

Word Selection in the Slovenian Sentence Matrix Test for Speech Audiometry

Tatjana Marvin, Jure Derganc, Samo Beguš, Saba Battelino

Background and Research Goals

- Creating the „*matrix sentence test*“ for Slovenian;
- Used for clinical assessment of hearing in people with a hearing disorder.



Matrix Sentence Test

- **Name-Verb-Numeral-Adjective-Noun**
- **Base matrix:** 50 words, 10 per position;
- A computer program chooses a random combination:

Name	Verb	Number	Adjective	Noun
Peter	got	three	large	desks
Kathy	sees	nine	small	chairs
Lucy	brought	seven	old	tables
Alan	gives	eight	dark	toys
Rachel	sold	four	heavy	spoons
William	prefers	nineteen	green	windows
Steven	has	two	cheap	sofas
Thomas	kept	fifteen	pretty	rings
Doris	ordered	twelve	red	flowers
Nina	wants	sixty	white	houses

- Guidelines by International Collegium of Rehabilitative Audiology (ICRA), (Akeroyd et al., 2015), standard ISO 8253-3:2012.

Criteria for Word Selection

- Name-Verb-Numeral-Adjective-Noun;
- 5 female + 5 male names;
- Highly frequent, non-offensive words;
- Grammatically correct sentences (SLO: verb in present tense, Numerals 5+);

eg. Jana kupi pet/šest/deset/itd. velikih škatel.

Jana buys five/six/ten/etc. big boxes

- ***Phonemic balance.***

Slovenian Phonemes

25 Letters → 29 Phonemes → >40 Allophones

Letter N → Phoneme /n/ → Allophones [n] (noga "leg"); [N] (banka "bank");

Slovenian: general tendency *letter ≈ phoneme*,

Exceptions:

letter E: /e/ in *led* "ice", /E/ in *žep* "pocket" or /@/ in *pes* "dog"

letter O: /o/ in *nos* "nose" or /O/ in *noga* "leg"

phoneme /dZ/ ≈ letters DŽ *džip* "jeep"

phoneme /@/ ≈ { } *vrt* "garden"

phoneme /j/ ≈ { } *pacient* "patient"

