

Evidence for the Sonority Scale in Kazakh
What Kazakh can tell us about Sonority

Jonathan North Washington
jonwashi@indiana.edu

3rd Annual IU Linguistics Department Graduate Student Conference

27 March 2009

1 *Introduction*

- Overview
- Sonority is important
- Kazakh is underdocumented
- Sonority phonology of Kazakh is striking

2 *Data*

- Onsets
- Coda clusters
- Syllable Contact

3 *Discussion*

- Summary
- Thoughts and future directions

Overview

What this research is about

Aim

Better understand full picture of sonority in Kazakh

Methodology

Examining sonority-related features and processes

Why?

- Better generalisations about Kazakh (poorly documented)
- Sonority phonology of Kazakh is striking
- Sonority is important

Sonority is important

- Sonority is ubiquitous
 - Assumed to be a driving factor in many syllable-structure-related processes = syllables often assumed to be sonority peaks
- Sonority isn't well understood
 - Analyses of different languages can arrive at different hierarchies
 - Ambiguity within individual languages regarding certain phonemes
 - Difficult to quantify: phonetic cues elusive (Parker, 2002)
- A better understanding of sonority is valuable
 - Example of disconnect between phonetics and phonology
 - Studying such disconnects can improve understanding of big-picture questions

Kazakh is underdocumented

- Kazakh has ~12 million native speakers
 - ≈ all native Finnish and Norwegian speakers combined
 - ≈ all native Greek speakers
 - Among top 70 most widely spoken native languages
- But English-language materials lacking
 - **Dictionaries** (Krippes, 1994) full of errors
 - [Teaching/reference] **grammars** (Krippes, 1996; Bekturova and Bekturov, 1996; Kirchner, 2006; Krueger, 1980; Somfai Kara, 2002; Ираида Кубаева, 2003) not to be trusted
(not based on linguistic work)
 - **Scholarly work** suffers from lack of native speaker linguists (cf. Greek, Finnish, Norwegian)
 - Only some well-informed linguistic work (Eulenberg, 1996; Vajda, 1994)

Sonority phonology of Kazakh is striking

- Historically **limited subset of phonemes** allowed **word-initially**
 - almost no sonorants or voiced stops...
- Common, but very restricted **complex codas**
 - C_2C_1 where C_2 must be a sonorant and C_1 a voiceless stop
- Falling **sonority across syllable boundaries** enforced by desonorisation
 - $/sy\text{ʉ}z/$ 'word' + $-/LAR/$ (plural) $\rightarrow /sy\text{ʉ}zdj\text{ə}r/$ 'words'
(historically by segmental & featural metathesis)
 - $*/j\text{a}\text{ʉ}mur/ \rightarrow /z\text{a}nb\text{ə}r/$ 'rain'

Onsets

Word-initial position in Proto-Turkic highly restricted

Proto-Turkic phonemes in word-initial position (Johanson, 2006)

	labial		dental		palatal		dorsal	
stops	/p/	/b/	/t/	/d/	/tʃ/	∅ ¹	/k/	/g/
	*	✓	✓	*	✓		✓	*
fricatives			/s/	/z/	/ʃ/	∅ ¹		
			✓	*	*			
nasals	/m/		/n/		/ɲ/		/ŋ/	
	*		1ω		*		*	
other	lateral		rhotic		glide		vowels	
	//		/r/		/j/ ¹		∅	
	*		*		✓		✓	

Onsets

Word-initial position in Proto-Turkic highly restricted (ctd)

Generalisation about Proto-Turkic word-initial phonemes

Only voiceless obstruents begin words.

Exceptions

- */j/, unless assumed to be obstruent
- single word */næ/ 'what'
- */b/, but Pre-Proto-Turkic word-initial */p/ → */h/ → ∅
- (/ʃ/ not attested, but predicted?)
- (vowels okay at beginning of word—not onsets)

Onsets

Kazakh frequency count

Phonemes beginning Kazakh words occurring >1.5Kx in 24M word corpus (x1000)

	labial		dental		palatal		dorsal	
stops	/p/	/b/	/t/	/d/	/tʃ/		/k/	/g/
	308	2136	1054	798	57		2571	169
fricatives	/f/	/β/	/s/	/z/	/ʃ/	/ʒ/		
	21	83	948	141	210	1812		
nasals		/m/		/n/				/ŋ/
		1050		458				0
glides		/w/				/j/		
		37				33		
	lateral		rhotic		vowels		glottal	/ts/
	/l/		/r/		∅		/h/	/tʃ/
	22		281		4454		0	+

Onsets

Kazakh frequency count (ctd.)

