Penn Linguistics Colloquium 31 23 February 2007

The Real Effect of Word Frequency on Phonetic Variation

Aaron J. Dinkin, University of Pennsylvania
dinkin@babel.ling.upenn.edu

Exemplar Theory or **Usage-Based Phonology** (e.g., Bybee 1999, 2000; Pierrehumbert 2002): Phonological knowledge consists of memorized phonetic tokens of individual lexical items.
Therefore: “high frequency words tend to lead Neogrammarian sound changes” (Pierrehumbert 2002); Bybee (2000) cites several examples.
However: Labov (2003) finds no effect of word frequency on fronting of back upgliding vowels in American English.
So: what *is* the relationship, if any, between word frequency and sound change?

Subject of this study: the short vowels /i e æ ʌ u/ of Northern American English
Phonetic data: F2 measurements from the Telsur survey (Labov et al. 2006)
Frequency data: derived from Brown Corpus of Standard American English (data from http://www.edict.com.hk/textanalyser/wordlists.htm)
Northern Cities Vowel Shift: increases F2 of /æ/, reduces F2 of /i e ʌ/.

Words coded for frequency as follows:
marked top 5000, 500, and 200 words in frequency rank in the Brown Corpus;
within top 5000, coded exact number of occurrences in corpus;
within top 500, also coded for “function word” status:
prepositions, conjunctions, determiners, verbal auxiliaries, *have*, *be*, etc.

Methodology: Multiple-regression analyses on F2 of each short vowel, against frequency and phonetic variables.

Results:

variable	coefficient	variable	coefficient
onset cluster	–489 Hz	labial onset	–119 Hz
liquid onset	–423 Hz	complex coda	–84 Hz
apical onset	–167 Hz	apical coda	–71 Hz
palatal onset	–151 Hz	/l/ coda	–69 Hz
nasal coda	+136 Hz	polysyllable	–66 Hz
labial coda	–122 Hz	**Top 5000**	**–57 Hz**

Table 1: effects of Brown frequency and phonetic variables on /i/ in the North.
$p < .01\%$ $n = 2492$ constant = 2147 Hz $r^2 = 32\%$

variable	coefficient	variable	coefficient
apical coda	–353 Hz	stop coda	+127 Hz
labial coda	–324 Hz	liquid onset	–125 Hz
labdent. coda	–279 Hz	complex coda	–96 Hz
intdent. coda	–271 Hz	polysyllable	–83 Hz
nasal coda	+218 Hz	/l/ coda	–67 Hz
palatal coda	–216 Hz	voiced coda	+60 Hz
velar coda	–204 Hz	apical onset	–39 Hz
onset cluster	–162 Hz	**Top 5000**	**–33 Hz**

Table 2: effects of Brown frequency and phonetic variables on /e/ in the North.
$p < .01\%$ $n = 2913$ constant = 2034 Hz $r^2 = 39\%$

variable	coefficient	variable	coefficient
nasal coda	+275 Hz	stop coda	+94 Hz
velar coda	–207 Hz	labdent. coda	–79 Hz
apical coda	–152 Hz	voiced coda	+75 Hz
liquid onset	–134 Hz	apical onset	–63 Hz
onset cluser	–123 Hz	complex coda	+42 Hz
labial coda	–123 Hz	**Top 5000**	**–23 Hz**
polysyllable	–99 Hz		

Table 3: effects of Brown frequency and phonetic variables on /æ/ in the North.
$p \leq .01\%$ $n = 5091$ constant = 2058 Hz $r^2 = 30\%$

variable	coefficient	variable	coefficient
/l/ coda	–287 Hz	palatal coda	+106 Hz
liquid onset	–147 Hz	polysyllable	+49 Hz
labial onset	–124 Hz	**Top 5000**	**+36 Hz**
onset cluser	–111 Hz	voiced coda	–32 Hz
apical coda	+110 Hz		

Table 4: effects of Brown frequency and phonetic variables on /ʌ/ in the North.
$p \leq .02\%$ $n = 1794$ constant = 1372 Hz $r^2 = 37\%$

variable	coefficient	variable	coefficient
apical onset	+253 Hz	**Top 200**	**+145 Hz**
palatal onset	+237 Hz	velar onset	+141 Hz
/l/ onset	–184 Hz	labial onset	–112 Hz

Table 5: effects of Brown frequency and phonetic variables on /u/ in the North.
$p < .01\%$ $n = 731$ constant = 1267 Hz $r^2 = 68\%$

More frequent /i/ and /e/ words are **ahead** of the NCVS, but more frequent /æ/ and /ʌ/ words **trail** the NCVS!

