Learning Nonlocal Phonology from Local Morphophonological Cues

Maria Gouskova, NYU[1]

Nonlocal phonological interactions such as vowel harmony and consonant dissimilation are usually analyzed by allowing the non-adjacent segments to interact directly—either by placing them on an autosegmental projection (McCarthy 1979) or by positing a special correspondence relation (Rose and Walker 2004). Regardless of how these interactions are analyzed, they present several learnability problems. The number of possible nonlocal interactions is rather large if the learner is completely unbiased, so if the learner searches for every possible interaction, it must be guided by heuristics. As with learning other phonological patterns, the learner must also contend with noisy data—how does the learner distinguish between accidental gaps and restrictions that are phonologically plausible but may nonetheless be violated? How does the learner contend with restrictions that hold of some part of the language but not generally?

In this talk, I demonstrate that in at least some languages, nonlocal interactions can be detected **inductively** in the process of learning local phonotactics. The learning model uses three insights from decades of phonological research: (1) segments that interact with each other nonlocally are part of a natural class (McCarthy 1988, Rose and Walker 2004, and others); (2) nonlocal interactions are easier to observe if the language has simple syllable structure (McCarthy 1989), (3) different phonotactic restrictions apply to segments at morphological boundaries vs. segments inside a morpheme (Trubetzkoy 1939, Chomsky and Halle 1968).

These ideas are implemented in a computational model that builds on Hayes and Wilson's (2008) UCLA Phonotactic Learner. The learning procedure starts with a baseline grammar without projections, which is then analyzed for the presence of certain *segmental trigram* constraints that act as cues to the presence of nonlocal interactions. If the constraints are not morphologically sensitive, the learner builds projections on the basis of the natural classes mentioned in the trigrams, and then searches these projections for constraints that have support in the learning data. If the cues refer to certain morphological information, the learner creates a sublexicon (Gouskova et al. 2015, Becker and Gouskova 2016), which serves as a new baseline for phonotactic learning; if this new baseline turns up trigrams that lead to projection building, those projections are tested in the sublexicon and then again in the whole language.

This learning model is tested in three case studies. First is Quechua, a language with well-studied categorical laryngeal co-occurrence restrictions (Gallagher 2010, 2016). Second is Shona, a language with gradient co-occurrence restrictions on vowel height (Beckman 1997, Hayes and Wilson 2008). Finally, I discuss the case of Aymara, whose laryngeal co-occurrence restrictions hold gradiently in morphemes but are violated in complex words (MacEachern 1997, Hardman 2001). All three languages have relatively simple syllable structure, so the interacting segments occur in trigrams often enough that the learner identifies cues to nonlocal phonological interactions in local phonotactics.

References

Becker, Michael, and Maria Gouskova. 2016. Source-oriented generalizations as grammar inference in Russian vowel deletion. *Linguistic Inquiry* 47:391–425.

[1]This talk reports on joint work with Gillian Gallagher of NYU, supported by NSF BCS-1724753.

Beckman, Jill. 1997. Positional faithfulness, positional neutralization, and Shona vowel harmony. *Phonology* 14:1–46.

Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Gallagher, Gillian. 2010. Perceptual distinctness and long-distance laryngeal restrictions. *Phonology* 27:435–480.

Gallagher, Gillian. 2016. Asymmetries in the representation of categorical phonotactics. *Language* 92:557–590.

Gouskova, Maria, Sofya Kasyanenko, and Luiza Newlin-Łukowicz. 2015. Selectional restrictions as phonotactics over sublexicons. *Lingua* 167:41–81. URL `http://ling.auf.net/lingbuzz/002673`.

Hardman, Martha James. 2001. *Aymara*, volume 35. München: Lincom Europa.

Hayes, Bruce, and Colin Wilson. 2008. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry* 39:379–440.

MacEachern, Margaret. 1997. Laryngeal Cooccurrence Restrictions. Doctoral Dissertation, UCLA, Los Angeles.

McCarthy, John J. 1979. Formal problems in Semitic phonology and morphology. Doctoral Dissertation, MIT. Published by Garland Press, New York, 1985.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 43:84–108.

McCarthy, John J. 1989. Linear order in phonological representation. *Linguistic Inquiry* 20:71–99.

Rose, Sharon, and Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80:475–532.

Trubetzkoy, N. S. 1939. *Grundzüge der Phonologie*. Prague: Travaux du cercle linguistique de Prague 7.