- Kazakh corpus of 24M words (782K unique)
- Phonemes beginning words occurring >1500 times in corpus (in thousands)
- Sanity check: Russian-only /tʃ/, /f/, /v/, /ts/
- Kazakh word-initial /ʒ/ ← */j/ and /m/ ← /b/
- Almost everything “allowed”, but similar tendencies as Proto-Turkic.
- Factors:
 - Relatively recent borrowings from Persian and Russian have maintained faithfulness
 - Corpus contamination

Coda clusters

C_1, C_2	p	k	t	s	ʃ
w	*	*	*	*	*
j	*	*	✓	*	*
r	*	*	✓	*	*
l	*	*	✓	*	*
n	*	*	✓	*	*
m	*	*	*	*	*
ŋ	*	✓	*	*	*

Table: Kazakh $C_2C_1]_\sigma$

Coda clusters

C_1, C_2	p	K	t	s	ʃ
w	*	*	*	*	*
j	*	*	✓	*	*
r	*	*	✓	*	*
l	*	*	✓	*	*
n	*	*	✓	*	*
m	*	*	*	*	*
ŋ	*	✓	*	*	*

Table: Kazakh $C_2C_1]_\sigma$

- Generalisations (\approx constraints):
 - C_2 must be sonorant
 - C_1 must be voiceless obstruent
 - ... a stop
 - C_1 and C_2 must be homorganic
- Onomatopoetic words excluded
 - (/tars/ 'bang', /zarq/ 'twinkle')
- Gaps:
 - /wp/, /mp/
 - Accidental historical gaps?
 - Kyrgyz fills /mp/ gap

Syllable Contact

Data

	-L	-N	-NVN	-M	-MVN	-D	-DVN
V	V.l	V.n	V.n	V.m	V.m	V.d	V.d
j, w	J.l	J.d	J.d	J.m	J.m	J.d	J.d
r	r.l	r.d	r.d	r.m	r.m	r.d	r.d
l	l.d	l.d	l.d	l.m	l.m	l.d	l.d
N	N.d	N.d	N.n	N.b	N.m	N.d	N.n
C _[+vc]	C.d	C.d	C.d	C.b	C.b	C.d	C.d
C _[-vc]	C.t	C.t	C.t	C.p	C.p	C.t	C.t

Table: Kazakh syllable contact

Syllable Contact

Data

	-L	-N	-NVN	-M	-MVN	-D	-DVN
V	V.l	V.n	V.n	V.m	V.m	V.d	V.d
j, w	J.l	J.d	J.d	J.m	J.m	J.d	J.d
r	r.l	r.d	r.d	r.m	r.m	r.d	r.d
l	l.d	l.d	l.d	l.m	l.m	l.d	l.d
N	N.d	N.d	N.n	N.b	N.m	N.d	N.n
C _[+vc]	C.d	C.d	C.d	C.b	C.b	C.d	C.d
C _[-vc]	C.t	C.t	C.t	C.p	C.p	C.t	C.t

Table: Kazakh syllable contact

- /l/ and /m/ desonorise after C of equal or lower sonority
- /n/ desonorises after any consonant
- (...trumped by nasal harmony)
- (Devoicing of onset after voiceless coda)

Syllable Contact

Discussion

Definition (Syllable Contact Law)

Tendency to avoid rising sonority across a syllable boundary
(Murray and Vennemann, 1983)

or

Tendency for sonority to fall across a syllable boundary (Baertsch and Davis, 2001, 30)

- Syllable Contact Law responsible for // and /m/ desonorisation
- What about /n/? Doesn't pattern like /m/.