Look at it another way: front vowels in more frequent words are backed; short vowels are fronted: **short vowels in frequent words are more centralized**.
(Function word / lexical word status has no statistical effect.)

This **remains true when not restricted to the North**:

vowel	/i/	/e/	/æ/	/ʌ/	/u/
effect of frequency	–61 Hz	–28 Hz	–18 Hz	+44 Hz	+80 Hz
n	10,182	11,466	17,147	6939	3197

Table 6: effects of Brown frequency on short vowel F2 in the whole Telsur corpus.
$p < .01\%$ in all cases; frequency variable is Top 200 for /u/, Top 5000 otherwise.

So NCVS is not subject to frequency effects, but **degree of centralization of short vowels in general** is.

Phillips (1984): "Changes affecting the most frequent words first typically involve either vowel reduction and eventual deletion or assimilation…. The thing to note about these sound changes is that they all have their basis in the physiology of speech."
Construe **lenition** as referring broadly to **reduction of articulatory effort**.
Then Phillips's principle is: **frequent words lead sound changes of lenition** (broadly construed), not sound changes in general.

- Most of the examples cited by Bybee (2000) fit this description.
- Fronting of /uw ow aw/ is not lenition, and frequent words don't lead (Labov 2003).
- NCVS is not lenition, and frequent words don't lead.

Generalize the principle to stable variation as well as changes in progress:
Frequent words are more subject to lenition than less frequent words.

- Bybee (2002): frequent words are favored for English *t*/*d*-deletion.
- Abramowicz (2006): no frequency effect for English (ing), which is not lenition.

"Phillips's principle" explains the findings of this paper:

- Short-vowel centralization reduces articulatory effort;
- therefore, frequent words with short vowels will be more centralized.
- This applies whether or not there is such a sound change in progress.

The real effect of word frequency on phonetic variation is Phillips's principle:
not **change** per se but **lenition** is favored by more frequent words.

Caveats and statistical anomalies:

- Gradient frequency variable in some cases shows a significant but minuscule effect.
- Some of the phonetic variables have effects that are bizarre and implausible-seeming.

But: the effect of Top 5000 (or Top 200 for /u/) is consistent and everywhere significant to $p \leq .01\%$.
Given the erraticness of other frequency measurements, perhaps the significant results of Top 5000 indicate something subtler going on.

References:

Abramowicz, Lukasz (2006). "Sociolinguistics Meets Exemplar Theory: Frequency and Recency Effects in (ing)". Talk presented at NWAV 35, Columbus, Ohio.

Bybee, Joan (1999). "Usage-Based Phonology". In Michael Darnell, Edith Moravcsik, Frederick Newmeyer, Michael Noonan, & Kathleen Wheatley (eds.), *Functionalism and Formalism in Linguistics, Volume I: General Papers,* 211–242. John Benjamins, Amsterdam.

Bybee, Joan (2000). "Lexicalization of Sound Change and Alternating Environments". In Michael B. Broe and Janet M. Pierrehumbert (eds.), *Laboratory Phonology V: Acquisition and the Lexicon*: 250–268. Cambridge University Press, Cambridge.

Bybee, Joan (2002). "Word Frequency and Context of Use in the Lexical Diffusion of Phonetically Conditioned Sound Change". *Language Variation and Change* 14:261–290.

Labov, William (2003). "Words Floating on the Surface of Sound Change". Talk presented at NWAVE 32, Philadelphia.

Labov, William, Sharon Ash, & Charles Boberg (2006). *Atlas of North American English*. Mouton/de Gruyter, Berlin.

Phillips, Betty (1984). "Word Frequency and the Actuation of Sound Change". *Language* 60:320–342.

Pierrehumbert, Janet (2002). "Word-Specific Phonetics". In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology VII*: 101–140. Mouton/de Gruyter, Berlin.