Michael Becker & Andrew Nevins
{becker, nevins}@fas.harvard.edu

NELS 40 @ MIT
November 13-15, 2009

Initial-syllable faithfulness as the best model of word-size effects in alternations*

- Surprisingly, whether a noun undergoes *voicing alternation* is not 50%∼50% (Ernestus & Baayen 2003); it is often a function of word-length (henceforth, "a size effect").
- The regulation of phonological behavior is a subject of current debate in the study of language: Do large-scale trends represent the result of grammaticalized usage pressures or not?
- We compare the ability of different measures to predict voicing alternations in Turkish and Russian.
- We show that monosyllabicity, a *discrete grammatical* factor, is the best predictor among these. It is better than gradient grammatical measures, and much better than word-similarity measures such as neighborhood density.

1 Sources of size effects

In Turkish, voicing alternations affect some short words,

(1) tatʃ ∼ tadʒ-ɨ 'crown NOM/POSS'
satʃ ∼ satʃ-ɨ 'hair NOM/POSS'

and some long words:

(2) amatʃ ∼ amadʒ-ɨ 'goal NOM/POSS'
anatʃ ∼ anatʃ-ɨ 'cub NOM/POSS'

*For their thoughtful comments and feedback, we thank Adam Albright, Jonathan Barnes, Maria Gouskova, John McCarthy, Andreea Nicolae, and Engin Sezer. Any remaining errors are due to an unbalanced diet.

The shorter words, however, are much less likely to alternate (Lees 1961; Inkelas & Orgun 1995; Inkelas et al. 1997; Hayes 1995; Pycha et al. 2007; among others).
What is the best characterization of this size effect?

(3) Initial-syllable faithfulness: Monosyllables are more likely to be protected from alternations than polysyllables (Becker, Ketrez & Nevins 2008)

(4) Moraic-based markedness:
- Minimal CVC words are syllabified early, thereby escaping an alternation pressure that applies to larger-than-minimal words (Inkelas & Orgun 1995; Inkelas et al. 1997; Pycha et al. 2007)
- Word-minimality (FtBin in OT) is a markedness pressure to expand sub-minimal words. In a serial model, FtBin can cause earlier syllabification of stops in CVC nouns, and greater faithfulness to syllabified consonants can prevent alternations.

(5) Neighborhood density
- Neighbors (Luce & Pisoni 1998): Words that are only one segment's deletion, addition, or substitution away.
- Short = competitive: Words in dense lexical neighborhoods are protected from alternations (Wedel 2002; Ussishkin & Wedel to appear). Shorter words have more neighbors, pressuring them to keep their shape constant in order to facilitate lexical access in the face of many phonologically-close lexical competitors (though see Pycha et al. 2007).

Other potential predictors:

(6) Structural/grammatical factors: Alternation is correlated with syllable count, mora count, segment count, etc.

(7) Lexicon-based/similarity-based factors: Alternation is correlated with cohort size, uniqueness point (Marslen-Wilson & Welsh 1978; Luce 1986; Gaskell & Marslen-Wilson 2002), etc.

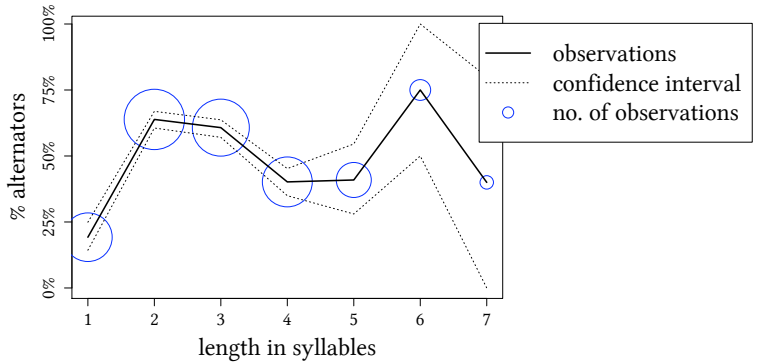
2 Turkish voicing alternations

2.1 The Turkish lexicon

Data: The 3049 stop-final nouns of TELL (Inkelas et al. 2000), an electronic lexicon.

