Paradigmatic space and defectiveness in Hungarian – an empirical study

Ágnes Lukács ♠ ♦

Péter Rebrus ♦ ♥

Miklós Törkenczy ♣ ♦ ♥

♠ HAS-BME Cognitive Science Research Group

♦ Research Institute for Linguistics, HAS, Budapest

♥ Department of Theoretical Linguistics, HAS–ELTE Budapest

♣ Department of English Linguistics, Eötvös Loránd University, Budapest

Dialogues – paradigm gaps: the phenomenon

(1) Two squirrels are talking:
   *Én majd erre a fára **ugrok**, maga csak* [OK]***ugorjon*** *oda*
   'I'll jump on this tree, you **should jump** on there.'

(2) Two hens are talking:
   *Én majd **kotlok** itt, maga csak* [?*]***kotoljon*** *ott.*
   'I'll brood here, you **should brood** over there.'

(3) Two snakes are talking:
   *Én **vedlek** a saját fészkemben, maga csak ***vedeljen** ott.*
   'I'll shed my skin in my own nest, you **should shed** yours there.'

Outline

- Paradigm gaps in Hungarian verbs
  - Morphophonological background
    - Stem and suffix classes
    - Locations and properties of gaps
  - A paradigm-based analysis
    - Distribution of allomorphs in the verbal paradigm
    - Analogical motivation of gap irreparability
- Empirical results on paradigm gaps
  - Testing gap locations and analogical relations
  - Testing gap properties: uncertainty and forced repair

Gap locations are independent of (morpho)syntactic category

- Defectiveness effects about 65 verbal stems:
  - intransitive: *csukl-ik* 'hiccup', *sikl-ik* 'glide', *ízl-ik* 'taste good', *patakz-ik* 'gush', *habz-ik* 'foam', *burjánz-ik* 'proliferate' etc.
  - (optionally) transitive: *vedl-ik* 'shed skin', *háml-ik* 'peel', *sínyl-i* 'suffer', *kétl-i* 'doubt' etc.

- Gaps occur with wildly different 'analytic' suffixes:
  - Subjunctive/Imperative   *csuk(o)l-**j***   'hiccup!'
  - Modal   *csuk(o)l-**hat***   'may hiccup'
  - Adverbial Participle   *csuk(o)l-**va***   'hiccuping'
  - Definite Person/Number   *csuk(o)l-**ja***   's/he hiccups it'

Gap locations are phonotactically and lexically motivated

- Defective stems end in CC:
  - with sonority rise: *csukl-, kotl-, vedl-, fesl-, bűzl-, háml-, sínyl-, habz-*
  - with relatively little sonority fall: *morajl-, porl-, nyálz-, párz-, burjánz-*
- The class of defective stems is lexically determined:
  - *roml-ik ~ romol-hat* vs. *háml-ik ~ *hámol-hat*
  - *vonz ~ vonz-hat* vs. *burjánz-ik ~ *burjánz-hat*
- Direction of *forced repair* is influenced by stem-final CC
- Interspeaker variability and uncertainty of forced repair and classification

---

Background: morphophonological **stem** classes (final CV patterns)

i. **stable VC**-stem (stable):   always **VC-**

   e.g. *rámol ~ rámol-ok ~ rámol-hat*

ii. **epenthetic**, non-*ik* (alternating):   **CC- ~ VC-**

   e.g. *söpör ~ söpr-ök ~ söpör-het*

iii. **epenthetic** *ik*-stem (alternating):   **CC- ~ VC-**

   e.g. *roml-ik ~ roml-ok ~ romol-hat*

iv. **defective** (CC-)stem (stable):   always **CC-**

   e.g. *háml-ik ~ háml-ok ~ *hám(o)l-hat*

v. **stable CC**-stem (stable):   always **CC-**

   e.g. *hord ~ hord-ok ~ hord-hat*

Background: morphophonological **affix** classes (initial CV patterns)

    a. **synthetic** (stable):   always **-V…**

           e.g. *ad-ok, ad-om ~ hord-ok, hord-om*

    b. **quasi-analytic** (alternating):   **-VC… ~ -C…**

           e.g. *ad-nak, ad-tok ~ hord-anak, hord-otok*

    c. <u>**analytic**</u> (stable):   always **-C...**

           e.g. *ad-va, ad-hat ~ hord-va, hord-hat*

Defectiveness

- Defectiveness occurs when a **defective** CC-**stem** (iv) and an **analytic suffix** (c) combine

