1. Background

Phonological opacity (Kiparsky 1973)
Assume rule of the form A → B / C_D. This rule is opaque if there are surface forms with either:
- A in environment C_D (underapplication)
- A → B in environments other than C_D (overapplication)

Opacity may pose learning difficulties (Kiparsky 1973)

Many theoretical battles have been fought over opacity
- Baković (2011) shows no theory really captures all its forms

This project:
- Examines an opaque pattern in Uyghur (Turkic: China) based on a large-scale corpus study
- Suggests this opaque pattern is learned based on lexically-specific constraints as function of frequency

2. Uyghur Backness Harmony

Basic pattern: Sounds in suffixes must agree in backness with /æ/ in stem:
- /æ/-day/*-dg ‘type-LOC’
- /æ/-ka/*-gæ ‘money-DAT’
- /æ/-business/*-du ‘podium-DAT’
- /æ/-strap-*la ‘surroundings-LOC’

The vowels /æ/ /æ/ are transparent.
- /æ/-dil-ta/*-tg ‘mosque-LOC’
- /æ/-menin/*-gæ ‘believer-DAT’
- /æ/-student-*la ‘student-PL’
- /æ/-amir/*-gæ ‘element-DAT’

If a stem contains no harmonizing elements, it is lexically specified for backness, with a tendency towards back suffixes.
- /æ/-gæ/*-gæ ‘us-DAT’
- /æ/-velop/*-lær ‘bicycle-PL’
- /æ/-sir-*lä ‘secret-PL’
- /æ/-hejt-*ta ‘festival-LOC’

3. Vowel raising

The vowels /æ/ /æ/ raise to /œ/ in medial, open syllables.
- bolq ‘child’
- bolq-*lar ‘child-PL’
- qara/*-gæ ‘look-GER’
- qara-*di ‘look-3.SG.PAST’
- mewa ‘fruit’
- mewa-*si ‘fruit-3.SG.POS’
- seza/*-lær ‘talk-GER’
- seza-*di ‘talk-3.SG.PAST’

Certain words and morphological constructions resist this raising.
- hava ‘weather’
- hava-*si ‘weather-3.POS’
- sawa/*-lær ‘reason’
- sawa/*-di ‘reason-3.POS’

/æ/ is more likely to raise than /a/

4. Corpus methodology

Examined raising patterns from corpora constructed from two online Uyghur newspapers (total of about 15 million words).
- Uyghur orthography transparently reflects harmony and raising of /æ/.
- An automated morphological parser was used to parse words (Washington et al. to appear)
- Maps from orthographic forms to stem + morphological tags.
- Modified to detect suffix backness as well as stem and morpheme identity.
- Text processing done on stem and surface forms to identify raising and opaque harmony.

5. Corpus results

183 display raising:
- BF stems (n=177)
- FB stems (n=6)

Raised forms are generally opaque, but a portion of these (n=87) vary in whether they display surface harmony or opaque harmony.
- /æ/-lær/*-gæ ‘population’
- /æ/-gæ/*-gæ ‘element-DAT’
- /æ/-marsha/*-gæ ‘private’
- /æ/-shagh ‘condition’

Linear regression shows that log word frequency and proportion of unraised tokens are significant predictors of opaque harmony.

6. Phonological analysis

Model using maximum entropy optimality theory with paradigm uniformity constraints (e.g. Steriade 2000).
- PU constraint: harmonize as the unraised form does
- Weights for PU learned as function of exposure to base

7. Discussion/Future Directions

Gradient opacity indicates that opacity is not learned as a completely productive process.
- Gradient captured by lexically-specific PU constraints that are weighted as function of base exposure
- Is this indicative of language change?
- Raising is relatively new. Could surface harmony become the norm?

Corpora and computational methods provide greater empirical support to phonological theory.

Future work: Explicit learning model!!

Selected References