Gradient opacity in Uyghur backness harmony: A large-scale corpus study

Connor Mayer (University of California, Irvine)

This paper demonstrates on the basis of a large corpus study that an opaque phonological interaction in Uyghur (Turkic: China) displays variability in rates of opacity that is conditioned by root frequency. This relationship can be accounted for if this opacity is modeled as a type of lexical exceptionality, rather than an ordered interaction between phonological processes.

Background. Uyghur displays *backness harmony*, whereby certain consonants and vowels in suffixes must agree with the final front /æ ø y/ or back /ɑ o u/ root vowel (e.g., k**ø**z-d**æ**/*-dɑ 'eye-LOC'; **ɑ**t-t**ɑ**/*-tæ 'horse-LOC'). The vowels /i e/ are transparent (e.g., **ɑ**mil-**ʁɑ**/*-gæ 'element-DAT'; m**æ**st͡ʃit-**kæ** 'mosque-DAT'). Uyghur also has a vowel reduction process that neutralizes /æ ɑ/ to the harmonically neutral [i] in medial, open syllables (e.g., bɑl**ɑ** 'child', bɑl**i**-lɑr 'child-PL'; sæll**æ** 'turban', sæll**i**-lær 'turban-PL'). The interaction of these processes in disharmonic roots can produce opaque harmonizing behavior (e.g., /ɑp**æ**t-i-GA/ → [ɑpiti**gæ**] 'disaster-3.POS-DAT'; /ʃæjt**ɑ**n-i-GA/ → [ʃæjtini**ʁɑ**] 'devil-3.POS-DAT'). In other cases, surface-true harmony may be obligatory (e.g., /ærz**ɑ**n-i-GA/ → [**æ**rzini**gæ**] 'cheap-3.POS-DAT') or optional (e.g., /æz**ɑ**n-i-GA/ → [æzini**ʁɑ**] or [**æ**zini**gæ**] 'call to prayer-3.POS-DAT').

Methodology. This study uses two large news corpora: *Uyghur Awazi* (6.1m words) and *Radio Free Asia's* Uyghur-language website (9.6m words). Data were collected using webscrapers, and parsed using a morphological transducer (Washington et al. 2020) that decomposes words into their roots and suffixes, and detects suffix backness and whether vowel reduction occurs.

Results. 195 disharmonic roots exhibited reduction of their second vowel. 140 displayed only opaque harmony, while 55 displayed some surface-true harmony. A logistic regression model fit to individual tokens of the 195 roots shows log token frequency is a significant predictor of rates of opaque harmony. Raised /æ/ was also more likely to harmonize opaquely (cf. Vaux 2008).

Analysis. If opaque harmony is modeled as a serial process where harmony precedes reduction, this requires lexically-conditioned variability in the order of rule application (under a rule-based analysis) or ordering of phonological strata (under, e.g., Stratal OT; Bermúdez-Otero 2003), both of which are unexpected. The relationship between frequency and rates of opacity also cannot be straightforwardly accounted for in such models. We adopt a parallel MaxEnt OT model (see tableau below) which treats opaque harmony as a type of *lexical exceptionality* to surface-true harmony. Uyghur speakers' grammars contain general preferences for surface-true harmony (modeled using simple AGREE constraints) as well as a set of lexically-indexed constraints (Pater 2009) that mandate the harmonizing behavior of particular roots (Rebrus & Törkenczy 2017). Larger weights on these indexed constraints correspond to higher rates of opaque harmony. When the model is fit to the corpus data, we observe a positive correlation between the log frequency of a root and the weight of its indexed constraint, aligning with the established relationship between frequency and exceptionality (e.g. Bybee 1985, Morgan & Levy 2016).

Conclusion. The representational challenges opacity poses for parallel models of phonology have been used as support for serial models. Although serial analyses of these patterns are usually straightforward, they may simply recapitulate the historical processes that led to the opacity, rather than modeling speakers' synchronic representations. Treating opacity in Uyghur as lexical exceptionality provides a better account of the patterns observed in the corpus data and allows it to be represented in a strictly parallel model. This approach may be fruitful for cases of opacity in other languages.

/sahabæ$_k$-lAr/	Pred. Freq.	Obs. Freq.	H	VAGREEB $w = 7.37$	VAGREEF $w = 7.82$	*UNREDUCED $w = 31.02$	HARMFRONT$_k$ $w = 7.11$
sahabæ-lær	0	0	31.02			1	
sahabæ-lar	0	0	45.95		1	1	1
sahabi-lær	0.44	0.44	7.37	1			
sahabi-lar	0.56	0.56	7.11				1