

## Automatic extraction of phonological frequency and applications to language acquisition

Marina Vigário & Sónia Frota (Universidade de Lisboa, Onset-CEL)

Frequency has been shown to play an important role in many areas of linguistics. Our goal in this presentation is twofold: (i) to describe a new electronic tool for extracting frequency information of phonological units at the word-level and below, from Portuguese written text - *FreP (Frequency in Portuguese)*; (ii) and to report on the application of this tool to acquisition studies.

Taking advantage of a highly predictable relation between orthography and (lexical) phonology in Portuguese, *FreP* allows the automatic extraction (identification and count) of the following phonological units: prosodic words and clitics, syllables, classes of segments (consonants, vowels, glides), consonantal segments (p, t, k, ..), and PoA features. Besides that, (i) it locates word stress, and provides information on the distribution of stress within words (i.e. number of words with final, penult and antepenult stress, and this also taking into consideration word size – e.g. number of disyllabic words with final and penult stress, of trisyllabic words with final, penult and antepenult stress...), (ii) it counts the frequency of different syllable types (CV, V, CVC...), and does so by position in the word (initial, internal and final), taking into account the presence/absence of word stress, or both (position in the word and presence/absence of word stress), (iii) it provides information on the size of words (number of words with one, two, three, N syllables), and does so for prosodic words as well as for clitics, (iv) within the class of phonological clitics, it sets enclitics and proclitics apart, supplying frequency information for both types of units separately, (v) it counts the number of PoA features, taking into account the position within the word (e.g. Labial in initial, internal or final position, Coronal in initial, internal or final position...), (vi) it provides a word frequency list (both token and type).

The tool has been already applied both to adult speech and child directed speech corpora, as well as to child speech corpora. A summary of the investigation conducted so far on several aspects of the acquisition and development of phonology will be given, which includes (i) the development of prosodic word shapes (Vigário, Freitas and Frota 2006), (ii) the development of syllable structure (Frota, Freitas, Vigário & Martins 2005, Freitas, Frota, Vigario & Martins 2006), (iii) the acquisition of Point of Articulation features (Costa, Freitas, Frota, Vigario & Martins 2006, 2007), and (iv) the development of the lexicon (Pós de Mina 2007, in progress). Possible extensions of *FreP* and its use in connection with acquisition databases will be discussed.

More information at:

<http://www.fl.ul.pt/LaboratorioFonetica/FreP/>