Statistical laryngeal harmony in Oromo: a Maximum Entropy model

Avery Ozburn, University of Toronto

In this work, I examine a novel case of laryngeal co-occurrence restrictions in Oromo, a Cushitic language spoken in Ethiopia. Through statistical morpheme-internal harmony in both ejectivity and voicing, Oromo raises crucial questions about the place of non-categorical patterns in phonology, and about the types of theoretical accounts that can deal with such phenomena.

Oromo has laryngeal contrasts in stops/affricates that vary from a two-way contrast for bilabials (/p', b/), to a three-way contrast for post-alveolars (/tʃ, tʃ', dʒ/) and velars (/k, k', g/), to a four-way contrast for dentals (/t, t', d, ɗ/) (Gamta 1989). The purpose of this research was to analyze harmonic restrictions on the distribution of these consonants in morphemes of shape CV(C)CV. A lexicon analysis tested whether Oromo voiceless stops (plain and ejective) are required to agree in [constricted glottis], as well as whether all stops are required to agree in voicing. Results show that these different categories of Oromo stops are not randomly distributed; chi-square statistics on laryngeal categories of first and second consonants show a very high level of significance ($p<0.00001$ in both cases). Observed over expected (O/E) values given below further show that agreement is highly over-represented. Thus, Oromo has statistical laryngeal harmony, but it is not categorical, since the disagreeing cases do not have O/E of 0.

Ejective harmony			*Voicing harmony*		
	C2 Ejective	*C2 Plain*		*C2 Voiced*	*C2 Voiceless*
C1 Ejective	1.53	0.35	*C1 Voiced*	1.46	0.57
C1 Plain	0.22	2.00	*C1 Voiceless*	0.38	1.58

In addition to this agreement, Oromo has further statistical asymmetries that mirror patterns that are categorical in other languages. In particular, within the disagreeing cases, there is a directionality effect, with greater under-representation of cases where C2 is ejective for ejective harmony and where C2 is voiced for voicing harmony. For the voicing case, there is a notable trend towards significance ($p<0.08$). This result mirrors a categorical regressive effect in languages such as Ngizim, where sequences of voiceless stops followed by voiced stops are forbidden but the opposite order is allowed (Hansson 2001, 2010). This statistical directionality effect, like the overall statistical agreement, requires explanation.

With such results, I consider the larger question of how these directionality effects can be treated within phonological theories of harmony, which often account only for categorical phenomena. If we want to understand how cross-linguistic regressive directionality arises, it is crucial to look at statistical cases like Oromo, since they suggest that such biases may go much deeper than approaches to categorical harmony would suggest. Thus, building on previous treatments of statistical consonant harmony through learning algorithms (e.g. Brown 2008), I consider how such models can capture asymmetries like directionality within a statistical pattern, and how they connect to current understandings of categorical consonant harmony.

More specifically, to account for the Oromo patterns, I applied the Maximum Entropy (Maxent) Model of phonotactics and phonotactic learning (Hayes and Wilson 2008), which has been successfully applied to gradient lexical phenomena, including other harmony systems. Since the primary interest of this project is in long-distance interactions between stops, the tier 'Stop' was used as a projection. Running the algorithm on the Oromo data, the grammar learned the constraints *[-cg][+cg] and *[-voi][+voi] on the stop tier, indicating as expected that there are restrictions on long-distance co-occurrences in ejectivity and voicing between stops. Crucially, however, the model did not learn the constraints in the opposite order: *[+cg][-cg] and *[+voi][-

voi] were not part of the grammar. Thus, the grammar correctly learned the directionality asymmetry. In fact, more than that, the model suggests that the only harmony constraints necessary for Oromo are those enforcing regressive harmony of the marked feature, despite clear evidence above that all disagreeing sequences are under-represented. This result is particularly interesting given that theories previously used to analyze categorical bans on all disagreeing sequences, such as Agreement by Correspondence (ABC) (Hansson 2001, 2010; Rose and Walker 2004), have been required to posit either bidirectional harmony of the marked feature value or regressive harmony of both the marked and unmarked values.

Connecting the learning algorithm to ABC provides fresh insight into issues within this theory. Though the particular constraints enforcing agreement are different, with Ident-CC[F] constraints in ABC versus *[-F][+F] constraints in the Maxent model, it is easy to re-analyze Ident-CC as prohibiting disagreeing sequences. Further, the projection tier for stops in the Maxent analysis mirrors the correspondence constraints used in ABC, in that both are ways to allow consonants to see similar consonants across other segments. As such, the Maxent analysis offers the opportunity to examine the benefits and issues of viewing correspondence as feature projection, and therefore also opens up the possibility of using some of the ideas behind ABC to also analyze statistical cases like Oromo. The advantage of such an account is that Maxent, unlike ABC, is able to provide insight into both statistical and categorical harmony systems. Moreover, this framework makes strong predictions about directionality; it must be coded in the 'agreement' component, as in Rose and Walker (2004), rather than as part of correspondence, as done by Hansson (2001, 2010), because correspondence is accomplished through projection, which cannot be directional. As such, while the Maxent analysis of Oromo seems disconnected from previous accounts of categorical consonant harmony, it in fact fits in well with a re-interpretation of the standard ABC approach, which in turn allows the ideas behind ABC to be extended to account for statistical harmony.

Thus, by examining this novel case of laryngeal harmony, I conclude that Oromo offers an important new avenue to understanding consonant harmony, including the types of attested statistical patterns, the presence of directionality asymmetries within disagreeing cases, and the meaning of 'agreement' and 'correspondence' within a learning algorithm framework.

References:

Brown, Jason Camy. 2008. *Theoretical aspects of Gitksan phonology*. PhD dissertation, University of British Columbia.

Gallagher, Gillian. 2010. Perceptual distinctness and long-distance laryngeal restrictions. *Phonology*, *27*(03): 435-480.

Gamta, Tilahun. 1989. *Oromo-English Dictionary*. AAU Printing Press.

Hansson, Gunnar Ólafur. 2001. *Theoretical and typological issues in consonant harmony.* PhD dissertation, University of California, Berkeley.

Hansson, Gunnar Ólafur. 2010. *Consonant harmony: long-distance interaction in phonology*. Berkeley, CA: University of California Press.

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

Rose, Sharon, and Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80: 475-531.