Phonological Alternation, Lexical Neighborhood Density and Markedness in Processing

Andrew Wedel
UC Santa Cruz/University of Arizona
wedel@ling.ucsc.edu

LabPhon8
Yale University
June 28-July 1, 2002

Abstract

Phonological alternation results in non-identical surface forms of morphemes. For final devoicing alternations in Turkish and Catalan, and vowel reduction alternation in Catalan, stems with fewer than about 4 phonemes are much less likely to contain alternating phonemes than larger stems. This suggests that phonological alternation may be crosslinguistically marked in small stems.

Small morphemes are more likely to be found in dense lexical neighborhoods. Lexical entries with many near neighbors have been shown to be processed less rapidly and accurately than those with few neighbors. Under several models of lexicon structure, phonological alternation may be expected to exacerbate the negative influence of high neighborhood density on processing, with the result that alternation should be relatively marked in dense neighborhoods. This hypothesis is supported by the finding that small morphemes that *do* show alternation inhabit less dense neighborhoods on average than those that do not.

Evidence is presented that the preference for non-alternation in small stems has been grammaticalized in Turkish. However, the grammar does not appear to make reference to neighborhood density *per se*, but rather syllable number—which is shown to be a good, but not perfect, predictor of neighborhood density.

1. Phonological alternation is not evenly distributed across the lexicon

- Phonological Alternation: Phonological context dependent change in the surface form of a phoneme.

    Example: Final Obstruent Devoicing in Catalan

    /vɛrd/ 'green' →   [vɛrt] 'green-MASC'   ~   [vɛrd-a] 'green-FEM'

For four alternations studied in three languages…

- Final Devoicing in Catalan
- Final Devoicing in Turkish
- Vowel Reduction in Catalan
- Palatalization in Czech (not discussed here further for lack of space)

…small morphemes are significantly less likely to contain alternating phonemes than large morphemes.

2. The Data

2.1 Distribution of Final Devoicing Alternation by Stem-size in Catalan

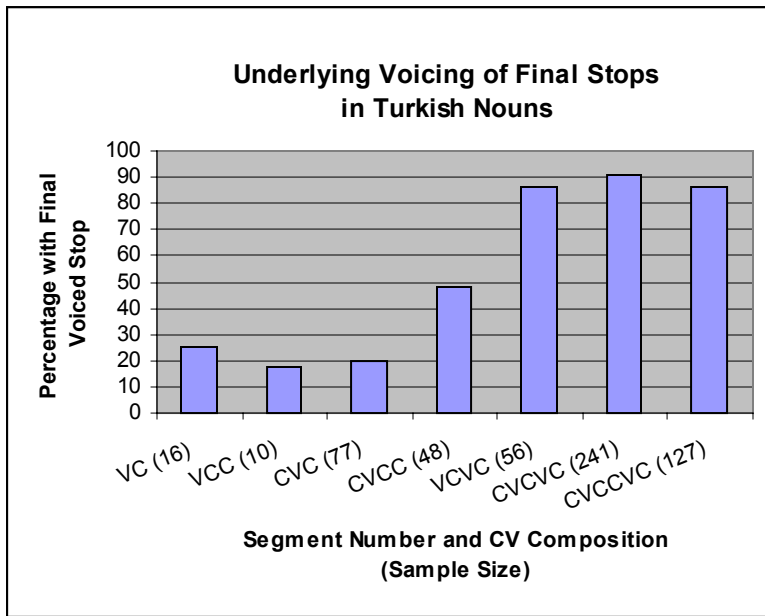Figure 1 (data from Mascaró and Rafel 1990)
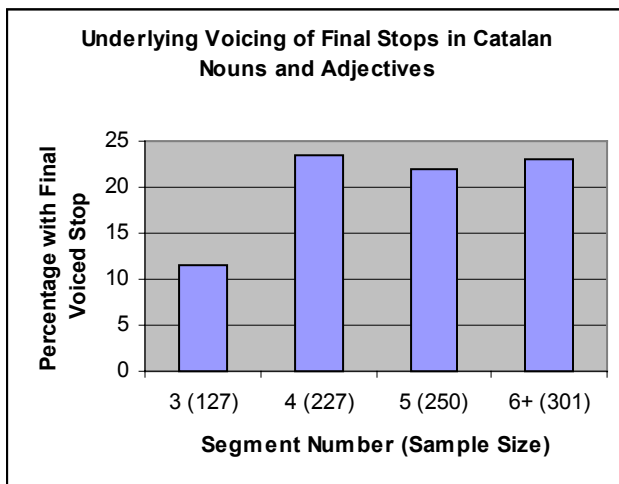
⇒ The 3 segment class is significantly distinct from any other ($p < .001$).

