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

Connor Mayer
University of California, Irvine

29mfm 2022

cjmayer@uci.edu | www.connormayer.com

1. Uyghur Backness Harmony

Basic pattern: Suffixes agree in backness with final vowel in root.

tʏr-dæ/*-dɑ ‘type-LOC’ pʊl-ʁɑ/*-gæ ‘money-DAT’
mʊnbær-gæ/*-ʁɑ ‘podium-DAT’ ætrɑp-tɑ/*-tæ ‘surroundings-LOC’

The vowels /i e/ are *transparent*

mæstʃit-tæ /*-tɑ ‘mosque-LOC’ mømin-gæ/*-ʁɑ ‘believer-DAT’
stʊdent-lɑr/*-lær ‘student-PL’ ɑmil-ʁɑ/*-gæ ‘element-DAT’

Roots with no harmonizers are lexically specified for backness

biz-gæ/*-ʁɑ ‘us-DAT’ welisipit-lær/*-lɑr ‘bicycle-PL’
sir-lɑr/-*lær ‘secret-PL’ hejt-tɑ/*-tæ ‘festival-LOC’

2. Vowel reduction

/æ ɑ/ raise to [i] in medial, open syllables.

bɑlɑ	‘child’	bɑlɨ-lɑr	‘child-PL’
qɑrɑ-ʃ	‘look-GER’	qɑrɨ-di	‘look-3.SG.PAST’
mewæ	‘fruit’	mewɨ-si	‘fruit-3.SG.POS’
søzlæ-ʃ	‘talk-GER’	søzlɨ-di	‘talk-3.SG.PAST’

Certain words and morphological constructions resist this raising, and /æ/ is more likely to raise than /ɑ/

3. Raising and harmony in disharmonic roots

Two possible interactions for disharmonic roots (vowels **FB** or **BF**)

Surface-true harmony		**Opaque harmony**	
UR	/ɑpæt-i-GA/	**UR**	/ɑpæt-i-GA/
Reduction	ɑpit-i-GA	**Harmony**	ɑpæt-i-gæ
Harmony	ɑpit-i-ʁɑ	**Reduction**	ɑpit-i-gæ
SR	[ɑpitiʁɑ]	**SR**	[ɑpitigæ]

Elicitation result: opaque harmony is most common, but roots can **vary** in whether they display surface-true or opaque harmony.

Opaque	/ɑpæt-i-GA/ →	[ɑpitigæ]	‘disaster-3.POS-DAT’
	/ʃæjtɑn-i-GA/ →	[ʃæjtiniʁɑ]	‘devil-3.POS-DAT’
Surface	/ærzɑn-i-GA/ →	[ærzinigæ]	‘cheap-3.POS-DAT’
Variable	/æzɑn-i-GA/ →	[æziniʁɑ]	‘call.to.prayer-3.POS.DAT’
		[æzinigæ]	

Some Uyghur roots vary in whether they display opaque interactions between vowel reduction and backness harmony. This variation is sensitive to root frequency and other properties.

This variability can be accounted for by modeling opacity as a conflict between morphological knowledge of the harmonizing class of a root, and surface phonotactic constraints.

4. Corpus methodology

Corpora constructed from two online Uyghur newspapers (~15m words).

Morphological parser was used to identify root and detect suffix backness (Washington et al. to appear)

قىزىڭلارغا → Morphological transducer → قىز<n><px2sg><frm><dat-b>

5. Corpus results

190 roots met the criteria to display opacity (*43,450 tokens)*
- BF stems (*n=185*): e.g. /ɑdæt/ ‘custom’, /sijɑsæt/ ‘politics’
- FB stems: (*n=5*): e.g. /ærzɑn/ ‘cheap’ /kæsipdɑʃ/ ‘colleague’

Raised forms are usually opaque, but a number of roots (*n=53*) **vary**

e.g., /ɑhɑlæ/ ‘population’
- Opaque harmony in 79% of cases: [ɑhɑli-lær-gæ]
- Surface harmony in 21% of cases: [ɑhɑli-lɑr-ʁɑ]

Results of statistical analysis (ask me for details)
1. **More frequent roots are more likely to harmonize opaquely**
2. **Roots that appear in raised forms more frequently are more likely to show surface harmony**
3. **/ɑ/-final roots are more likely to show surface harmony** (Vaux 2000)

6. Modeling challenges and proposal

Standard serial theories of opacity cannot model this variability
- E.g., Stratal OT analysis (Bermúdez-Otero 2003) would require probabilistic re-ordering of strata
- Connections to frequency are also unexpected!

