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

Maryualê M. MITTMANN (UFMG)
Bruno ROCHA (UFMG)

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), according to which, the linguistic behavior is accomplished through speech acts. A speech act is understood as the simultaneous performance of three acts: locutionary, illocutionary and perlocutionary (Austin, 1962). 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).

Prosody works as an interface between the locutionary and illocutionary acts, since 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. 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). A utterance is considered simple when is composed by a single prosodic unit. If the utterance is parsed into two or more prosodic units is, then it is considered complex. The possible internal units are associated with informational functions, through which information is patterned within the utterance. 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).

Among the textual information functions, the most important is the "topic", since it is the most frequent textual unit (about 50% of textual units in a BP sample). 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).

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.

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 Language into Act Theory (Moneglia; Cresti, 1997). The utterances containing at least one topic unit were extracted from the sample and we proceeded the analysis through Praat software (Boersma; Weeninck, 2011) through the following steps: (1) Extraction of F0 values in Hertz; intensity and duration (syllabic and vocalic). (2) Stylization of F0 contour and manipulation of prosodic parameters through speech

resynthesis for the identification of relevant perceptual movements. (3) Identification of intonational form types.

PB and PE present the three intonational forms described for Italian (Firenzuoli, Signorini, 2003). A fourth intonational form (type 4) was found in both Portuguese varieties. 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. 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. 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. Type 4 topic 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 final tonic syllable presents a lengthening and an increase of intensity. The F0 contour on the final portion seems not to be of relevance, since it can be either flat, or slightly falling or rising. The relevant parameters for this prosodic form of topic seem to be the F0 peak on the first tonic syllable and the increase in the intensity and duration of the last tonic syllable. The high pitch of the first tonic syllable may spread on the post-tonic syllable.

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. Type 3 is the least common form in Portuguese 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 F0 values of the first tonic syllable onset, with the Brazilian variety presenting higher values than the European.

References

AUSTIN, J. L. How to do things with words. Oxford: Oxford University Press, 1962.

BOERSMA, P.; WEENINK, D. Praat: doing phonetics by computer. Retrieved from <http://www.praat.org/>, 2011.

CRESTI, E. Corpus di Italiano parlato. v. 1. Firenze: Accademia della Crusca, 2000.

CRESTI, E.; MONEGLIA, M. Informational patterning theory and the corpus-based description of spoken language: The compositionality issue in the topic-comment pattern. In: A. Panunzi (Ed.); Bootstrapping Information from Corpora in a Cross-Linguistic Perspective. Firenze: FUP, 2010.

FIRENZUOLI, V.; SIGNORINI, S. L'unitá informativa di topic: correlati intonativi. in Atti delle Giornate del Gruppo di Fonetica Sperimentale - XIII, Pisa, Novembre 2002 ETS, Pisa. Anais... p.177-184. Retrieved September 6, 2007, from <http://lablita.dit.unifi.it/preprint/preprint-02coll11.pdf>, 2003.

MONEGLIA, M.; CRESTI, E. Formazione dell'atto linguisitico complesso e intonazione: l'ontogenesi delle relazioni informative in italiano. In: E. Cresti; M. Moneglia (Eds.); Ricerche Sull'acquisizione Dell'italiano. p.63-114. Roma: Bulzoni editore, 1993.

SIGNORINI, S. Topic e soggetto in corpora di italiano parlato spontaneo. Phd. Thesis, Dottorato di Ricerca in linguistica Italiana, Università degli Studi di Firenze, Departamento di Linguistica, Firenze, 2005.