In search for empirical evidence for optimal sonority distance

Ruihua Yin
University of Queensland

1. Introduction

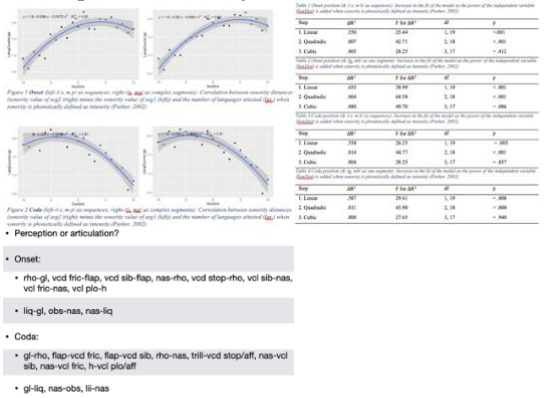
A long-term underlying assumption in relation to the sonority distance between two consonant clusters is that clusters with large sonority distances are preferred by languages, as captured in the unmarked status of onset obstruent-glide (OG) or coda glide-obstruent (GO) clusters. The unmarked status of such clusters is theorised in terms of sonority distance constraints on consonant clusters. Cross-linguistically, it is long assumed consonant clusters with largest sonority distance like OG-/-GO are regarded most frequently attested compared with relatively unmarked clusters like obstruent-liquid (OL-), obstruent-nasal (ON-), nasal-obstruent (-NO), and liquid-obstruent (-LO) etc. This cross-linguistic preference towards high sonority distance has been often observed by [1], [2], and [3], etc., and is also put forward as the Sonority Dispersion Principle which states that large sonority distances in the onsets are preferred by languages [2]. To what degree will this unmarked status of OG- or -GO hold against a large-scale cross-linguistic database? The current study aims to test these theoretical predictions of the unmarkedness status of consonants clusters on large cross-linguistic empirical data.

2. Methodology

1) Phoneme sequences from 496 languages were obtained from two large lexical databases, CLICS2 [4] and AusPhon-Lexicon [6]; 2) *permissible* consonant clusters in each language were obtained; 3) by adopting two widely accepted sonority scales, [gl(4) > nas(3) > liq (2) > obs (1)] (sonority is phonologically defined [2]), and [gl (10) > rho (9) > ... > vcl plo (1)] (sonority is phonetically defined [5]), permitted sonority distances in each language were calculated; 4) the number of languages with each sonority distance type was counted; 5) lastly, sequential polynomial regression was performed to investigate the correlation between the number of languages attested at each sonority distance.

3. Results

There is a significant correlation between sonority distance and the number of languages attested at each sonority distance. The number of languages attested generally increases as sonority distance increases, however, it does not monotonically keep increasing, i.e., the correlation between the number of languages and sonority distance does not show linear model. Rather, the highest significant value in the correlation is attested when a quadratic component is added. Specifically, at the sonority distance value of 4/5, the highest number of languages is attested. This trend stands true, both in onsets and codas, regardless of sonority scales or assumption on complex segments (see statistical details and clusters with optimal sonority distances below).



4. Discussion

OG- and -GO have long been regarded as most common, and thus conceptualised in theories as most unmarked. However, the finding that there are optimal sonority distances where the largest number of languages are attested indicates that not large sonority distances are always favoured by languages as previously assumed. Rather, there are optimal sonority distances that are mostly favoured by languages cross-linguistically (see these clusters in the table above). Therefore, the current large-scale cross-linguistic study calls for empirical articulatory and perceptual search for most frequently attested consonant clusters.

5. References (selected)

[1] Baertsch, K., *An optimality theoretic approach to syllable structure: The split margin hierarchy.* (PhD), Indiana University, Bloomington, 2002.

[2] Clements, G., The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the grammar and physics of speech* (pp. 283-333). Cambridge: Cambridge University Press, 1990

[3] Gouskova, M., Relational hierarchies in OT: The case of syllable contact. *Phonology, 21*(2), 201-250, 2004

[4] List, Johann-Mattis, Greenhill, Simon, Anderson, Cormac, Mayer, Thomas, Tresoldi, Tiago, & Forkel, Robert. 2018. Database of cross-linguistic colexifications, 2018.

[5] Parker, Steve., *Quantifying the sonority hierarchy.* (Ph.D.), University of Massachusetts at Amherst, 2002.

[6] Round, Erich., *The AusPhon-Lexicon project: 2 million normalized segments across 300 Australian languages.* Paper presented at the 47th Poznań Linguistic Meeting, Poznań, 2017a.