- Defectiveness is (partly) phonologically motivated:

$$*... C_1C_2 + C ...$$

where C_1C_2C is phonotactically ill-formed

e.g.   **há<u>ml</u>-<u>h</u>at*   **hám<u>ol</u>-hat*   **há<u>ml</u>-<u>o</u>hat*

cause: **phonotactics**   **stem** paradigm   **suffix** paradigm

Verbal paradigm: stem-final VC vs. CC

suffix types: stem types:	a. Synthetic	b. Quasi-analytic	c. **Analytic**
i. stable VC	rám*ol*ok	rám*ol*nak	rám*ol*hat
ii. epenthetic non-*ik*	söp*rö*k	söp*ör*nek	söp*ör*het
iii. epenthetic -*ik*	om*l*ok	om*l*anak / omo*l*nak	om*ol*hat
iv. **defective**	hám*l*ok	hám*l*anak	--
v. stable CC	hor*d*ok	hor*d*anak	hor*d*hat

Verbal paradigm: the two patterns

suffix types: stem types:	a. Synthetic	b. Quasi-analytic	c. **Analytic**
i. stable VC	V	V	V
ii. epenthetic non-*ik*	C	V	V
iii. epenthetic -*ik*	C	C / V	V
iv. **defective**	C	C	--
v. stable CC	C	C	C

The analysis of Hungarian paradigm gaps

(Rebrus & Törkenczy to appear)

Repair must
- ➢ not violate **phonotactics** &
- ➢ be **local** &
- ➢ be **conservative**.

where
- *locality*: reassignment can only target a point in the paradigm where a gap occurs
- *conservativity*: reassignment must be into another existing stem-class

Hungarian gaps are irreparable

- stem classes: closest neighbours

iii. epenthetic	⟨C	C/V	V⟩	
iv. defective CC	⟨C	C	--⟩ →	**?**
v. stable CC	⟨C	C	C⟩	

- potential repair is
  - **unphonotactic**: *hámlhat* → ⟨C C **C**⟩
    or
  - **non-conservative**: *hámolhat* → ⟨C C **V**⟩
    or
  - **non-local**: *hámolnak hámolhat* → ⟨C C/**V** **V**⟩

Intraparadigmatic relations

- Conservatism crucially refers to existing stem classes, but pressupposes no systematic constraints on paradigm space

- But: stem paradigms are patterned

  there are implicational (analogical) relations between forms:

  > The patterns of Q-forms are determined
  > by the patterns of Basic-forms and A-forms

Analogical relations: Q-form is determined by Basic form and Analytic form

suffix types: stem types:	Basic (3Sg.Pres)	Synthetic	Quasi-analytic	Analytic
i. stable VC	V	V	V	V
ii. epenthetic non-*ik*	V	C	V	V
iii. epenthetic -*ik* (E)	C	C	C / V	V
iv. defective (D)	C	C	C	--
v. stable CC (CE/CD)	C	C	C	C

Hypotheses about paradigmatic space

1. The stem alternant that appears in the A-form always appears as a stem alternant in the Q-form.

   epenthetic stems: A: *ugor-hat* and Q: *ugor-nak*
   stable CC-stems: A: *hord-hat* and Q: *hord-anak*
   defective stems: no A-forms (vacuously satisfied)

2. There is no optional A-form:

   epenthetic stems: A: *ugorhat* / *ugr(o)hat*.

3. There always exists a Q-form: no defectivity with a Q-suffix.

Psycholinguistic experiment

- 100 adult participants (MA: 27.4)
- **lexical decisions** about 12 forms of 120 stems (1440)
  - 26 epenthetic *ik*-stems (E)
  - 34 defective (D)
  - 60 -CC control matched individually on frequency to the defective and epenthetic stems (CD, CE).
  - For each stem, VC and CC alternants generated with 4 quasi-analytic (Q) and 2 analytic (A) suffixes
  - Task: to decide for each alternant whether it is an existing word of the language or not.

