Scalar similarity and gradient featural representation in Lezgian laryngeal harmony

Huteng Dai
huteng.dai@rutgers.edu
Department of Linguistics and Center for Cognitive Science
Rutgers University

Overview

- Similarity is crucial in evaluating both input-output and surface correspondence in Agreement by Correspondence (Rose and Walker, 2004).
- However, the **representational structure** necessary to compute similarity is often left undefined.
- I encode similarity as a weighted featural lattice in Gradient Harmonic Grammar (Smolensky et al., 2014).
- The gradient representational system correctly predicts patterns of laryngeal consonant co-occurrence in Lezgian, providing a reanalysis of (Ozburn and Kochetov, 2018).

Lezgian

- Co-occurring stops in Lezgian generally agree in LARYNGEAL.
- Yet surprisingly $T^h \leftrightarrow T$ and $T \leftrightarrow D$ (T'=Ejective, T=Voiceless, T^h=Aspirated, D=Voiced) are also overrepresented in the corpus.
- *Over-represented laryngeal co-occurrences in Lezgian O/E>1*

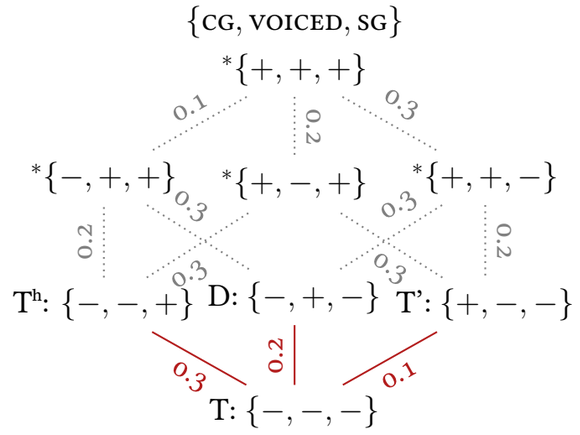
Ejective-Ejective	[q'ats'un]	'get dirty'
Voiceless-Voiceless	[qaqa]	'ready'
Voiced-Voiced	[midad]	'grieve'
Aspirated-Aspirated	[tʃ^{h}ip^h]	'fool'
Aspirated-Voiceless	[k^{h}utsun]	'to flush'
Voiceless-Voiced	[kudaj]	'hot'

Proposal

- The challenge is to define similarity as to allow overrepresented but disharmonic exceptions while penalizing the underrepresented structures.
- Perceptual distance/similarity is defined as the summed weights of unshared features: $\textit{distance}_{\mathbf{w}}(x,y) = \sum_{f\in\mathcal{F}} w_f \cdot \boxed{\delta_f(x,y)}$, ← categorical similarity

$$\delta_f(x,y) = \begin{cases} 0, & \text{if } x \text{ and } y \text{ share the feature } f \\ 1, & \text{else} \end{cases}$$

- CORR$[d \leq k]$ is proposed with respect to the similarity metric: if $\textit{distance} \leq k$, two segments must be in correspondence.
- The weights on features factor into the computation of IDENT-IO and IDENT-CC ($\mathcal{P} = w \times w_f$).



Input	Output	CORR$[d. \leq 0.1]$ $w = 20$	ID-CC[LARYNGEAL] $w = 18$	ID-IO[LARYNGEAL] $w = 5$	CORR$[d. \leq 10]$ $w = 0.1$	$\mathcal{H}$
i. /efsigun/	a. [efs$_x$ig$_y$un]$[d. = 0.2]$ ☜				1	−0.1
	b. [efs$_x$ik$_x$un]$[d. = 0]$			0.2		−1.0
	c. [efs$_x$ik'$_x$un]$[d. = 0.1]$			$0.2 + 0.1$		−1.5
ii. /qafs'un/	a. [q'$_x$afs'$_x$un]$[d. = 0]$ ☜			0.2		−1.0
	b. [q$_x$afs'$_x$un] $[d. = 0.1]$		0.1			−1.8
	c. [q$_x$afs^h$_x$un] $[d. = 0.3]$		0.3	$0.3 + 0.1$		−7.4
	d. [q$_x$afs'$_y$un] $[d. = 0.1]$	1			1	−20.1

Conclusion

- The current study accounts for laryngeal co-occurrence patterns in Lezgian by introducing a gradient featural similarity lattice, over which the intra-featural similarity is evaluated.
- Moreover, the proposed lattice offers a universal structure to link language-specific phonetics to phonological features.

References:

Ozburn, Avery and Kochetov, Alexei (2018). Ejective harmony in lezgian. *Phonology*, 35(3):407–440.

Rose, Sharon and Walker, Rachel (2004). A typology of consonant agreement as correspondence. *Language*, pages 475–531.

Smolensky, Paul, Goldrick, Matthew, and Mathis, Donald (2014). Optimization and quantization in gradient symbol systems: a framework for integrating the continuous and the discrete in cognition. *Cognitive science*, 38(6):1102–1138.

Acknowledgements: I thank Adam McCollum, Adam Jardine, Bruce Tesar, Brian Pinsky, Robin Karlin, Chaoyi Chen, Sreekar Raghotham, and audiences at RULing 2019, for their comments and insights.