

The Correlation of Markedness and Frequency: Emergent or Innate?

Andrew Martin, UCLA

A growing number of studies in phonological theory address the hypothesis that some typological asymmetries traditionally ascribed to principles of Universal Grammar may in fact be emergent properties resulting from local interactions, without the need for any representation of markedness in the synchronic grammar. Recent work in this vein has shown that properties of vowel systems (de Boer 2001), syllable structure preferences (Redford et al. 2001), and strict constraint domination (Wedel 2004) can all evolve independently in computer simulations in the absence of *a priori* principles of UG.

In this talk I will test the empirical predictions made by this emergentist hypothesis on an actual case of historical change. Words beginning with /b/ in Classical Latin were less common than words beginning with other voiced stops, due to a sound change that occurred in the transition from Proto-Indo-European to Latin. By the time Latin had evolved into Modern French, this was no longer the case—/b/-initial words now outnumber /d/-initial words (Boersma 1997; Figure 1).¹ The higher frequency of /b/-initial words in French is to be expected, given that /b/ is considered to be the least-marked voiced stop on both typological (Hayes and Steriade 2004) and phonetic (Ohala & Riordan 1979) grounds. The question is, how did the lexicon of Latin/French change from an unnatural distribution, where unmarked /b/ was infrequent, to a more natural distribution, in which /b/ is frequent? I will compare two possible theories.

A traditional grammatical markedness account of this fact, which I will call the *preference theory*, would make reference to universal principles of markedness known to native speakers—these speakers know that /b/ is less marked than /d/, and thus prefer to retain (or borrow, or coin) words that start with /b/ over words that start with /d/. Over time, this gradient preference would lead to the present statistical bias in French. The *emergence theory* (as articulated in Blevins 2004 and Wedel 2004), on the other hand, explains the same facts as the gradual accumulation of learners' errors. If, for example, initial /d/ is misperceived by learners as /t/ more often than initial /b/ is misperceived as /p/, this would eventually lead to the same bias towards /b/-initial words (see Wedel 2004 for details of a simulation in which statistical biases of this type can emerge from biases in production and perception).

The two theories make different empirical predictions—the emergence theory, based as it is on the occurrence of errors in transmission, predicts that changes within individual words are responsible for the statistical change in the lexicon. Thus, some /d/-initial words in Latin should end up with different initial segments in French. The preference theory, however, does not predict this—as long as more /b/-initial words enter the language than /d/-initial words, no change in individual words is necessary to result in the French statistical bias.

In order to assess these predictions, I used a Classical Latin word list and, for each /b/- and /d/-initial word, determined whether the word had a descendent in Modern French. Figure 2 shows that overall, few Latin words survived; of those that did, /b/-initial words were more likely to be retained into French than /d/-initial words. Furthermore, only two of these words (out of 80) show evidence of a change in the initial segment (*deorsum* > *jusant* and *diurnum* > *jour*). Figure 3, which tabulates the etymologies of French /b/- and /d/-initial words, shows that /b/-initial words have also been more frequently borrowed and morphologically derived in French than /d/-initial words—none of these words shows a change in initial segment from its source. It is thus highly unlikely that the bias in French is the result of accumulated misperceptions. The preference theory better explains the data—over time, speakers have favored the creation or retention of words with the less-marked /b/, a process which is plausibly linked to the tendency of children to avoid using words with marked sounds (Leonard et al. 1981, Schwartz and Leonard 1982). Furthermore, this preference for /b/ cannot be learned, because the Latin lexicon did not favor /b/-initial words, suggesting that at least some markedness preferences are due to *a priori* knowledge on the part of speakers.

¹ These figures do not include prefixed words, which skew the numbers towards /d/ due to the highly productive prefixes *dé-* and *des-* in French.

