CONSTRUCTING PSEUDOWORDS FOR EXPERIMENTAL RESEARCH: A CASE STUDY

Anthi Revithiadou, Dimitra Ioannou, Maria Chatzinikolaou, Katerina Aivazoglou
Aristotle University of Thessaloniki
Contact: revith@lit.auth.gr

4th Conference on Language Disorders in Greek
TEI of Patras 28-29 September 2012

1. BACKGROUND

- Greek has morphology-determined stress. Stress is lexically-encoded and is assigned on the basis of a grammar-specific principle, e.g., headedness (Revithiadou 1999)

- In inflected words, *stem accent* prevails over inflectional suffix accent, e.g. /buɣáð-ón/ [buɣáðon] `laundry-gen.pl´ (Ralli & Touratzidis 1992; Revithiadou 1999)

- Predictable aspect: The 3σ-*window* yields APU, PU & U stress patterns (Malikouti-Drachman & Drachman 1989; Drachman & Malikouti-Drachman 1999)

(1) *feminine nouns in -a*

a.	θálasa	/θalas-a/	`sea-nom.sg´
b.	kopéla	/kopél-a/	`girl-nom.sg´
c.	aɣorá	/aɣor´-a/	`market-nom.sg´

(2) *masculine nouns in -os*

a.	pólemos	/polem-os/	`man-nom.sg´
b.	sirókos	/sirók-os/	`SE wind-nom.sg´
c.	uranós	/uran´-os/	`sky-nom.sg´

- The ***phonological default*** (=non-lexically inflicted stress) has been claimed to fall on the **APU** syllable (cf. Malikouti-Drachman & Drachman 1989; Ralli & Touratzidis 1992)

Problems with APU stress as the default: (a) It is not the preferred stress pattern in reading tasks (Protopapas et al. 2006); (b) It is marginal in suffixless words, e.g. acronyms (Revithiadou et al. 2011; Topintzi & Kainada 2011), and in inflected words (Apostolouda 2012)

AIM: IN THIS PAPER, WE PRESENT A METHODOLOGY FOR THE CONSTRUCTION OF PSEUDOWORDS THAT EXPLOITS CORPUS - BASED TOOLS FREELY AVAILABLE ON THE INTERNET. THE CONSTRUCTED DATA ARE INTENDED FOR AN EXPERIMENT THAT AIMS ATTESTING THE STATISTICALLY PREFERRED POSITION OF STRESS (=*EMERGING DEFAULT*) IN GREEK

2. AIM OF RESEARCH

Q1: Which stress pattern represents the *emerging default* (= statistically preferred) stress in Greek?

Q2: Does stress position hinge on type of inflection/ morphological classhood?

In order to answer these questions, we designed and carried out a production experiment:

Target groups: (a) preschoolers and (b) elementary school students (a´& b´grade)
→ developing or no reading skills
Items: 200 pseudonouns from five major morphological classes: *-os, -o, -a, -as, -i* fem

The words were orally presented to the participants by a Robot-like character which uttered them with equal stress prominence. The participants were prompted to produce the input word with a specific stress pattern (see Revithiadou et al. in progress)

Methodological issue:
The words must be unfamiliar but still `sound´ Greek enough to the young speakers´ ears

Focus of this study: The construction of pseudowords for a production experiment on morphology-oriented stress

3. METHOD 1: COMPILING A MORPHOLOGY - ORIENTED CORPUS FOR PSEUDOWORD CONSTRUCTION

THE CORPUS

Clean Corpus (a component of the "ILSP Psycho-Linguistic Resource", http://speech.ilsp.gr/iplr, cf. Protopapas et al. 2010)

Variables
Clean provides a set of quantitative measures for each word. The relevant ones for the purposes of this study are:

(3)
a. Bigram frequencies (phonemes only): i. Logmean bigram token frequency; ii. Logmean bigram type frequency

b. Neighborhoods & cohorts: i. N phonological neighbors (replace only); ii. N phonological neighbors (replace, delete, insert, transpose); iii. Phonological Levenshtein distance 20

The variables in (3) allow us to control whether the constructed words are close to but yet not too distant from existing ones

Problem with Clean Corpus: No morphological categorization (e.g., nouns, verbs, pronouns, etc.) of listed words,which is required in the present research due to the morphology-based nature of Greek stress