2.2 Distribution of Final Devoicing Alternation in Turkish

a.   /sebep/ 'reason'   → [sebep] 'reason-NOM'   ~   [sebebi] 'reason-ACC'

b.   /ip/ 'string'   → [ip] 'string-NOM'   ~   [ipi] 'string-ACC'

Figure 2 (data from Redhouse Turkish-English Elsözlügü 1997)

Underlying Voicing of Final Stops in Turkish Nouns

Percentage with Final Voiced Stop vs. Segment Number and CV Composition (Sample Size): VC (16), VCC (10), CVC (77), CVCC (48), VCVC (56), CVCVC (241), CVCCVC (127)
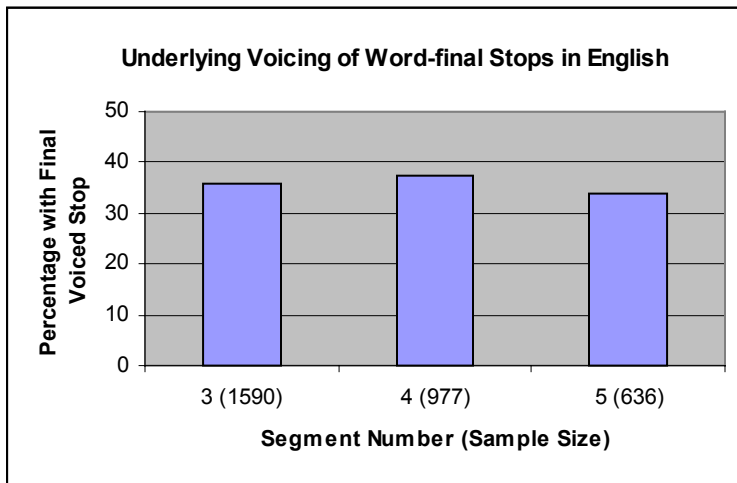
⇒ The three smallest sets class together statistically, as do the three two syllable sets.

⇒ The CVCC set is significantly different from all others.

As a control, we can look at the distribution of underlying voiced and unvoiced stops by stem-size in a language without phonological final devoicing such as English.

Figure 3 (data from http://www.speech.cs.cmu.edu/cgi-bin/cmudict)

Underlying Voicing of Word-final Stops in English

Percentage with Final Voiced Stop vs. Segment Number (Sample Size): 3 (1590), 4 (977), 5 (636)

⇒ The three classes are not statistically distinct

2.3 Distribution of Vowel Reduction Alternation in Catalan
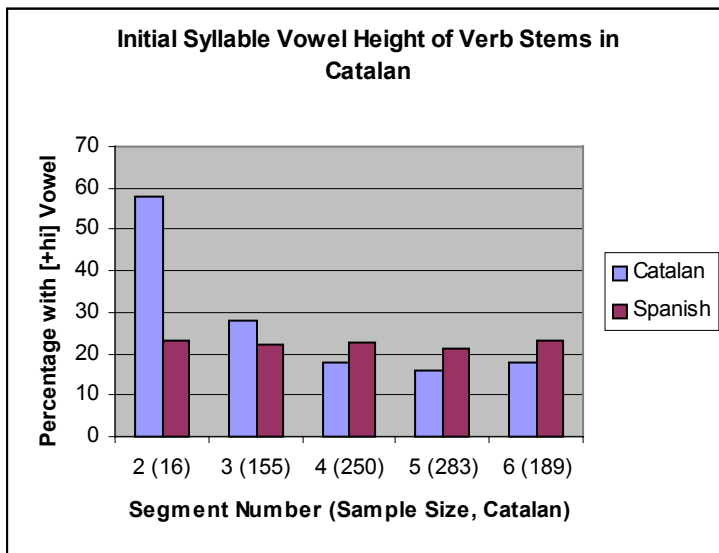
Stress-shift induced vowel reduction in Catalan

+Stress   −Stress

i ⟶ i

e ⇢
ε ⇢ ⟶ ə
a ⇢

u ⟶
o ⇢ ⟶ u
ɔ ⇢
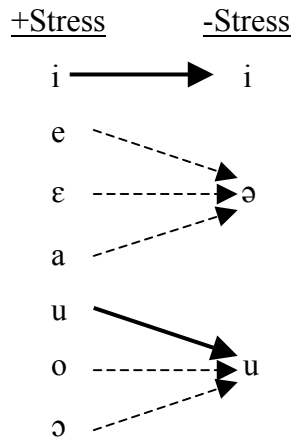
⇒ Two vowels, /i/ and /u/, do
not change their surface
form upon stress shift.