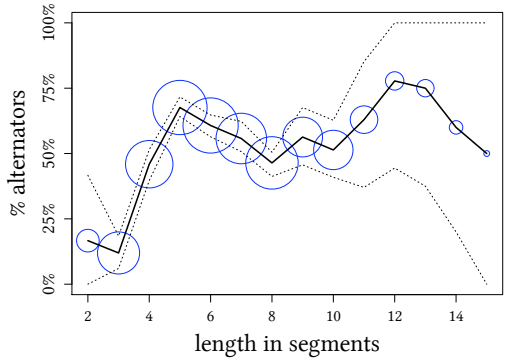
The correlation between length in syllables and alternation is mostly concentrated in the mono/polysyllabic distinction:

(8) Turkish alternation rates by syllable, with confidence intervals



(9) Confidence intervals were calculated by taking 100 random samples from the data and taking 1.5*the interquartile range on either side of the mean prediction at each point.

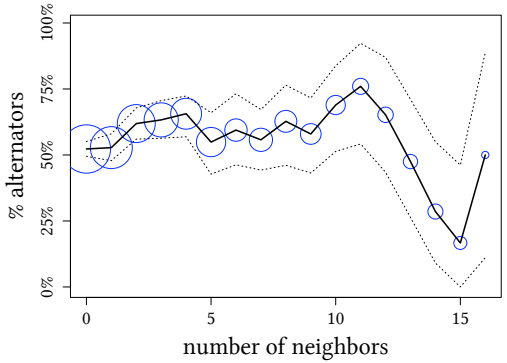
(10) Turkish alternation rates by segment, with confidence intervals



Counting by syllables is better than counting by segments: length in segments reliably correlates with alternation only in the 3-5 range:

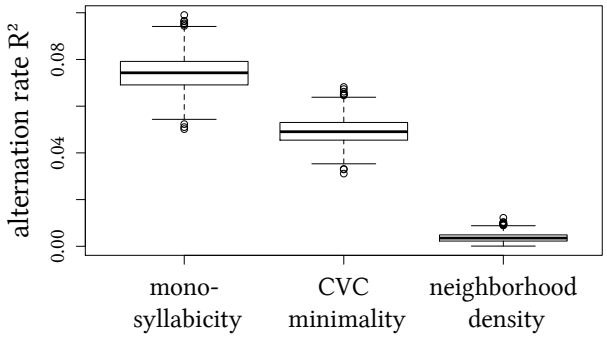
Neighborhood density, however, is hardly correlated with alternations at all: The line is rather flat for words with less than 12 neighbors, which make up 95% of the lexicon. This confirms the finding in Pycha et al. (2007).

(11) Turkish alternation rates by neighborhood size, with confidence intervals



How well do various size measurements correlate with alternations?

(12) Turkish: Correlation scores for three models of alternations



The ranges of R^2 value from three logistic regression models, each done over 1000 random samples.

(13) Monosyllabicity consistently achieves a significantly higher R^2 value than CVC minimality (Wilcoxon rank sum test, $W = 10^6$, $p < .001$).

(14) Syllable/mora/segment *count* have R^2 of less than 1%, because the additional increase in size does not correlate with an increase in alternation rate.

The alternation rates in the Turkish lexicon are best modeled by monosyllabicity, rather than by word-minimality or neighborhood density.

2.2 Turkish-speaking humans

Is the Turkish pattern just limited to the lexicon, or are speakers attuned to it, and use it in their treatment of novel items?

