Does contrast play a role in the evolution of phonological categories? Results from three complementary methodological approaches

Andrew Wedel
University of Arizona

I am interested in processes influencing maintenance and loss of contrast between categories that are primarily behaviorally-defined. 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 phonemes, 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 system of phonemic distinctions in a language often remain quite stable through change, as in the case of chain-shifts (reviewed in Hock and Joseph 1996).

A long-standing intuition holds that the greater the contribution a particular phonemic category makes to overall *lexical* 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 lexical memory (Wedel 2004, Blevins and Wedel 2009).

The abstract properties required by this rich-memory model of the lexicon are:

1. Storage of some degree of non-contrastive detail of experienced tokens (reviewed in Johnston 1997) at multiple levels of analysis (Bybee 2002), here modeled through an exemplar-based computational architecture (Pierrehumbert 2001, 2002; see also Wedel 2004, 2006, 2007, Blevins and Wedel 2009);
2. Feedback between perception and production behavior (e.g., Goldinger 2000, Oudeyer 2002);
3. A bias in production and/or perception toward local category centers (Kuhl 1991, Goldstone et al. 2001).

I have recently extended this research program with two complementary methodologies to examine the relationship between lexical contrast and sound change: statistical analysis of attested contrast mergers in relation to distributional and frequency data from corpora, and laboratory studies of artificial lexicon learning. Initial results from both methods are consistent a role for local lexical contrast in the trajectory of change in the global system of contrasts.

In addition to reporting current results from this research project, I will argue that corpus-based, laboratory and simulation studies together provide complementary types of evidence that together allow us 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.