Figure 1. Voiced-stop-initial words in Latin and French (no prefixed forms)²

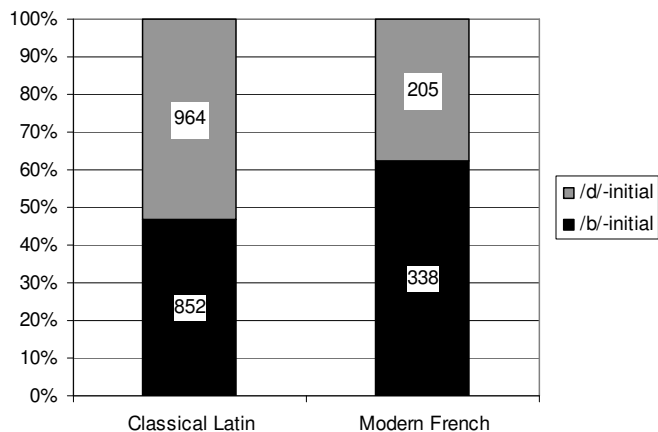
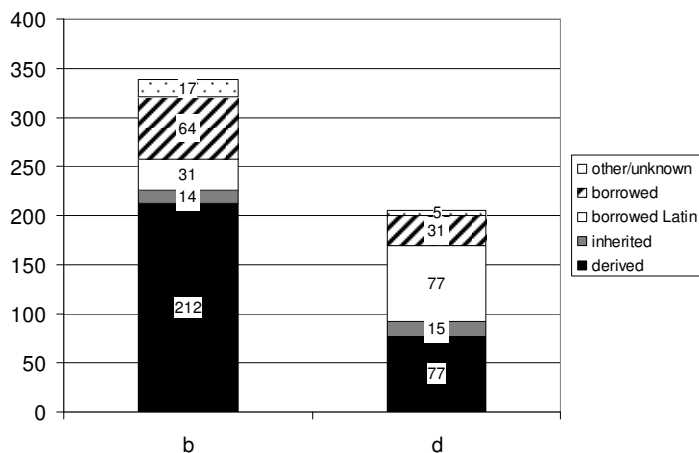


Figure 2. Latin word survival rates

	% survived
/b/-initial	5.4% (46/852)
/d/-initial	3.5% (34/964)

($p < .05$ by Fisher's Exact Test)

Figure 3. Modern French etymologies (no prefixed forms)



References

- Blevins, Juliette. 2004. *Evolutionary Phonology*. Cambridge: Cambridge University Press.
- de Boer, Bart. 2001. *The origins of vowel systems*. Oxford: Oxford University Press.
- Boersma, Paul. 1997. Sound change in functional phonology. Unpublished manuscript, University of Amsterdam, The Netherlands.
- Hayes, Bruce and Donca Steriade. 2004. The Phonetic Bases of Phonological Markedness. In Hayes, B., Kirchner, R. and Steriade, D. (eds.) *Phonetically Based Phonology*. Cambridge University Press. 1-33.
- Leonard, L.B., R.G. Schwartz, B. Morris and K. Chapman. 1981. Factors influencing early lexical acquisition: lexical orientation and phonological composition. *Child Development* 52: 882-887.
- Ohala, John and Carol Riordan. 1979. Passive vocal tract enlargement during voiced stops. In J. J. Wolf and D. H. Klatt (eds.), *Speech Communication Papers*, New York: Acoustical Society of America. S. 89-92.
- Redford, M.A., Chen, C. C. and Miikkulainen, R. 2001. Constrained Emergence of Universals and Variation in Syllable Systems. *Language and Speech* 44: 27-56.
- Schwartz, R.G. and L.B. Leonard. 1982. Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language* 9: 319-336.
- Wedel, Andrew. 2004. Self-organization and categorical behavior in phonology. PhD. Dissertation. UC Santa Cruz.

² Data sources:

Latin: 53,000-word lemmatized word list derived from a 2.1-million-token corpus of Classical Latin texts (available at the Perseus Digital Library: <http://www.perseus.org>). The data represents every word beginning with *b* or *d*.

French: *Le Trésor de la Langue Française* (<http://atilf.atilf.fr/tlf.htm>). The data represents every tenth word from the *b* and *d* sections of the dictionary.