Stimuli

		Defective (D)		Epenthetic (E)		Stable CC (CD, CE)	
		-CC	-VC	-CC	-VC	-CC	-VC
-nAk	Q1	*hámlanak*	*hámolnak*	*ugranak*	*ugornak*	*toldanak*	*tolodnak*
-tOk	Q2	*hámlotok*	*hámoltok*	*ugrotok*	*ugortok*	*toldotok*	*tolodtok*
-nA	Q3	*hámlana*	*hámolna*	*ugrana*	*ugorna*	*toldana*	*tolodna*
-ni	Q4	*hámlani*	*hámolni*	*ugrani*	*ugorni*	*toldani*	*tolodni*
-hAt	A1	*hámlhat*	*hámolhat*	*ugrhat*	*ugorhat*	*toldhat*	*tolodhat*
-vA	A2	*hámlva*	*hámolva*	*ugrva*	*ugorva*	*toldva*	*tolodva*

Canonical forms (i.e. forms generally considered grammatical) highlighted in yellow

Canonical vectors

- Responses by stem/participant are 12-tuples (Boolean vectors):

  ⟨ CC1Q1 VC1Q1 CC1Q2 VC1Q2
     CC1Q3 VC1Q3 CC1Q4 VC1Q4
     CC1A1 VC1A1 CC1A2 VC1A4 ⟩

  where Q1 Q2 Q3 Q4 are Q-suffixes, A1 A2 are A-suffixes

- Canonical vectors:
  - ➢ stable CC-stems (CD, CE):

    ⟨ 1 0 1 0 1 0 1 0 | 1 0 1 0 ⟩
  - ➢ epenthetic stems (E):

    ⟨ 1 1 1 1 1 1 1 1 | 0 1 0 1 ⟩
  - ➢ defective stems (D):

    ⟨ 1 0 1 0 1 0 1 0 | 0 0 0 0 ⟩

Results

- The assumptions were tested by comparing actual distribution of 'accept' (1) and 'reject' (0) answers to random distribution with χ^2-tests.