Syllable Contact

The problem of /n/

*ONS/V >> *ONS/J >> *ONS/R >> *ONS/L >> *ONS/N >> *ONS/Obs

- Assume /n/ is a normal nasal
- Set of constraints blocking syllable onsets: “high sonority = bad onset”

Syllable Contact

The problem of /n/

faithfulness constraints



*ONS/V >> *ONS/J >> *ONS/R >> *ONS/L >> *ONS/N >> *ONS/Obs

- Assume /n/ is a normal nasal
- Set of constraints blocking syllable onsets: “high sonority = bad onset”
- Kazakh generally doesn’t like sonorant onsets

Syllable Contact

The problem of /n/



- Assume /n/ is a normal nasal
- Set of constraints blocking syllable onsets: “high sonority = bad onset”
- Kazakh generally doesn’t like sonorant onsets
- // is basically tolerated
- /m/ is basically tolerated
- /n/ remains subject to not wanting to be an onset, while /m/ remains subject to the Syllable Contact Law

Summary

What we've learned about sonority in Kazakh

- **Word onsets, historical:**
 - No sonorants allowed except /j/, which might've been obstruent
 - Only unvoiced obstruents allowed, except /b/
- **Word onsets, current:**
 - Highly dispreferred (<100K): /w/, /j/, //, /ŋ/
 - Plus less common (<500K): /p/, /G/, /z/, /ʃ/, /n/, /r/
 - Versus most common (>1M): /K/, /b/, /ʒ/, /t/, /m/
- **Complex codas:**
 - Sonorant + homorganic voiceless stop
- **Syllable contact:**
 - // desonorises after // and Cs of lower sonority
 - /m/ desonorises after /m/ and Cs of lower sonority
- $V > /j/, /w/, /r/ > // > \text{nasals} > \text{some obstruents} > \text{vcls stops}$

Thoughts and future directions

What now?

- Word onset \neq syllable onset
 - Qualify the difference with data and attempt to treat it
- Finishing touches on OT analyses of
 - complex coda phenomena
 - syllable contact phenomena
- Combine analyses?
- Do similar work on Kyrgyz
 - fairly closely related language
 - more complicated coda cluster phenomena (more conservative)

Thoughts and future directions

What now?

- Word onset \neq syllable onset
 - Qualify the difference with data and attempt to treat it
- Finishing touches on OT analyses of
 - complex coda phenomena
 - syllable contact phenomena
- Combine analyses?
- Do similar work on Kyrgyz
 - fairly closely related language
 - more complicated coda cluster phenomena (more conservative)
- What else? Interesting data, but theoretically useful?

References

- Baertsch, Karen and Stuart Davis (2001). *Turkic C+III(uster) Phonology*. CLS 37:29–43.
- Bekturova, Sh. and K. Bekturov (1996). *Manual of the Kazakh Language*. Almaty: Rauan.
- Ираида Кубаева (2003). *Казахский язык: Просто о невероятно сложном*. *Kazakh Language Made Easy*. Almaty: Atamura.
- Eulenberg, Alex (1996). *Voicing and consonantal strength in Kazakh suffixes*. Indiana University .
- Johanson, Lars (2006). *History of Turkic*. In Lars Johanson and Éva Á. Csató (Eds.) *The Turkic Languages*, chap. 5, pp. 81–125. New York: Routledge.
- Kirchner, Mark (2006). *Kazakh and Karakalpak*. In Lars Johanson and Éva Á. Csató (Eds.) *The Turkic Languages*, chap. 19, pp. 318–332. New York: Routledge.
- Krippes, Karl A. (1994). *Kazakh (Qazaq) - English Dictionary*. Kensington, MD: Dunwoody.
- (1996). *Kazakh Grammar with Affix List*. Kensington, MD: Dunwoody.
- Krueger, John R. (1980). *Introduction to Kazakh*. Bloomington, Indiana: Indiana University Research Institute for Inner Asian Studies.
- Murray, R. W. and T. Vennemann (1983). *Sound change and syllable structure in Germanic phonology*. *Language* 59(3):514–528.
- Parker, Steve (2002). *Quantifying the sonority hierarchy*. Ph.D. thesis: University of Massachusetts, Amherst .
- Somfai Kara, Dávid (2002). *Kazak*. München: Lincom Europa.
- Vajda, Edward J. (1994). *Kazakh Phonology*. In Edward H. Kaplan and Donald W. Whisenhunt (Eds.) *Studies on East Asia*, vol. 19: “Opuscula Altaica: Essays Presented in Honor of Henry Schwarz”, pp. 603–650. Western Washington University.