comprehensibility of an utterance, the foreign accent and the credibility of the message. Four bizarre-but-true news read in Italian by four non-native speakers were artificially modified with Praat and WaveSurfer. Each piece of news was transplanted, so that segmental and prosodic features of a text read by a native speaker were transferred onto the same text uttered by a non-native speaker. The corpus was administered to 265 native Italian listeners, who were requested to indicate the degree of comprehensibility, the level of foreign accent and the truthfulness of each item. The results point out the existence of a close inverse relationship between comprehensibility and credibility. The presence of foreign accent, providing an impediment to the understanding of the message, tends to create an attitude of distrust in the listener. The most important features for the foreign accent reduction are the suprasegmental ones and, in particular, the durations of the phones and the pitch movement.

Keywords: foreign accent; comprehensibility; credibility; L2 Italian; prosody.

1. Introduction

Our recent study on socio-cultural effects of foreign accent on communication effectiveness (De Meo et al., 2012) revealed the relevance of comprehensibility factors - such as disfluency, frequency of silences, pitch range variation, silent pauses, segmental errors - on message credibility. A hundred seventy-five native Italian listeners, after hearing a set of 10 news uttered in Italian by one native speaker of Italian and four non-native speakers of L1 Chinese, Vietnamese, Arabic and Japanese, were asked to assess the comprehensibility, i. e. listener’s estimation of difficulty in understanding an utterance (Munro & Derwing, 1999), and the truthfulness of each news item. The four non-native speakers, all late bilinguals with a basic (A2) and a mid (B1) level of competence as laid out in the Common European Framework of Reference, and an average stay time in Italy of 6 months, were chosen for the study after a global foreign accentedness rating test which was administered to 70 male and female native Italian listeners. Listeners rated the degree of foreign accentedness of a short read text on a 4-point scale (0 = native speaker; 3 = strong foreign accent). The results allowed to select four L2 speakers of Italian with a strong foreign accent.

Ten bizarre-but-true news from around the world read by the native speaker and the four non native speakers were presented to native listeners in form of radio news magazines, each combining the four voices reading different news, same news sequences but random voice order, pretending to administer a survey on media reliability, in order to avoid to focus the attention on foreign voices.

Obviously each single piece of news revealed to have its own degree of credibility, in accordance with the textual content of the message. However results showed that, within the same text, ratings were significantly different depending on the auditory comprehensibility level.

The study showed that when there are no comprehensibility problems the assessment of real/false is maintained around 50%, so in a range of randomness. On the contrary, when the level of comprehensibility lowers, due to various acoustic factors (disfluencies, errors, percentage of silence, tonal variation, etc.), the judgments of “false” increase rapidly, reaching 90% when the statement proves to be poorly understandable for the 40% of listeners. Therefore, there seems to be a threshold of comprehension tolerance, i.e. a level of difficulty in understanding an utterance at which the listener’s effort to understand the message leads him to believe that what he has just heard is not credible.

Following this line of research, our current study intends to carry out a perceptual test on artificially modified speech, in order to evaluate the role played by both segmental and suprasegmental features in the achievement of an L2 comprehensible communication and find out if there is a relationship between the perceived degree of foreign accent and credibility.

2. Methods and materials

The corpus used for this study, taken from the one used in De Meo et al. (2012), consists of 4 news artificially modified with Praat (Boersma & Weenink, 2012). Each single piece of news was manipulated, so that disfluencies and errors were removed, and the prosodic features of the native speaker’s utterances were transferred onto the same utterances produced by the non-native speakers (prosodic transplantation technique).

2.1 Corpus and Informants

1) Informants: 5 female voices
   • 1 Italian speaker (L1)
   • 4 L2 Italian speakers (Chinese, Vietnamese, Japanese, Arabic L1s)

2) Corpus: 18 audio files (bizarre-but-true news)
   • 8 original news (4 L1, 4 L2)
• 10 L2 artificially modified (4 with removed disfluencies and cloned pauses, 2 with removed errors, 4 with cloned durations and pitch contour).

2.2 The transplantation technique
The rhythmic-prosodic transplantation technique is based on the algorithm PSOLA (pith-synchronous overlap-add, Moulins & Charpentier, 1990), implemented in Praat and illustrated in Yoon (2007) with regard to the English productions of Korean speakers. The prosodic features that can be transplanted from one voice to another are essentially four: the length of the segments, the pitch contour of the utterance, the intensity contour and the silent pauses.

The procedure of transplantation must follow a well-defined sequence of steps, since each of them is preparatory for the subsequent ones. The five phases are: anomalies treatment (disfluencies removal, pause cloning, errors elimination), segmentation and labelling, transplantation of the duration, intensity transplantation, pitch contour superimposition.