- We rejected a hypothesis if the number of refuting cases was significantly larger than would be expected from random answers.
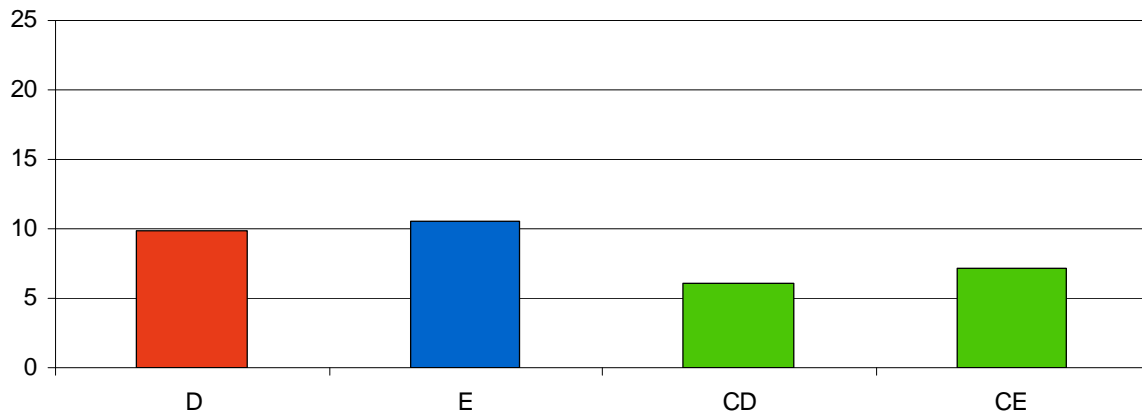
Hyp1.
The stem alternant that appears in the A-form always appears as a stem alternant in the Q-form.
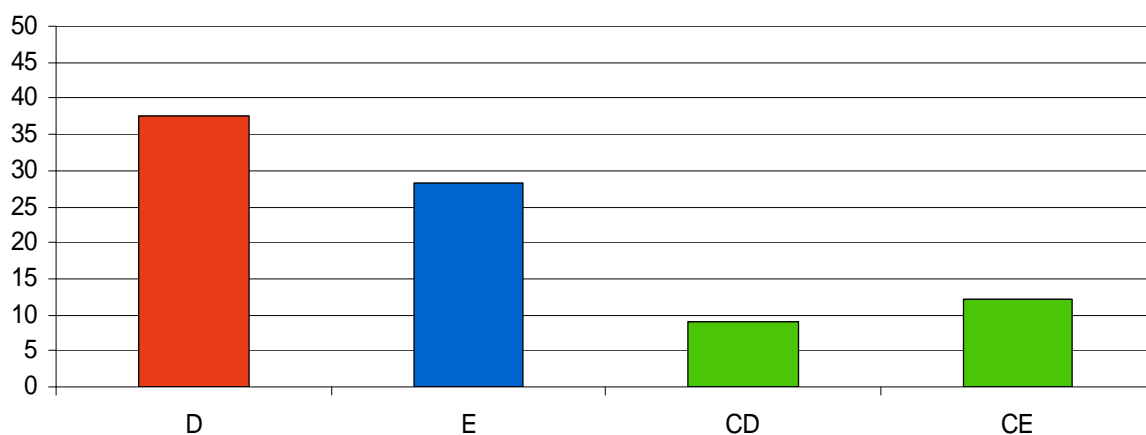
Percentage of refuting cases by stem type.
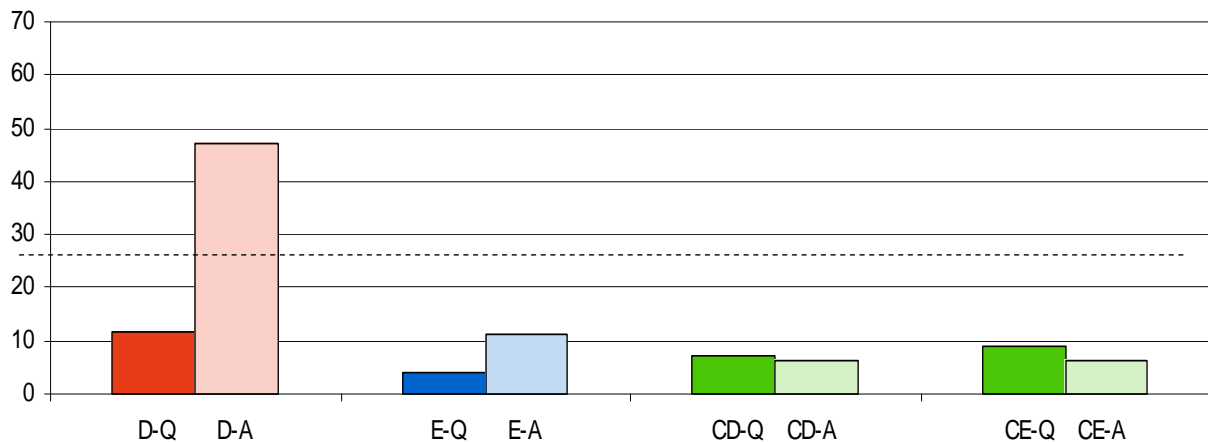
All significantly less than would be expected by chance at p<0.001

Hyp 2.
There is no optional analytic form

Percentage of refuting cases by stem type.

