

## **Intonational Prominence and Information Status in Brazilian Portuguese**

Luciana LUCENTE (UNICAMP)  
Julia HIRSCHBERG (Columbia University)  
Plinio BARBOSA (UNICAMP)

The relation between intonational focus and information status, especially to given and new information (Terken & Hirschberg, 1994; Gravano & Hirschberg, 2006) has been accepted for some time. According to definitions of Chafe (1974) and Prince (1981) the speaker assumes that given information is what is already in the hearer consciousness at the time of the utterance, and new information is what the speaker assumes s/he is introducing into the hearer consciousness by what s/he says.

The present work aims at exploring the relationship between given and new information in Brazilian Portuguese (henceforth BP) using the VoCE spontaneous speech corpus. The VoCE corpus is a corpus essentially composed of spontaneous speech data from radio broadcasts and podcasts from the internet. This corpus has been being collected since 2006 and contains approximately 10 hours of conversation among announcers and guests.

The information status research is being developed in the framework of the DaTo (Dynamic Tones of Brazilian Portuguese Intonation) intonation annotation system (Lucente & Barbosa, 2008; 2010). The DaTo system attempts not only to present a system for intonational annotation, but also to describe functional aspects of intonational contours in spontaneous speech, e. g., their pragmatic function. The labels set for intonational contours annotation is: LH (rising), >LH (late rising), vLH (compressed rising), HLH (falling-rising), HL (falling), >HL (late falling), vHL (compressed falling) and LHL (rising-falling). The static labels L (low) and H (high) have the boundary function, used to annotate final boundaries. The DaTo system presupposes an annotation based on: i) automatic detection of (filled and silent) pauses (stress group boundaries); ii) speech segmentation in V-V (vowel-to-vowel) units, iii) orthographic transcription; and iv) pragmatic information.

As research about information status and intonation in the DaTo framework follows a pragmatic/cognitive approach, we attempt to analyze whether given/new information status is related to i) a specific DaTo intonational contour; ii) intonational prominence, and iii) grammatical function. The choice of these three parameters is based on the hypothesis that speakers use them to differentiate given information from new information.

The first step before data analysis was the classification of nouns as given (G) or new (N) following Prince's (1981) taxonomy. Such classification assumes a cognitive basis, for which the information status is defined according to the information achievement in both speaker and hearer minds (Chafe, 1974). The next step was to relate the information status classification with the DaTo intonational contours (LH, >LH, vLH, HL, LHL and HLH) associated with each given/new token. The given/new classification was also related to the grammatical function (subject/object) and surface position (initial, medial, final) in the discourse segments (DS), according to Grosz and Sidner (1986) theory of discourse structure. The annotation referring to intonational contours, information status and DSP (Idem) segmentation was done in separate tiers using Praat software package. A classification of the material in nouns preceded or followed by adverbs, and the presence or absence of the referred expressions (Gundel et al, 1993) was also carried out.

Our preliminary results have shown significant correlations between information status and grammatical function, prominence and intonational contours. After a Machine Learning classification using Weka's C4.5 we found that: i) given information tends to occur in subject

position and new information in object position; and ii) the relationship between information status and prominence degree is ambiguous; the intonational prominence in given and new information can vary according to the token position in the utterance or the presence of an adverb or a referring expression preceding the token; iii) the tendency of LH/>LH and LHL/HL contours occurring with on given and new information, respectively; this occurrence is also related with information position in the discourse structure.