This technique seems to be a rather effective tool for the study of the spoken L2, since the manipulation of an utterance allows to evaluate the role played by individual acoustic parameters at the pragmatic-communicative level.

2.3 Perceptive test
The whole corpus was administered in a randomized order to 265 native listeners (male and female, mean age 21, university students) organized into 5 groups, so that nobody could listen to the same news more than once. As the purpose of the survey was to assess the credibility, the repeated exposure to a same input would have affected the reliability of the test results.

For each utterance, listeners were asked to evaluate the comprehensibility (poor, sufficient, good), assess the degree of perceived foreign accent (native accent, mild foreign accent, strong foreign accent) and judge on its truthfulness (true/false).

3. Results and discussion
In this section we will examine the results of the abovementioned test, in order to evaluate the relevance of each manipulated factor on the perceptual level. The discussion will be organized into three parts corresponding to the different steps of the synthesis procedure. For the data analysis the One-Way ANOVA was performed.

3.1 First step: Removing disfluencies and cloning native silences
Figures 1, 2, 3 show the average percentage values of the judgements given to the utterances, both original and modified, produced by the native (NS) and the non-native speakers (NNS), with respect to comprehensibility, degree of foreign accent and credibility.

As for the foreign accent, both the NS and the NNSs were correctly recognized by almost all the listeners (Figure 2). The modifications carried out on NNSs’ utterances produced a decrease of about 20% of the judgments of “strong foreign accent” (from 79% to 60%). In addition, it is worth noting that the 5% of the listeners assumed to have heard a native voice. Data are statistically significant (p<0.001).

The removal of disfluencies and the repositioning of the silences determine a statistically significant improvement (p<0.001) of the NNSs’ utterances comprehensibility (Figure 1). As a result of the manipulation, the majority of the listeners (78%) judged the non-native productions at least sufficiently comprehensible. Obviously, the NS proved to be highly comprehensible.
Removal of disfluencies and changes of silences (Figure 3) determined a significant increase (p<0.005) in the level of news credibility (+26%), taking the NNSs’ values to levels very similar to those obtained by the native speaker.

3.2 Second step: Errors removal

For the second step of the study data are limited to the A2 level NNSs, since in the other speakers’ productions there were no particular segmental irregularities. Using WaveSurfer, phones perceptually detected as wrong by three native trained phoneticians were artificially modified or substituted through a self-transplantation procedure, i.e. using adequate micro-segments produced by the speaker within the same utterance.

Because of the large variability and unpredictability of the errors, this phase is technically the most problematic. The typology and frequency of errors require operations that may damage the quality of the synthesized audio file and interfere with the perceptive evaluation.

Our data show that the segmental modifications give anyhow rise to a slight but significant improvement in terms of foreign accent assessment (p<0.005; “strong” from 77% to 66% and “mild” from 20% to 31%). No significant variations were observed for the comprehensibility and the credibility (p>0.05).

3.3 Third step: Duration and pitch transplantation

The final step of the transplantation procedure involved the cloning of the duration of each segment and, subsequently, the superimposition of the intonation contour from the NS’s utterances to the NNSs’ ones. The perceptive test outcomes are generally satisfactory. Figure 4, concerning comprehensibility, shows that, if compared to the first step of the procedure (disfluencies removal and silences cloning), the negative judgments decreased by 10% (from 23% to 12%) in favour of the “sufficient” ratings, while the “good comprehensibility” values did not change (p>0.05). It should be noted that the results of the overall transplantation process, when compared to the original utterances, reveal a remarkable improvement: the “poor comprehensibility” lowers by 42% and the “good comprehensibility” rises by 16% (p<0.001).

The most evident effects of this last step are those related to the degree of foreign accent (Figure 5), with a gain of about 30% for the judgment of “native” and a 60% reduction with regards to the judgment of “strong foreign accent” (p<0.001).

Finally, the values of credibility (Figure 6) do not undergo further significant variations (p>0.05).

4. Conclusion

In conclusion, the study confirms the existence of a close relationship between comprehensibility and credibility, both for original and manipulated audio files (p<0.001). The more the utterance is easy to understand the more the listener is led to believe true what he/she has just heard. In this perspective, for our informants, beginners (A2) and low-intermediate (B1) speakers of L2 Italian, whose speech is characterized by disfluencies, anomalous silences, segmental errors, and inappropriate pitch contour, foreign accent provides an impediment to the understanding of the message and, consequently, tends to create an attitude of distrust in the listeners. However, it is worth to emphasize that it is not the “foreignness” as such to cause a lowering of credibility, but it is rather the difficulty of decoding the message determined by the presence of anomalies typical of an early L2 speech.
5. References