All significantly less than would be expected by chance at p<0.001

Hyp 3. Defectivity with Q- and A-suffixes: There always exists a Q-form
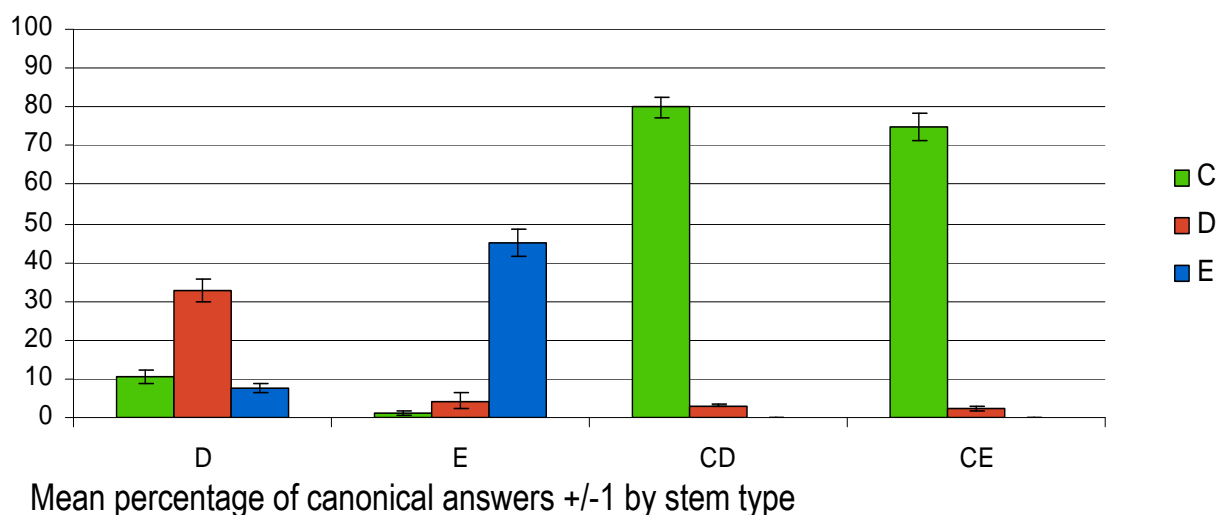

Percentage of refuting cases by stem type.

All significantly less than would be expected by chance at p<0.001. D-A significantly more than chance (p<0.001).
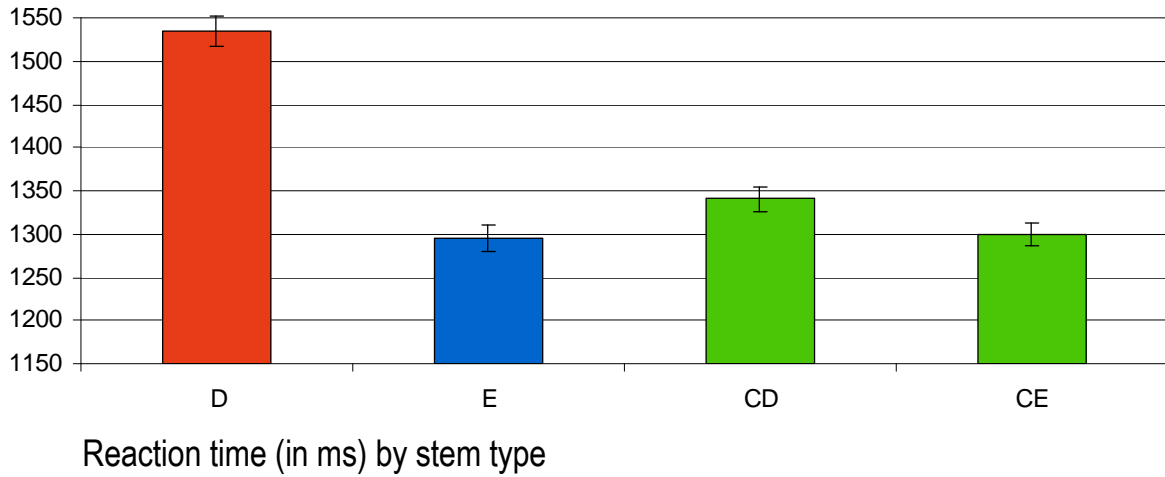
Classification, uncertainty and forced repair

1. There is **no stem** class that would fall **outside** the paradigmatic space defined by the analogical relations.

2. Assignment of stems into classes by participants shows differing degrees of **uncertainty** (stable CC < E < D) :
   - longer **reaction times**
   - greater **variability** in response patterns

3. Forced repair of D-stems
   - more 'accept' answers than expected in class D
   - direction depends on stem shape (sonority relations)