Stress-shift under suffixation in Catalan verbs.

	verb stem		+ 1sg-pres /u/		+ 1sg+pst /í/
a.	/somi/ 'dream'	→	[sumíu]	~	[somií]
b.	/vol/ 'fly'	→	[vólu]	~	[vulí]

Figure 8 (Sabater and Freixinet 1990, The American Heritage Pocket Spanish Dictionary 2001)

Initial Syllable Vowel Height of Verb Stems in Catalan

Y-axis: Percentage with [+hi] Vowel (0 to 70)
X-axis: Segment Number (Sample Size, Catalan): 2 (16), 3 (155), 4 (250), 5 (283), 6 (189)
Legend: Catalan, Spanish

⇒ Decreasing size of Catalan verb stems is
associated with increasing use of non-
alternating vowels. The 2 and 3 segment
size classes are significantly distinct
from one another and all larger size
classes.

⇒ In Spanish, which does not show vowel
reduction, there are no significant
distinctions between any of the size
classes.

3. What might be the source of this effect?
What makes small words different?

3.1 The Neighborhood Density Effect

- Words are recognized more rapidly and accurately if they are in sparse lexical neighborhoods; that is, if they have few phonologically similar neighbors in the lexicon (Goldinger, Luce and Pisoni 1989, Cluff and Luce 1990).

3.2 Small morphemes tend to inhabit dense neighborhoods

Figure 4 (From Frauenfelder, Baayen and Hellwig 1993)

- Density is crosslinguistically inversely correlated with morpheme size (Frauenfelder, Baayen and Hellwig 1993).

⇒ Small words are less contrastive, on average, than larger words .

The mean number of neighbors by root-length in English
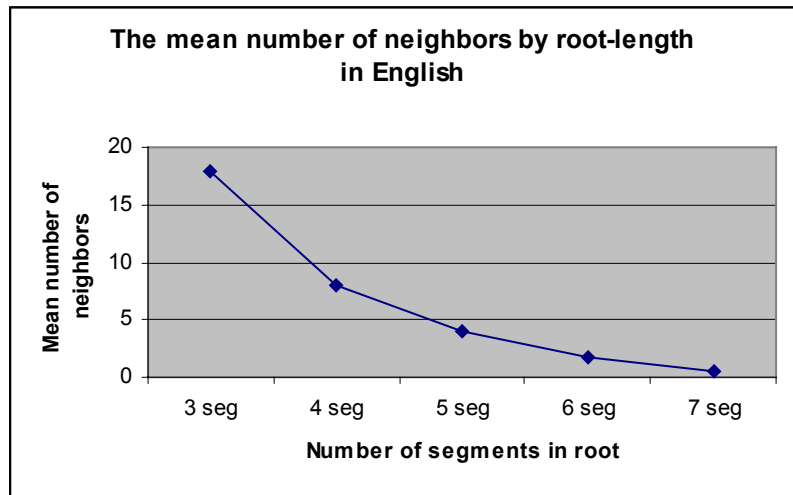
3.4 Alternation and Neighborhood Density

- Under the hypothesis of 'one morpheme--one underlying representation', allomorphy due to phonological alternation is predicted to result in a mismatch between a phonetic string and the corresponding lexical entry, as illustrated for final obstruent-devoicing in Catalan in (i).

(i)

Lexical entry:	/vɛrd/ 'green'
	││││̸
Phonetic form:	[vɛrt] 'green-MASC'

⇒ A mismatch between an underlying representation and a phonetic form is predicted to result in lower activation of the intended correspondent, and, in dense neighborhoods, a greater activation of competitors (cross-activation).

⇒ In contrast, when neighborhood density is lower, alternation is predicted to result in less cross-activation because of greater distance to the nearest lexical neighbors.