→ Verb stress ≠ Noun stress (Revithiadou 1999)

Solution: NClean Corpus
- A finer-grained corpus consisting of only **nouns** was culled up from the Clean Corpus (version: ignoring stress)
- NClean-specific values for the variables in (3) were calculated anew

4. METHOD 2: CONSTRUCTING PSEUDOWORDS FOR EXPERIMENT-SPECIFIC PURPOSES

THE ITEMS

Factors controlled for: (a) Type of inflection/ morphological classhood; (b) Word size (2σ, 3σ words) –20 items of each size, 40 items of each class → 200 pseudowords; (c) Syllable structure: CV.CV(C), CV.CV.CV(C), CCV.CV(C), CCV.CV.CV(C)

Procedure:
Step 1: NClean Corpus nouns were categorized according to their size and syllable structure
Step 2: Mean values and STDs for the variables in (3) were calculated afresh for each category. The acceptable range was set from mean –STD to mean +STD

		BGtokfreqPho	BGtypfreqPho	nNeiPho	nNeiRDITPho	PLD20
		0,135 to 1,867	0,305 to 1,996	8 to 28	9 to 33	0,942 to 1,562
Feminine pseudonouns in -a						
ζακα	zaka	1.675	1.948	13	14	1.350
κιντα	kida	0.408	0.617	13	13	1.450
χιπα	Xipa	1.224	1.208	12	13	1.400
λεσα	lesa	1.213	1.726	9	10	1.500
κοφα	kofa	0.554	0.890	13	15	1.250
χιζα	Xiza	0.330	0.643	11	12	1.400
ροβα	rova	0.656	0.974	10	11	1.450
φιδα	fiTa	0.565	0.613	11	11	1.450
μαδα	maDa	0.592	0.753	17	21	1.050
τουνα	tuna	1.724	1.059	11	11	1.450

Step 3: Words were constructed and tested by the **NumTool** (http://speech.ilsp.gr/iplr/NumTool.aspx, Protopapas et al. in press), which provides quantitative measures of the variables in question for each submitted word string

NUM Tool

Enter up to 20 words or nonwords:

κεβη
σπελος
ζαβοτα
στιβορο

Spelling	Phonetic	Logmean bigram token frequency (phonemes only)	Logmean bigram type frequency (phonemes only)	N phonological neighbors (standard: replace only)	N phonological neighbors (replace, delete,insert, transpose)	Phonological Levenshtein distance 20
κεβη	cevi	0.732	0.829	13	16	1.200
σπελος	spelos	0.949	1.146	2	4	1.850
ζαβοτα	zavota	0.420	0.783	1	1	2.000
στιβορο	stivoro	1.366	1.541	1	1	1.950

Step 4: Words that fell within the defined range (see Step 2) were selected as items for the experiment

This study demonstrates the usefulness of corpora and associated quantitative tools in constructing experimental material that complies to the phonotactic restrictions of Greek. Moreover, it shows that the incorporation of morphological information enhances their applicational power leading towards more targeted results

Acknowledgements: This research was supported by the `Excellence 2011' Program [Action B, Ref.No 878883] awarded to Dr. Anthi Revithiadou and funded by the Research Committee of the Aristotle University of Thessaloniki

Selected references: [1] Drachman, G. & A. Malikouti-Drachman. 1999. Greek word accent. In Hulst, H. van der (ed.), *Word Prosodic Systems in the Languages of Europe*. Berlin & New York: Mouton de Gruyter. 897-945. [2] Malikouti-Drachman, A. & G. Drachman. 1989. Stress in Greek [**Τονισμός στα Ελληνικά**]. *Studies in Greek Linguistics 1989*. University of Thessaloniki. 127-143. [3] Protopapas A., M. Tzakosta, A. Chalamandaris, & P. Tsiakoulis. 2010. IPLR: An online resource for Greek word-level and sublexical information. *Language Resources & Evaluation*. [4] Revithiadou, A. 1999. *Headmost Accent Wins*. Ph.D thesis, LOT Dissertation Series 15, The Hague: Holland Academic Graphics. [5] Touratzidis, L. & A. Ralli 1992. A computational treatment of stress in Greek inflected forms. *Language and Speech* 35: 435-453.