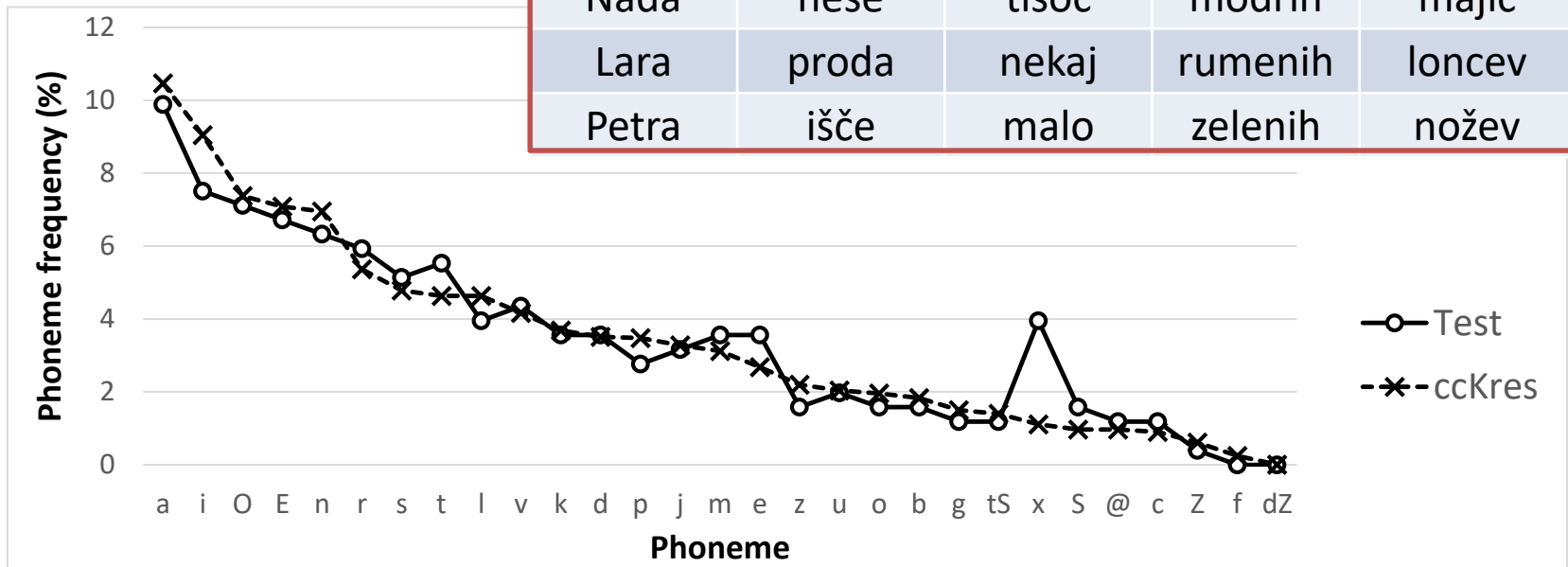
Obtaining Phonemic Balance in Slovenian

Orthographic data (the letter distribution) + Adjustment of distribution for phonemes /e/, /E/, /@/, /o/, /O/, /dZ/, /j/

1. Orthographic data: the corpus of written Slovenian ccKres (10 million words, a balanced genre structure);
2. Adjust number of phonemes /j/, /@/;
3. Count occurrences of /dZ/;
4. Distribution of phonemes denoted by letters O and E: Mihelič (2006), allophone distribution in 300.000 phonetically transcribed sentences;
 - Letter O: /o/ (21 %), /O/ (79 %).
 - Letter E: /e/ (25 %), /E/ (66 %) and /@/ (9 %).

The Slovenian Matrix Test

Name	Verb	Numeral	Adjective	Noun
Gregor	kupi	pet	velikih	stolov
Tone	dobi	šest	lepih	copat
Jure	najde	sedem	novih	škotel
Urban	skrije	osem	čudnih	avtov
Sašo	vzame	enajst	starih	zvezkov
Branka	ima	sto	dobrih	koles
Jana	pelje	tristo	dragih	kamnov
Nada	nese	tisoč	modrih	majic
Lara	proda	nekaj	rumenih	loncev
Petra	išče	malo	zelenih	nožev



Thank you!

Procedure:

- All letters in the corpus cckres are transformed into lower case characters. Next, the standard Slovenian diacritic marks on the letters “a”, “e” and “o” are discarded (“á” → “a”, “à” → “a”, “é” → “e”, “ê” → “e”, “è” → “e”, “ô” → “o”, “ó” → “o”). Finally, all the characters that are not in the standard Slovenian alphabet (except for “đ”) are discarded from the corpus.
- The number of phonemes /dʒ/ is determined by counting the total occurrences of “dž” and “đ”.
- The number of phonemes /j/ is adjusted by adding the occurrences that are pronounced, but not expressed in writing between the two vowels in the following combinations: “ia”, “ie”, “io”, “ea”, “oi”. The number of phonemes /j/ is reduced in the instances where the latter is found in spelling, but is not pronounced: nj#, njC, lj#, ljC.
- The number of phonemes /o/ and /O/ is determined by dividing the number of letters “o” according to the distribution in Mihelič (2006): /o/ (21 % of letter “o” occurrences), /O/ (79 % of letter “o” occurrences).
- The number of phonemes /e/, /E/ and /@/ is determined by first summing the number of letters “e” plus the number of occurrences of /@/ that are not expressed in writing. According to Toporišič (2000), the phoneme /@/ can be found in combinations with “CrC”, “Cr#”, “#rC”, “vn#”, “jn#”, “ln#”, “lm#”, “jm#”, “lmN”, “jmN”, “jnN”, “lnN”, “vnN”, where “C” stands for any consonant, “N” for any obstruent, and “#” for a word boundary. The total count of these occurrences is divided into the phoneme counts according to the distribution of these three phonemes in Mihelič (2006): /e/ (25 %), /E/ (66 %) and /@/ (9 %).

