Andries W. Coetzee   University of Michigan

Lexical frequency and variation

The problem. Variable phonological processes are influenced by the same grammatical factors as categorical processes. In English, t/d variably deletes from word-final clusters – cf. (1). Table 1 (next page) shows that the frequency of deletion is at least partially determined by phonological context. Several formal models have been developed over the past decade or so that can account fairly well for this grammatical influence on variable processes (Anttila 1997; Boersma & Hayes 2001; Coetzee 2006; etc.).

(1)   *Pre-C context*   *Pre-V context*   *Pre-Pause context*
               west bank ~ wes bank   west end ~ wes end   west ~ wes

However, usage frequency also influences the application frequency of a variable process. t/d-deletion is more likely in more frequent words – *west* and *vest* are very similar, but *west* is more likely to undergo t/d-deletion, corresponding to its higher usage frequency (Table 2). Current models of variation are all strictly grammatical, and cannot account for this frequency influence. I propose a model that allows grammar and lexical frequency to co-determine the application frequency of a variable process.

(2)   *PRE-C: No word-final [-Ct/d] before a C-initial word.
               *PRE-V: No word-final [-Ct/d] before a V-initial word.
               *PRE-##:   No word-final [-Ct/d] before a pause.
               Chicano English ranking:
               MAX-L1 ≫ *PRE-C ≫ MAX-L2 ≫ *PRE-V ≫ MAX-L3 ≫ PRE-## ≫ MAX-L4.

The proposal. (i) *Variable lexical indexation*. I assume that faithfulness constraints can be indexed to lexical classes, and that these constraints are interspersed between the markedness constraints, as shown in (2). An indexed constraint only evaluates words that share its indexation. The novel proposal here is that words do not have to belong to one lexical class exclusively. Since a word can vary its affiliation, it can be evaluated by different indexed constraints on different occasions, resulting in variation. Assume that /west/ can be assigned to L1, L2, L3, or L4. The faithful candidate of /west bank/ violates *PRE-C, and the deletion candidate one of the indexed MAX-constraints, depending on /west/'s lexical class affiliation. If it is assigned to L1, the faithful candidate is optimal, but any other indexation results in deletion. Pre-vocalically (/west end/), the faithful candidate violates *PRE-V. Now two indexations result in preservation (L1, L2), and two in deletion (L3, L4) (cf. tableau below). Pre-pausally only an L4-affiliation results in deletion. The grammatical influence on variation is hence captured – deletion is observed under 3/4 indexations pre-consonantally, 2/4 pre-vocalically position, and only 1/4 pre-pausally.

(ii) *Frequency and lexical class affiliation*. In the current model, the lexical class of a word is determined at each evaluation occasion. I propose that this process is influenced by the word's usage frequency. Every word is stored with its own probability distribution function. These functions range from 0 to 1, with the range divided into regions corresponding to the lexical classes. In the example here, values from 0 to .25 correspond to L1, .25 to .5 to L2, etc. Every time a word is submitted to the grammar, a value is chosen randomly from its probability distribution to determine its lexical class affiliation for that evaluation occasion. If a value under .25 is selected it will be evaluated by MAX-L1, etc.

The shape of a word's distribution function is determined by its frequency. Frequent words have left-skewed distributions so that their distribution mass is concentrated at the higher end. A frequent word will hence more likely select a value resulting in it being classified as L3 or L4 than L1 or L2. Consequently, a frequent word is more likely to be protected by low ranking faithfulness, and hence to undergo deletion. Infrequent words have right-skewed distributions. By similar reasoning, they are more likely to be assigned to L1 or L2, and hence to resist deletion (cf. figure below). Since usage frequency determines the shape of the distribution functions, lexical frequency gets to influence the likelihood of deletion.

Conclusions. There is mounting evidence that lexical factors (usage frequency) play a role in phonology. An adequate model of phonology must include a mechanism through which such lexical factors can contribute to phonological performance. Lexically indexed constraints allow lexical information an indirect entrance into the grammar, which I exploit here to allow grammar and the lexicon to co-determine the frequency with which variable processes apply.
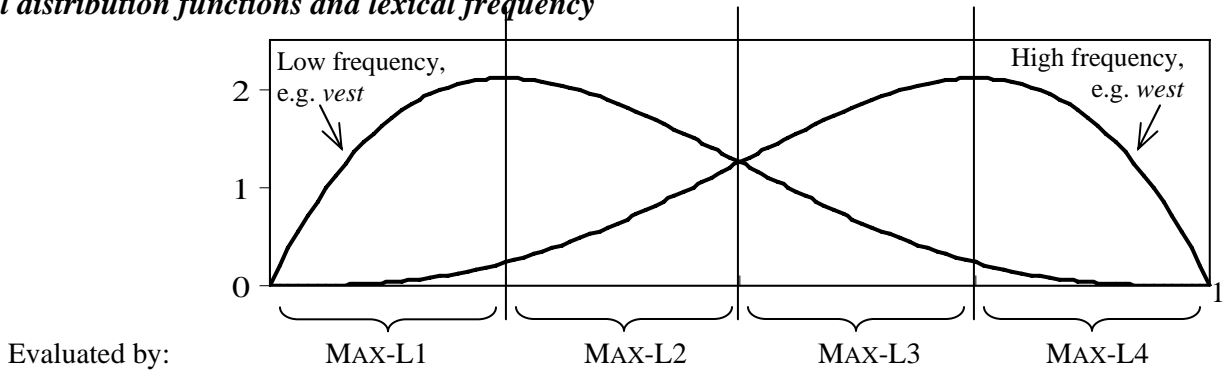
Table 1: t/d-deletion in Chicano English (Santa Ana 1991:76)

	Pre-C	*Pre-V*	*Pre-Pause*
n	3,693	1,574	1,024
% deleted	62	45	37

Table 2: Influence of usage frequency on t/d-deletion in Chicano English (Bybee 2000:70)

Frequency per million	*Deletion*	*Retention*	*% Deletion*
> 35	898	752	54.4%
< 35	137	262	34.4%

Lexical distribution functions and lexical frequency

Example derivations in Pre-V context

	MAX-L1	*PRE-C	MAX-L2	*PRE-V	MAX-L3	*PRE-##	MAX-L4
/west^{L1} end/ ☞ west end				*			
wes end	*!						
/west^{L2} end/ ☞ west end				*			
wes end			*!				
/west^{L3} end/ west end				*!			
☞ wes end					*		
/west^{L4} end/ west end				*!			
☞ wes end							*

References

Anttila, Arto. 1997. Deriving variation from grammar. In Frans Hinskens, Roeland van Hout and Leo Wetzels, eds. *Variation, Change and Phonological Theory,* Amsterdam: John Benjamins. p. 35-68.

Boersma, Paul and Bruce Hayes. 2000. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry*, 32: 45-86.

Bybee, Joan L. 2000. The phonology of the lexicon: evidence from lexical diffusion. In Michael Barlow and Suzanne Kemmer, eds. *Usage-Based Models of Language*. Stanford: CSLI Publications. p. 65-85.

Coetzee, Andries W. 2006. Variation as accessing "non-optimal" candidates. *Phonology*, 23:337-385.

Santa Ana, Otto. 1991. *Phonetic Simplification Processes in the English of the Barrio: A Cross-Generational Sociolinguistic Study of the Chicanos of Los Angeles*. Ph.D. Dissertation, University of Pennsylvania.