Asymmetries between prosodic words and clitics in the frequency of phonological patterns: implications for language acquisition

Marina Vigário, Sónia Frota & Fernando Martins

University of Lisbon

PaPI 2011 - Phonetics and Phonology in Iberia, Universitat Rovira i Virgili, Spain, June 21-22, 2011

1. Introduction

- It is well-known that prosodic words (PW) ≠ phonological clitics (CL)

Word-stress

Domain for phonological rules
Phonotactic restrictions
Pitch accent & focal accent

Typically belong to large, open classes
Often content words
May be morphologically complex
Never appear in isolation

Belong to small, closed classes
Only grammatical function
Morphologically simple

- Frequency-based asymmetries between lexical and function words: function words are short, highly frequent, display highly frequent syllable shapes at prosodic edges (e.g. Shi, Weker & Morgan 1999; Gervain, Nespor, Maizua, Horie & Mehler 2008; Christophe, Milloffe, Bernal & Lidz 2008)

- Phonological and acoustic cues – role in the bootstrapping of syntactic properties in language acquisition (Morgan & Denuth 1996; Nespor, Guasti & Christophe 1996; Shi et al. 1999; Christophe, Nespor, Guasti & van Ooyen 2003; Christophe et al. 2008, a.o.)

- Frequency-based bootstrapping mechanism (e.g. Gervain et al. 2008)

- Importance of multiple acoustic and phonological cues for distinguishing lexical and grammatical words (Shi et al. 1999, a.o.)

- PW/CL distinction is phonological/phonetic in nature

- Frequency-based asymmetries between lexical and function words: function words are short, highly frequent, display highly frequent syllable shapes at prosodic edges (e.g. Shi, Weker & Morgan 1999; Gervain, Nespor, Maizua, Horie & Mehler 2008; Christophe, Milloffe, Bernal & Lidz 2008)

- PW/CL distinction is phonological/phonetic in nature

Hyp.: 1. PW ≠ lexical words & CL ≠ function words
2. lexical words & function words
3. Quest for a frequency-based distinction between PW and CL in EP in multiple dimensions

2. Method

- Corpus of over half a million orth. words, spoken EP

- FreP (Martins, Vigário & Frota 2009, v. 2.1.09) automatically
  > separates PW from CL
  > provides frequency values for
    - Word-stress
      - Domain for phonological rules
      - Phonotactic restrictions
      - Pitch accent & focal accent
    - Types
      -PW/CL distinction is phonological/phonetic in nature
      - Typically belong to large, open classes
      - Often content words
      - May be morphologically complex
      - Never appear in isolation
    - CL: segmental inventory reduced in nearly 50%
    - PW: >30
      - larger the corpus the larger the diff.
    - CL_fin: of the three consonants allowed word-fin., only 2 appear CL-final; [w], [e] frequent
    - PW_fin: [w], [e] rare/impossible
    - 99.8% of [i]-initial syll. begin a word; in most cases the word is a CL (77.8%)
    - 95.2% of [u]-initial syll. begin a word; in most cases the word is a CL (91.5%)
    - 99.9% of [aw]-initial syll. begin a word; in most cases the word is a CL (87.8%)
    - 85% of [k]-initial syll. are CL-initial; 95.7% of [k]-initial syll. belong to CL
    - 75% of [d]-initial syll. are CL-initial
    - 100% of words starting in [6–j–] are CL
    - 100% of words ending in [a], [e], [–o–] are CL
    - above 96% of the words ending in [w], [i], [r] (stressless syll.) are CL
    - 100% of the words starting in [b, f, z, g, k, w, g, w, e, E, i, e, –o–] are PW
    - 100% of the words starting in [b,f, t, d, n, s, z, l, r, S, Z, J, L, k, g, R, l, G, k_w, g_w, e, E, O, i, e, –o–] are PW

3. Results

- Type/Token ratio
  - PW: 0.089
  - CL: 0.00036
- Token/Type ratio
  - PW: 11.2
  - CL: 2783.0
- Proportion PW/CL
  - PW: 30% polysyllabic
  - CL: maximally disyllabic
- Proportion of the most frequent word shapes inverted
  - PW: 44% disyll / 26% monos
  - CL: 7% disyll / 93% monos
- Diversity in word shapes
  - PW: more greater richness
  - CL: maximally disyllabic
- Segmental inventory of the language
  - PW: 30% polysyllabic
  - CL: segmental inventory reduced in nearly 50%

4. Discussion and conclusions

- Phonological frequency patterns in EP provide abundant cues for the separation of PW and CL
- There is evidence that much of what is at stake is perceived by infants at or before 9 months: infants are able to use probabilistic information, they show sensitivity to differences in word size and to differences between lexical and grammatical words in type/token ratio (e.g. Jusczyk, Luce & Charlier-Luce 1994; Biesemans-Babc, Bertocci & Mehler 1993; Saffran et al. 1996; Shi et al. 1999; Mattys & Jusczyk 2001; Saffran 2002, a.o.)
- Thus, it is proposed that the frequency of phonological patterns may play a role in the acquisition of the distinction between PWs and CLs in EP; and these in turn may function as precursors of the distinction between lexical and function words.
- Such a distinction may be crucial in the course of language acquisition to bootstrap the basic word order parameter (Gervain et al. 2008) and establish word classes (Christophe, Milloffe, Bernal & Lidz 2008).