In Becker, Ketrez & Nevins (2008), 24 speakers were asked to choose alternating or non-alternating forms for 72 novel nouns:

(15)

shape	no. of items	% alternating
CVC	24	35%
CVCC	24	44%
CVVC	2	49%
CVCVC	22	61%

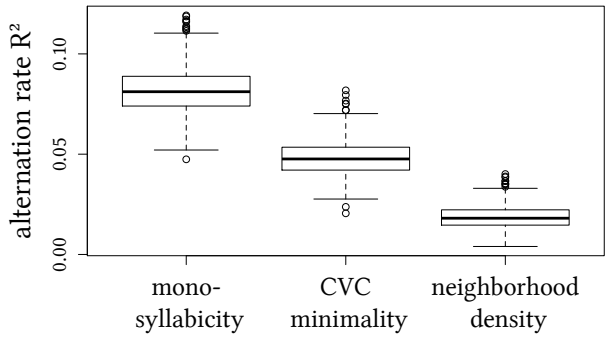
What correlates best with the speaker's behavior?

(16)

predictor	alternation R^2	alternation R^2 with place
monosyllabicity	8%	14%
length in segments	8%	12%
CVC minimality	5%	8%
neighborhood density	2%	4%
uniqueness point	1%	5%
cohort size	<1%	2%

Monosyllabicity consistently achieves a significantly higher R^2 value than CVC minimality in 1000 random samples (Wilcoxon rank sum test, $W = 10^6$, $p < .001$).

(17) Turkish wugs: Correlation scores for three models of alternations



To conclude:

(18) Monosyllabicity is not just the best predictor of alternations in the lexicon, it is also the best predictor of human behavior in tasks that probe alternation rates.

(19) Cohort size *is* an important factor in lexical access and processing, but not in paradigmatic relations among words.

3 Russian voicing alternations

In Russian, like in Turkish, voicing alternations affect some short words,

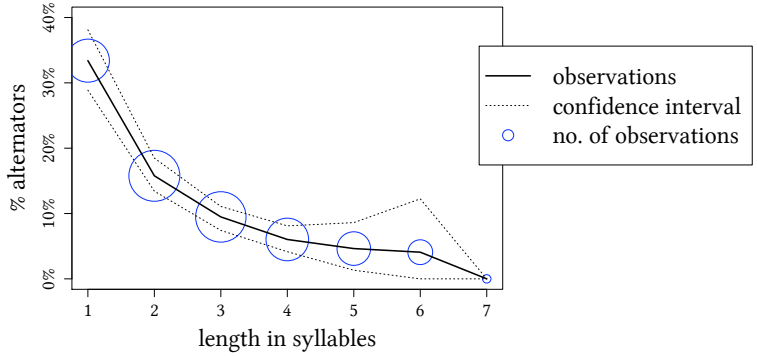
(20) snó**p** ~ snó**b**-a 'snob NOM.SG/GEN.SG'
snó**p** ~ sna**p**-á 'sheaf NOM.SG/GEN.SG'

and some long words:

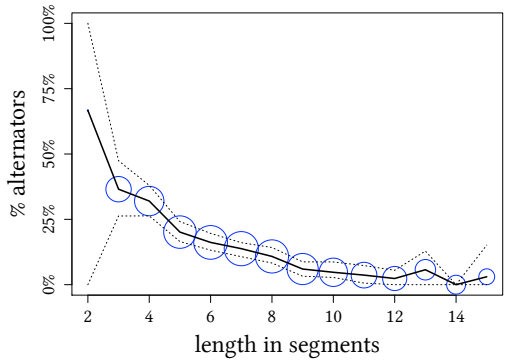
(21) sugró**p** ~ sugró**b**-a 'snowdrift NOM.SG/GEN.SG'
xaló**p** ~ xaló**p**-a 'slave NOM.SG/GEN.SG'

Data from Sharoff's (2005) electronic dictionary (3905 stop-final nouns): In Russian, shorter words are *more* likely to alternate — the mirror image of Turkish.

(22) Russian alternation rates by syllable, with confidence intervals

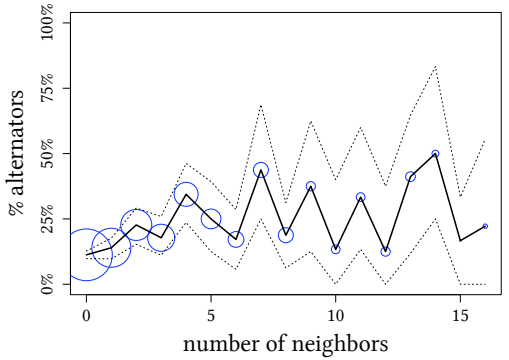


(23) Russian alternation rates by segment, with confidence intervals



Contrary to Ussishkin & Wedel's (to appear) predictions, neighborhood density is actually weakly *positively* correlated with alternations (this is clearest in the 0-4 range, which makes 93% of the data).