- Under the assumption of structured lexical entries (cf. Wunderlich 1996), alternation creates phonologically similar sub-entries within a lexical entry, as illustrated in (ii)

(ii) Possible structured lexical entry for *green* in Catalan

$$\left[\quad \overset{\text{/vɛrD/}}{\underset{\text{/vɛrt/}\swarrow \qquad \searrow \text{/vɛrd-/}}{}} \quad \right]$$
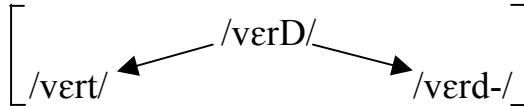
$\Rightarrow$ If complexity within a lexical entry slows processing in any way (for example, if sub-entries compete with one another), alternation may be marked in dense neighborhoods, where lexical access is already inefficient.
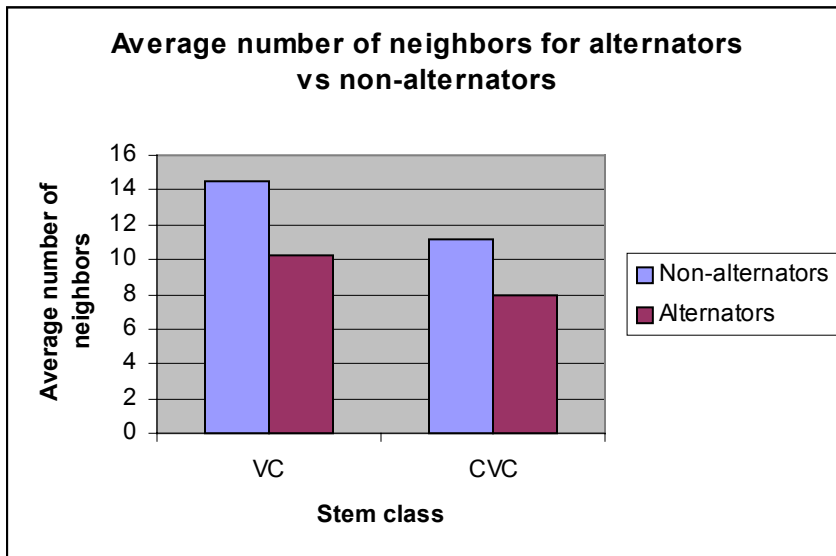
4. Predictions of the hypothesis

- If efficiency of access is the source of the pattern rather than morpheme-size per se, small words that exhibit alternation should inhabit sparser neighborhoods than small words that do not.

4.1 Catalan Final Devoicing

4.2 Turkish Final Devoicing

Figure 14.

Average number of neighbors for alternators vs non-alternators

$\Rightarrow$ Alternators have significantly lower neighborhood densities than non-alternators.

$\Rightarrow$ The data from Catalan and Turkish support the hypothesis that phonological alternation is marked in dense neighborhoods.

- High token frequency of a lexical item mitigates the inhibitory effect of high neighborhood density on access (cf. Dirks et al. 2000). Hence, in high density neighborhoods frequent lexical items should be more likely to alternate than less frequent items.

$\Rightarrow$ This prediction has not yet been tested.

5. Lexical access and the grammar

- In Catalan and Czech (data not shown), the relative avoidance of alternation in small stems appears to be *gradient*: the change in likelihood of alternation with size is proportionally small (see Figs 1 and 3).

- In Turkish the difference is more nearly *categorical*: ~90% of large stems alternate, while only ~20% of small ones do (see Fig. 2).

- This split is in fact grammaticalized in Turkish: foreign names and nonce words, for example, are spontaneously inflected in a way that appears to reproduce the pattern shown in Fig. 2 (AW, ongoing field research, Zimmer and Abbot 1976).

(15)

Alternation in two syllable Turkish stems | No alternation in one syllable Turkish stems

a. [worik] 'Warwick' ~ [wori-in] 'Warwick-GEN'   d. [rik] 'Rick'   ~ [rik-in] 'Rick-GEN'
b. [patrik] 'Patrick'   ~ [patri-in] 'Patrick-GEN'   e. [mit͡ʃ] 'Mitch' ~ [mit͡ʃ-in] 'Mitch-GEN'
c. [demit͡ʃ] (name)   ~ [demid͡ʒ-in] 'name-GEN'   f. [mark] 'Mark' ~ [mark-in] 'Mark-GEN'

⇒ Can neighborhood density relationships be grammaticalized?