Proposal: Uyghur backness harmony has **zones of variation** (Hayes 2016)
- harmony only semi-predictable from phonological properties of roots
- These zones require **lexical knowledge** of harmony class

If we treat **opacity** as another **zone of variation**, we can model it using the same mechanisms!

7. Phonological modeling

Maximum entropy optimality theory with indexed constraints (e.g., Pater 2009, Moore-Cantwell and Pater 2016, a.o.).
- Indexed constraints mandate front/back allomorphs
- Phonological constraints mandate surface harmony
- Most of the time these agree, but in opaque forms they don’t!

/sɑhɑbæ-lAr/	**Pred. freq.**	**VAgree Back** *w=4.7*	**VAgree Front** *w=10.4*	**Harmonize Back**$_{sahabe}$ *w=2.7*	**Harmonize Front**$_{sahabe}$ *w=7.1*	***Unraised** *w=24.8*
sɑhɑbi-lær	**0.44**	1		1		
sɑhɑbi-lɑr	**0.56**				1	
sɑhɑbæ-lær	**0**			1		1
sɑhɑbæ-lɑr	**0**		1			1

HARMONIZE weighted by **speaker certainty** of harmonic class

e.g., HARMONIZEBACK$_x$ ∝ $P(HC=back|x)$

Components
- **Phonotactic probability** based on UR that *x* is a back harmonizer, weighted by root activation (cf. Becker and Gouskova 2016, Breiss 2021)
- **Structural knowledge** about *x’s* harmony class
  - How many times have we seen *x* with a back suffix?
- **Bias** towards back suffixes (default class in Uyghur)

8. Discussion

Treating opacity as a consequence of lexical listing of morphological class lets us capture it in a parallel model
- Unifies it with other zones of variation in Uyghur
- Straightforward to model variability and the influence of frequency-based effects (cf. Coetzee and Kawahara 2013)

Uyghur backness harmony has a **dual life** as a **phonological process** and a **morphological class system**
- Like grammatical gender, but with strong predictability from phonotactics (cf. Becker and Dow 2013, Kupisch et al. 2022)
- Similar to proposal for Hungarian (Rebrus & Törkenczy 2017)

Future work: More colloquial corpora; experimental validation; further refine properties of model

Selected References

Becker, Dow (2013). Gender without morphological segmentation in French. *Phonology 2013*, UMass Amherst. **Becker, Gouskova (2016).** Source-oriented generalizations as grammar inference in Russian vowel deletion. *LI 47*:3. **Bermúdez-Otero (2003).** The acquisition of phonological opacity. *Variation within OT.* **Breiss (2021).** Lexical conservatism in phonology. PhD Thesis. **Coetzee, Kawahara (2013).** Frequency biases in phonological variation. *NLLT 31:1*. **Hayes (2016).** Comparative phonotactics. *50th CLS*. **Kupisch et al. (2022).** Structural and phonological cues for gender assignment… Glossa 7:1. **Rebrus, Törkenczy (2017).**Co-patterns, subpatterns and conflicting generalizations in Hungarian vowel harmony. *Approaches to Hungarian, 15*. **Washington et al. (to appear).** Free/open-source technologies for Turkic languages… *TURKLANG 2019*. **Vaux, (2000).** Disharmony and derived transparency in Uyghur vowel harmony. *NELS 30*.

Statistical analysis

Mixed effects logistic regression model fit to tokens

- Dependent variable: Does token exhibit opaque harmony?
- Fixed effects
  - intercept (β=6.38, z=5.06, p = 0)
  - log token frequency per million words (β=0.51, z=2.87, p < 0.005)
  - proportion of tokens that are raised (β=-3.25, z=-2.99, p < 0.005)
  - final vowel identity (reference level **æ**) (β=-3.94, z=-2.55, p < 0.05)
- Random intercepts
  - Root identity (σ=3.57)
  - Corpus (σ=1.20))