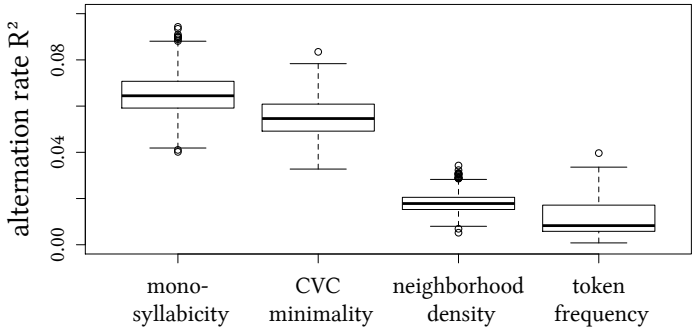
(24) Russian alternation rates by neighborhood size, with confidence intervals



(25) Token frequency is similarly poorly correlated with alternations in Sharoff's (2005) dictionary.

Assessing the confidence in the correlations with 1000 samples for each predictor:

(26) Russian: Correlation scores for four models of alternations



Monosyllabicity is more strongly correlated with voicing alternations than CVC minimality (Wilcoxon rank sum test, $W = 8^*10^5$, $p < .001$).

Because shorter words have more neighbors, and shorter words are more likely to alternate in Russian, neighborhood density is positively correlated with alternations in Russian, not negatively as in Turkish.

4 Comparing Turkish and Russian alternations

Differences between Russian and Turkish:

(27) Direction of the size effect: Size is positively correlated with alternations in Turkish, negatively correlated in Russian.

(28) Turkish has a three-way contrast: Nicolae & Nevins (before lunch) show that Turkish uses [sg] and [voice] (with alternators unspecified), whereas Russian only uses [voice] (with alternators specified for [voice]).

(29) Orthography: Turkish reflects alternations in the spelling, Russian doesn't.

(30) Lower overall alternation rate in Russian.

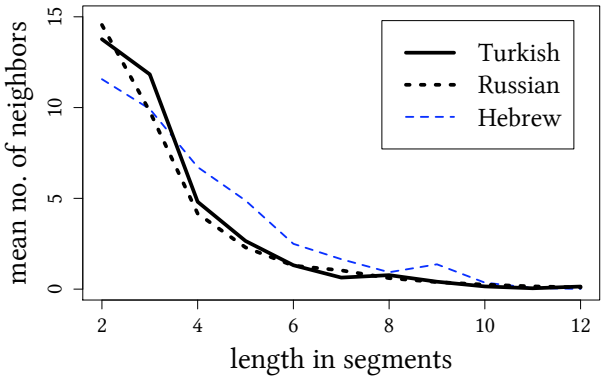
Turkish and Russian commonalities:

(31) In both languages, neighborhood density is poorly correlated with alternations.

(32) In both languages, monosyllabicity is well correlated with alternations.

(33) The relationship between size and neighborhood densities are essentially the same in both languages (unlike Hebrew, where monosyllables are rarer; data from Bolozky & Becker 2006).

(34) Distribution of neighborhood densities



5 Monosyllabicity as an initial-syllable effect

What is the typology of size effects, particularly with respect to the laryngeal alternations?

5.1 Instructive examples

English (Hayes 2009; p. 195):

(35) Plurals and denominal verbs create a three-way voicing distinction in labiodental and inter-dental fricatives:
[sɪv ∼ sɪvz], [motif ∼ motifs], [dwoɹf ∼ dwoɹvz].

(36) Alternations (including historically innovative ones) are restricted to monosyllables for many speakers, i.e. polysyllabicity is negatively correlated with alternations.