- The single syllable, CVCC stem class in Turkish contains significantly fewer alternating members than the two-syllable, VCVC stem class. If the avoidance of alternation in small stems is directly based on neighborhood density, we predict that CVCC stems must have more neighbors than VCVC stems. But this is not the case:
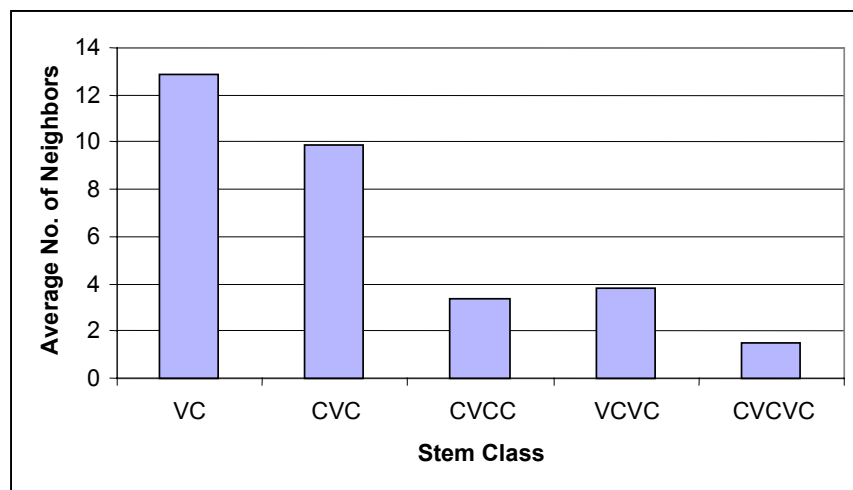
⇒ The CVCC and VCVC stem classes' average neighborhood densities are not statistically distinct.

- In contrast, the likelihood of alternation in four-phoneme Catalan stems shows no relation to syllable number.

- Grammars tend to display relatively simple, quantal behavior (cf. Pierrehumbert 2001, Gordon 2002). Although gradient skewing of the distribution of alternation in the lexicon may be based on inequalities in lexical access efficiency, the grammar may not able to phonologize this pattern directly.

- On this view, the Turkish solution—grammaticalization of alternation on the basis of syllable number—reproduces the pattern expected from Catalan with only a small overshoot, including the CVCC stems unnecessarily in the non-alternating class.

6. Conclusions

- Examination of four phonological alternations in three languages suggests that phonological alternation is more marked in lexical items of about 3 phonemes or fewer.

- The observation that alternating stems in Catalan and Turkish have lower average neighborhood densities relative to non-alternating stems supports the hypothesis that the markedness of alternation in small stems is related to access efficiency.

References

The American Heritage Pocket Spanish Dictionary 2001.Houghton Mifflin Company: New York.

Cluff, M. and Luce, P. A. 1990. Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance*, 16:551-563.

Dirks, D., Takayanagi, S., Moshfegh, A., Noffsinger, P., and Fausti, S. 2001. Examination of the Neighborhood Activation Theory in Normal and Hearing-Impaired Listeners. *Ear and Hearing* 22 (1):1-13.

Frauenfelder, U. H., Baayen, R. H. and Hellwig, F. M. 1993. Neighborhood density and Frequency across Languages and Modalities. *Journal of Memory and Language*, 32: 781-804.

Goldinger, S. D., Luce, P. A. and Pisoni, D. B. 1989. Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, 28: 501-518.

Gordon, M. 2002. A phonetically-driven account of syllable weight, *Language* 78, 51-80.

Luce, P. A. and Pisoni, D. B. 1998. Recognizing Spoken Words: the neighborhood activation model. *Ear and Hearing* 19:1-36.

Mascaró, J., Rafel, J. 1990. *Diccionari Catalá Invers amb Informació Morfológica*. Liberigraf, Barcelona, Spain.

Pierrehumbert, J. 2001. Why phonological constraints are so coarse-grained. In J. McQueen and A. Cutler (Eds) SWAP special issue, Language and Cognitive Processes 16 5/6, 691-698.

Redhouse Turkish-English Elsözlügü 1997. Redhouse Yayinevi, Istanbul, Turkey.

Sabater M. and Freixinet J. 1990. *Hippocrene Concise Dictionary, Catalan-English/English-Catalan*. Hippocrene Books: New York.

Wunderlich, D. 1996. Minimalist morphology: the role of paradigms. In: Geert Booij & Jaap van Marle (eds.) *Yearbook of Morphology 1995*, 93-114. Dordrecht: Kluwer.

Zimmer, K. and Abbot, B. 1976. The k/0 Alternation in Turkish: Some Experimental Evidence for Its Productivity. *Journal of Psycholinguistic Research*. 7:35-46.