Words influence the evolution of sound category contrast:
Evidence from simulation, corpus and laboratory studies

Andrew Wedel
University of Arizona

I am interested in processes influencing maintenance and loss of contrast between categories that are primarily behaviorally-defined. On the one hand, many categorial distinctions are supported by perceptually stable facts-about-the-world, as in the differences between the categories of water and air. At the other extreme however, we find categories, like sounds, that seem to function behaviorally through the very fact of their difference. For example, there is nothing particularly natural about a given boundary between adjacent vowels, nor any significant perceptual discontinuity that would by itself support multiple categories across the vowel space. Why doesn’t random noise in acquisition and usage rapidly erode these categorial distinctions? Instead, even though the phonetic properties that map to a particular category can shift over time, the abstract system of sound distinctions in a language often remains quite stable through change, as in the case of chain-shifts (Hock and Joseph 1996).

A long-standing intuition holds that the greater the contribution a particular sound category makes to overall word contrast, the less likely it is to be lost over time (Martinet 1955, Hockett 1955). However, it has been notoriously difficult to find satisfactory tests of this hypothesis within natural language data (King 1967, Surendran & Niyogi 2006). Working instead with computational simulations of toy models, I have previously shown that predictions of this ‘functional load’ hypothesis can be successfully modeled if we assume rich memory and multiple interacting levels of analysis, in conjunction with a selectional bias at some level favoring differentiation among word-pronunciations (Wedel 2004, Blevins and Wedel 2009). Here I will use computational simulations of this model to show how pressure for differentiation at the word level interacts with pressure for consolidation at the sound level to produce a compositional, particulate system (Abler 1989) of the sort we find in language.

I have recently extended this research program with two complementary methodologies to test this model: statistical analysis of attested sound mergers in relation to distributional and frequency data from corpora, and laboratory studies of artificial lexicon learning. Results from both methods are strongly consistent a role for local word contrast in the trajectory of change in the global system of sound contrasts. The functional load hypothesis is nearly a century-old, but this represents -- to my knowledge -- the first statistically significant evidence in its favor.

Using results from this research project as an example, I will in addition make the argument that corpus-based, laboratory and simulation studies together provide complementary types of evidence that together allow linguists to more effectively test hypotheses. Analysis of corpus data is closest to the real object of study, yet indirect and poorly controlled. Computational simulation studies represent the inverse: fully manipulable and transparent, but with an abstract and highly simplified relationship to real language. Laboratory studies occupy a third position, providing evidence about linguistic behaviors in usage that allow direct exploration of hypotheses generated from both natural language data and computational studies.