Catalan (Wedel 2002; Ussishkin & Wedel to appear):

(37) Alternations are positively correlated with size, as in Turkish:
[sɛrt ∼ sɛrtə] 'certain M./F.' vs. [likit ∼ likiðə] 'liquid M./F.'

(38) Catalan contrasts voiceless unaspirates with voiced stops (which can spirantize), as in Russian.

5.2 Grammar or word-similarity?

(39) "Words undergo, or resist, morphophonemic alternation in a manner unrelated to the noun's relationship to other lexical items." (Pycha et al. 2007; p. 385)

(40) Whether a word-final stop alternates does not depend on the word's neighborhood density; i.e. does not depend on global phonemic similarity to other words in the lexicon. Rather, such patterns should be accounted for in grammatical terms, specifically, monosyllabicity.

5.3 The role of initial-syllable faithfulness

Initial-syllable faithfulness (Steriade 1994; Beckman 1997, 1998; Casali 1998) protects the word-initial syllable from neutralizations and alternations. This includes the initial syllable's onset and nucleus, but also its coda (see Beckman 1998's analysis of Tamil).

In monosyllables, the only syllable is also the initial one.

(41) Turkish: an alternating monosyllable

/ʤoB+I/ [ʤop]	*VTV	OO-Id(voice)$_{\sigma_1}$	OO-Id(voice)
a. ☞ ʤo.bu		*	*
b. ʤo.pu	*!		

(42) Turkish: an alternating polysyllable

/ʃaraB+I/ [ʃa.rap]	*VTV	OO-Id(voice)$_{\sigma_1}$	OO-Id(voice)
a. ☞ ʃa.ra.bɨ			*
b. ʃa.ra.pɨ	*!		

Note: In *ʤop* ∼ *ʤo.bu*, the root-final stop is in the initial syllable only in the base. For defining strong positions relative to the base, see Kager (1999); Jesney (2009).

5.4 The grammar projects lexical statistics

Constraint cloning (Pater 2006, 2009; Becker 2009): Given evidence for inconsistent ranking, lexical items get listed with the most specific ranking they require.

(43) Grammar with cloned initial-syllable faithfulness:
Ident(voice)$_{\sigma_1\ \text{p of monosyllables}}$, Ident(voice)$_{\text{q elsewhere}}$
≫ *VTV ≫
Ident(voice)$_{\sigma_1\ \text{1-p of monosyllables}}$, Ident(voice)$_{\text{1-q elsewhere}}$

(44) In Turkish, p > q, whereas in Russian, p < q; these allow generalizations to novel items in wug-tests.

This analysis prevents a host of possible grammars: Syllable-counting based alternations, distinction between di-syllables and tri-syllables, etc.

5.5 Further questions to pursue

(45) What do Russian speakers do with novel words?

(46) Ident(voice)$_{\sigma_1}$ should protect monosyllables from alternations, all else being equal. Are Turkish and Catalan more natural than Russian and English?

(47) We are running an artificial grammar learning experiment, testing generalization tasks in languages with asymmetric distributions of alternations.

6 Conclusions

- Our statistical analysis shows that the skewed distributions of voicing alternations in Russian and Turkish (and speakers' knowledge of these patterns in novel generalization tasks) are best modeled in formal means by initial-syllable faithfulness, which directly allows for differential treatment of monosyllabic and polysyllabic words
- Other analyses (in terms of moras, segments, or lexicon-based measures) offer less coverage of the variation in the data, or fail to extend to the distribution of alternations in Russian.
- Initial-syllable faithfulness, originally motivated largely by static distributional evidence, thus emerges as a multi-purpose formal tool for keeping track within the grammar of distributional asymmetries in alternations as well.

References

Becker, Michael (2009). *Phonological Trends in the Lexicon: The Role of Constraints.* Ph.D. dissertation, UMass Amherst.

Becker, Michael, Nihan Ketrez & Andrew Nevins (2008). The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish devoicing neutralization. ROA-1001.

Beckman, Jill (1997). Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* **14**. 1–46.

