Learning vowel harmony from real speech in an exemplar-based model

The general level of awareness and degree of popularity of instance-based approaches to phonetics and phonology have steadily risen since their introduction over a decade ago (Goldinger 1996, 1998; Johnson 1997). Part of this increase in visibility derives from the simplicity and elegance with which exemplar-driven accounts incorporate the roles of variation and frequency across a variety of phenomena (Bybee 2001, Pierrehumbert 2002, Gahl 2008). Also relevant is the fact that these models lend themselves particularly well to a simulation/modelling-based methodological approach; their implementation is not particularly onerous (although there are several choices that may be more-or-less hard to motivate, cf. Johnson, 2007) and the increasing availability of sizeable, good quality corpora make a data-driven approach attractive. Such implemented models serve both as existence proofs for and sources of additional hypotheses (Wedel 2004, Oudeyer 2006). Notwithstanding all of this, the literature discussing actual implementations of exemplar-driven models suffers from several deficiencies. While it is no longer true that there is a lack of production models (Pierrehumbert 2001 and Kirchner, Moore & Chen 2010 address this particular lacuna), existing models typically deal with “toy scale” problems, in two senses. On the first hand, the data are insufficiently veridical; the linguistic phenomena addressed are either fictitious, or “sanitized” to the point of being comparatively uninteresting (e.g. diachronic lenition or chain shifts in small—sometimes 2-element—systems), particularly to a “typical” phonologist who is likely to be concerned with productivity and generalization in specific, extant, synchronic processes. On the second hand, the data representation is of insufficient dimensionality (Port 2007): models based on small formant-based datasets fail to take into account, hence to illustrate, the complex set of cross-domain interactions from which exemplar models draw most of their predictive power. The one clear exception to the issues discussed above is (Kirchner, Moore & Chen, 2010), in which the authors model a (fictitious, but realistic) synchronic process of intervocalic spirantization, on the basis of real speech tokens.

The work we present here adds to the tiny body of literature represented by Kirchner et al., in that we present an implemented exemplar-based model which learns a productive, general phonological pattern from tokens of real speech. Our empirical domain is the well-known pattern of Turkish vowel harmony. The existing literature on exemplar-based approaches to vowel harmony is sparse, essentially a small set of posters and a single theoretical paper (Cole 2009) that aims to (i) rectify perceived flaws in generative accounts of vowel harmony with respect to the distribution of attested vowel harmony systems, (ii) show that an evolutionary account of vowel harmony, whereby vowel co-occurrence restrictions derive from the phonologization of vowel-to-vowel coarticulation (cf. Ohala
(1994), provides a better account of the typological distribution of vowel harmony, and (iii) show that vowel harmony can be explained without recourse to features, but that domain effects require the interaction of segment- and morpheme-level units. Although we do not address (i) and (ii) in the work to be presented here, our model indirectly addresses aspects of (iii). Drawing inspiration from Silverman (2006), Johnson (2007) and Lodge (2009), and going further back from Klatt’s (1979) LAFS model, we eschew segmental representations and work directly on word-sized tokens of speech. Thus, we show by existence proof that the patterns of alternation that characterize Turkish vowel harmony can be induced from strictly word-sized exemplars. Trained on an LPC-encoded corpus of spoken Turkish, our model uses (lexical) semantic and grammatical information to build categories within and across which it can generalize in producing or perceiving novel tokens. After training, the model displays a fully general and productive use of synchronic palatal and labial vowel harmony, as shown both by traditional machine learning error and measures and impressionistically from synthesized speech of novel outputs. We conclude by discussing the predictions that exemplar-based models make for the acquisition path, and address the relatedness and convergence between these methods and state-of-the-art methods in the statistical and machine-learning literatures.

Word count: 652

References