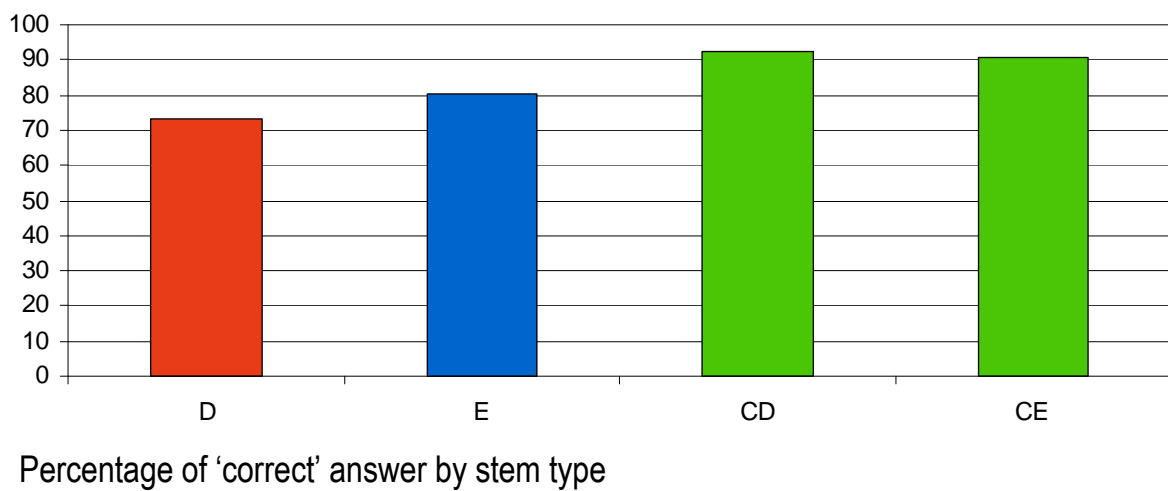# There is no stem class outside those defined by canonical vectors

- Response patterns: the 3 canonical vectors have the highest frequencies of occurence
- The 30 most frequent vectors (those with a relative frequency > 0.5%) are within a Hamming distance of 1 from canonical vectors
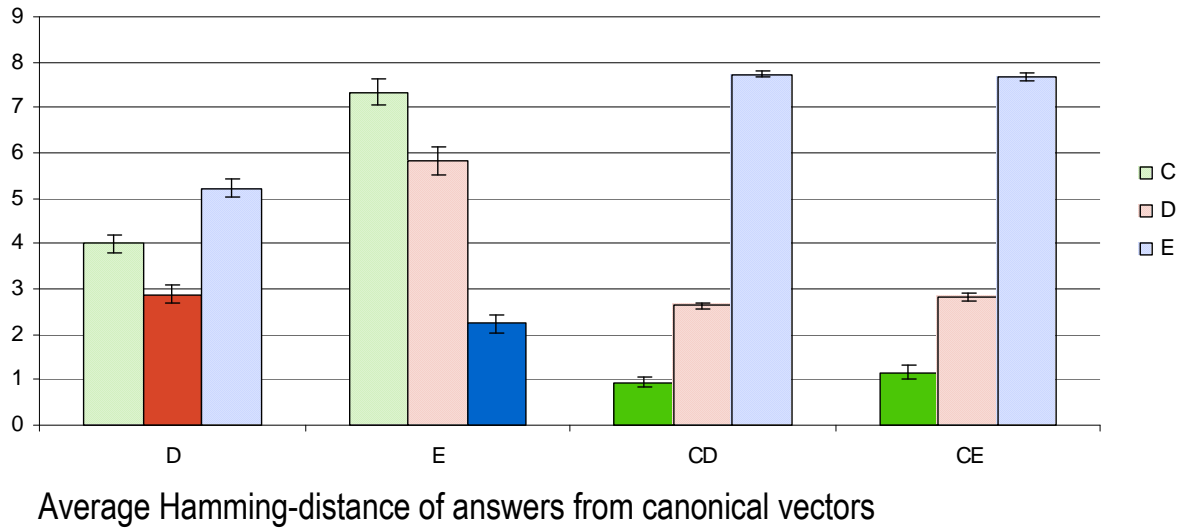