Beckman, Jill (1998). *Positional Faithfulness.* Ph.D. dissertation, UMass, Amherst.

Bolozky, Shmuel & Michael Becker (2006). Living Lexicon of Hebrew Nouns. Ms. UMass Amherst.

Casali, Roderic (1998). *Resolving Hiatus.* Garland, New York.

Ernestus, Miriam & R. Harald Baayen (2003). Predicting the Unpredictable: Interpreting Neutralized Segments in Dutch. *Language* **79**. 5–38.

Gaskell, M. Gareth & William D. Marslen-Wilson (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology* **45**. 220–266.

Hayes, Bruce (1995). On what to teach the undergraduates: Some changing orthodoxies in phonological theory. In Ik-Hwan Lee (ed.) *Linguistics in the Morning Calm 3*, Seoul: Hanshin. 59–77.

Hayes, Bruce (2009). *Introductory Phonology*. Wiley-Blackwell.

Inkelas, Sharon, Aylin Kuntay, John Lowe, Orhan Orgun & Ronald Sprouse (2000). Turkish Electronic Living Lexicon (TELL). Website, http://socrates.berkeley.edu:7037/.

Inkelas, Sharon & Cemil Orhan Orgun (1995). Level ordering and economy in the lexical phonology of Turkish. *Language* **71**. 763–793.

Inkelas, Sharon, Cemil Orhan Orgun & Cheryl Zoll (1997). The implications of lexical exceptions for the nature of the grammar. In Iggy Roca (ed.) *Derivations and Constraints in Phonology*, Oxford: Clarendon. 393–418.

Jesney, Karen (2009). Positional faithfulness, non-locality, and the harmonic serialism solution. Proceedings of the 39th Meeting of the North East Linguistic Society.

Kager, René (1999). Surface opacity of metrical structure in Optimality Theory. In Ben Hermans & Marc van Oostendorp (eds.) *The Derivational Residue in Phonological Optimality Theory*, Amsterdam: John Benjamins. 207–245.

Lees, Robert (1961). *The Phonology of Modern Standard Turkish.* Bloomington: Indiana University Press.

Luce, Paul A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception and Psychophysics* **39**. 155–158.

Luce, Paul A. & D.B. Pisoni (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear and Hearing* **19**. 1–36.

Marslen-Wilson, William D. & Alan Welsh (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* **10**. 29–63.

Nicolae, Andreea & Andrew Nevins (before lunch). The phonetics and phonology of fricative non-neutralization in Turkish. Talk given at NELS 40.

Pater, Joe (2006). The locus of exceptionality: Morpheme-specific phonology as constraint indexation. In Leah Bateman & Adam Werle (eds.) *UMOP: Papers in Optimality Theory III*, Amherst, MA: GLSA. 1–36.

Pater, Joe (2009). Morpheme-specific phonology: Constraint indexation and inconsistency resolution. In Steve Parker (ed.) *Phonological Argumentation: Essays on Evidence and Motivation*, Equinox. 1–33.

Pycha, Anne, Sharon Inkelas & Ronald Sprouse (2007). Morphophonemics and the Lexicon: A Case Study from Turkish. In M. J. Solé, P. Beddor & M. Ohala (eds.) *Experimental Approaches to Phonology*, Oxford University Press. 369–385.

Sharoff, Serge (2005). Methods and tools for development of the Russian Reference Corpus. In D. Archer, A. Wilson & P. Rayson (eds.) *Corpus Linguistics Around the World*, Rodopi. 167–180. http://www.comp.leeds.ac.uk/ssharoff/frqlist/frqlist-en.html.

Steriade, Donca (1994). Positional neutralization. Ms., UCLA.

Ussishkin, Adam & Andrew Wedel (to appear). Lexical access, effective contrast and patterns in the lexicon. In Paul Boersma & Silke Hamann (eds.) *Perception in Phonology*, Mouton de Gruyter.

Wedel, Andrew (2002). Phonological alternation, lexical neighborhood density and markedness in processing. Handout from presentation at LabPhon 8, Yale University.