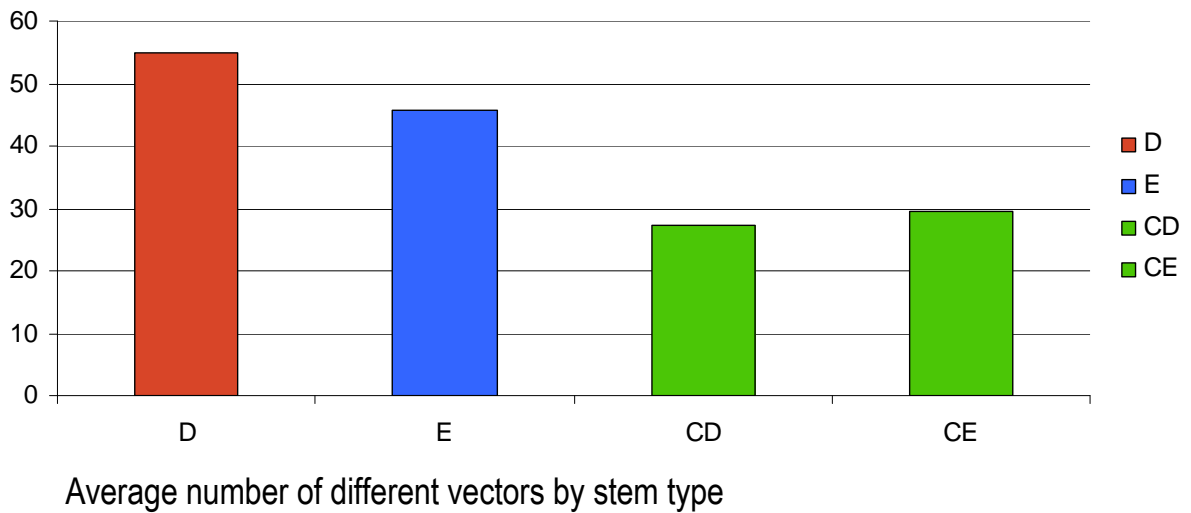Mean percentage of canonical answers +/-1 by stem type

Uncertainty: reaction times

Reaction time (in ms) by stem type

Uncertainty: variability/1

Percentage of 'correct' answer by stem type

Uncertainty: variability/2

Average Hamming-distance of answers from canonical vectors

Uncertainty: variability/3

Average number of different vectors by stem type

Uncertainty: variability/4

Entropy by stem type
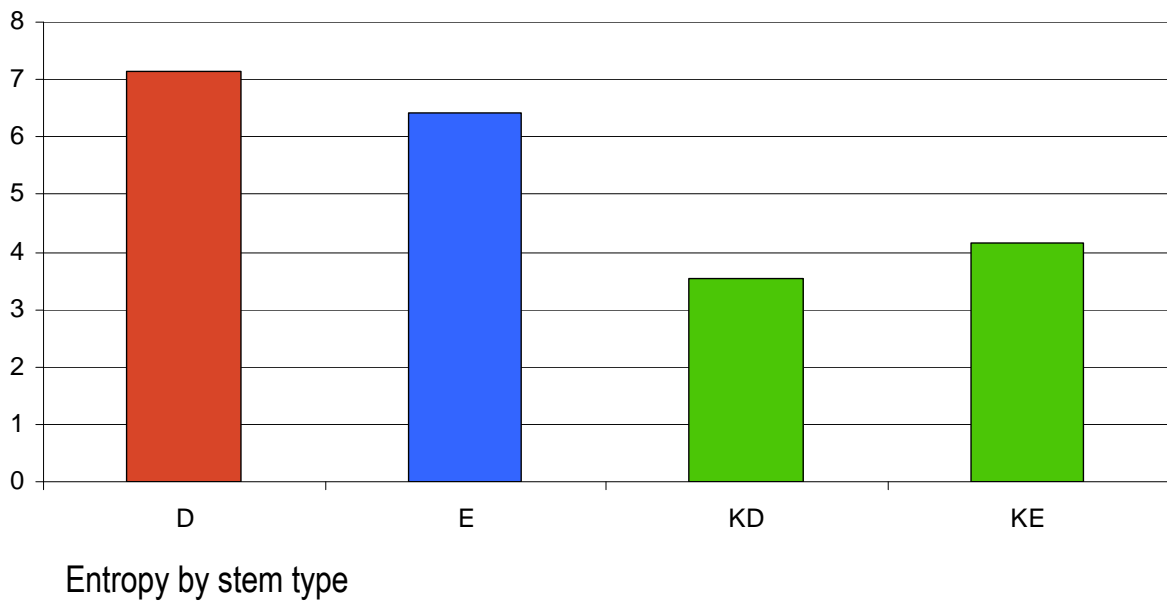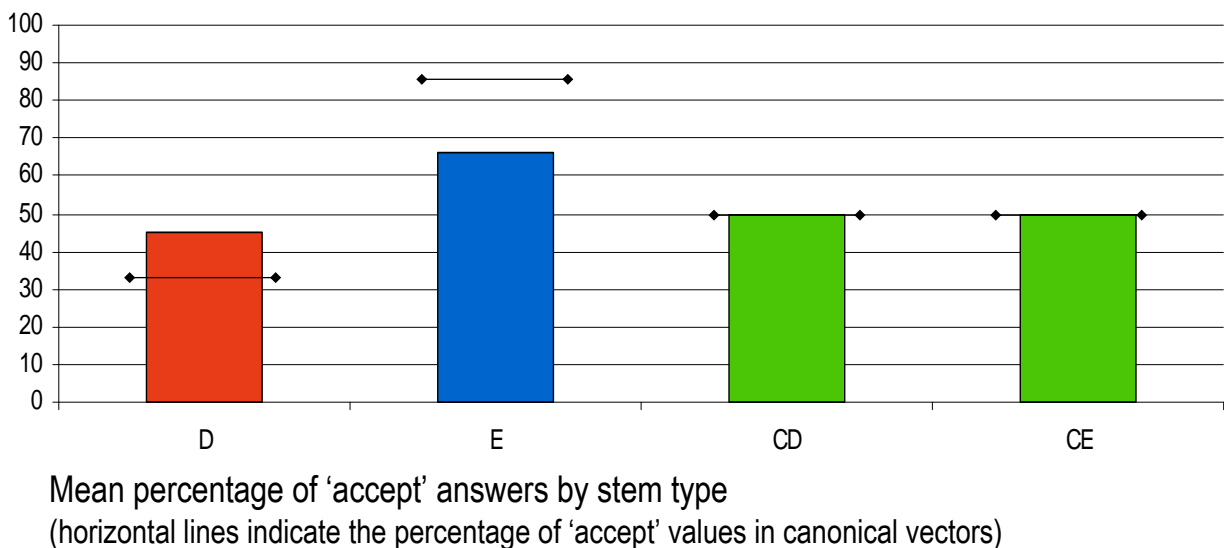
Forced repair: 'accept' answers

• Speakers are motivated to accept non-canonical forms in the Defective class, but not in others (locality)

Mean percentage of 'accept' answers by stem type
(horizontal lines indicate the percentage of 'accept' values in canonical vectors)

Forced repair: C_1C_2 phonotactics

- Tentative sonority scale:

  glide < hom. nasals < liquids < non-hom. nasals < fricatives < plosives

  j nt nd nk ng l r   m ny   s sz z zs t d k g b p

   0 0.5 1   2   3   4

- Sonority slope value = $C_2 - C_1$

  e.g. *csukl*-: **−3**, *fesl*-: **−2** , *háml*-: **−1**, *habz*-: **−1**

    porl-: **0**, *morajl*-: **1**, *párz*-: **2**, *burjánz*-: **2.5**

Forced repair: C_1C_2 phonotactics

- Hypothesis:

  the greater the sonority difference in stem-final CC, the greater the acceptability of Defective CC-stems + Analytic suffixes.

- Correlations (Pearson) with sonority difference:

  sonority $\times$ number of accept answers (only with CC-stems)

overall:	r=0.453; p<0.001
Defective stems:	r=0.329; p<0.001
Defective stems + Analytic suffixes:	r=0.693; p<0.001