

The three faces of Kazym Khanty schwa

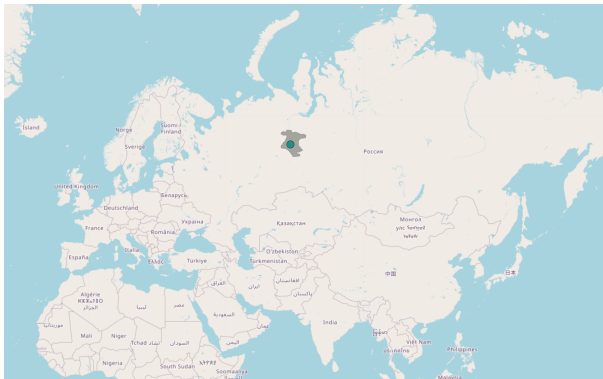
Sasha Shikunova

HSE University (Moscow)

Atelier de phonologie, October 18th 2023

Kazym Khanty

- » Minority Uralic language spoken in Khanty-Mansi and Yamalo-Nenets okrugs of Russia
- » Kazym Khanty – a Northern Khanty dialect, data collected in Kazym village (Khanty-Mansi okrug)



<https://uralic.cldd.org/languages/27>

Khanty schwa

» Schwa can be inserted to avoid complex onsets, which are prohibited

(1) *aškola* from Russian *škola* 'school'

» Schwa can alternate with zero

(2) a. *juχət-λ* 'come-NPST[3SG]'

b. *juχt-λ-ən* 'come-NPST-2SG'

» Schwa can be a stable vowel that never alternates with zero

(3) a. *orət-s* 'drag-PST[3SG]'

b. *orət-s-ən* 'drag-PST-2SG'

Search for the simplest analysis

- Phonologically, is it possible to treat the schwa as a single entity, whose surface realisation depends on the context?
- That is, can we avoid postulating two different schwa phonemes with distinct behaviour?
- Is it possible to avoid allomorphy where schwa alternates with zero?

Data

Phonotactic restrictions

- No complex onsets: #CC is absent in native words and broken up in loanwords
- Metathesis or epenthesis of schwa

(4) Rescuing illicit clusters in loanwords

- a. *kinška* from Russian *knižka* 'book'
- b. *aškola* from Russian *škola* 'school'

- Sonority Sequencing Principle: coda clusters must have decreasing sonority (evidenced by coda clusters that do occur)

(5) Coda clusters in Khanty

- a. *jeɾt* 'rain'
- b. *artna* 'debt'

(Solovar 2014)

Verbal inflection

- » base-(derivation)-tense-(inversive)-agreement
- » Tense: past vs non-past

NPST	-λ-
PST	-(ə)s-

- » Agreement: subject and object (only showing subject agreement paradigm)

1SG	-əm	1DU	-əmən	2PL	-əw
2SG	-ən	2DU	-ətən	2PL	-əti
[3SG]	-∅	3DU	-əηən	2PL	-ət

- (6) a. *mǎn-əs* 'walk-PST[3SG]'
- b. *mǎn-s-əmən* 'walk-PST-1DU'

Verbal bases, C#

- » There are two types of verbal bases where schwa can occur
- » One with an alternating schwa (*ir(ə)t-* 'turn') and the other with a stable schwa (*orət-* 'drag')

Form	No schwa	Alternating schwa	Stable schwa
	ort- 'divide'	ir(ə)t- 'turn'	orət- 'drag'
NPST[3SG]	ort-əλ ~ orλ	irət-λ	orət-λ
PST[3SG]	ort-əs ~ ors	irt-əs	orət-s
NPST-2SG	or-λ-ən	irt-λ-ən	orət-λ-ən
PST-2SG	or-s-ən	irt-s-ən	orət-s-ən
NPST-1DU	or-λ-əmn	irt-λ-əmn	orət-λ-əmn
PST-1DU	or-s-əmn	irt-s-əmn	orət-s-əmn

Verbal bases, C#

- » Tense endings *-s/-λ* 'PST/NPST' have no schwa, except upon careful pronunciation and in PST.3SG
- » Agreement endings *-ən/-əmn* '2SG/1DU' do have a schwa

Form	No schwa	Alternating schwa	Stable schwa
	ort- 'divide'	ir(ə)t- 'turn'	orət- 'drag'
NPST[3SG]	ort-əλ ~ orλ	irət-λ	orət-λ
PST[3SG]	ort-əs ~ ors	irt-əs	orət-s
NPST-2SG	or-λ-ən	irt-λ-ən	orət-λ-ən
PST-2SG	or-s-ən	irt-s-ən	orət-s-ən
NPST-1DU	or-λ-əmn	irt-λ-əmn	orət-λ-əmn
PST-1DU	or-s-əmn	irt-s-əmn	orət-s-əmn

Verbal bases, V#

- » With vowel-final bases, agreement suffixes lose the schwa ⇒ this schwa is not stable
- » After some bases like *ari-* 'sing', a glide appears in 3SG but not elsewhere

Form	a#	i#	i<j>#
	xunta- 'run'	ji- 'become'	ari- 'sing'
NPST[3SG]	xunta-λ	ji-λ	arij-λ
PST[3SG]	xunta-s	ji-s	arij-s
PST-2SG	xunta-s-n	?	ari-s-ən
PST-2SG	xunta-s-mən	ji-s-mən	?

Nominal inflection

- » Nominal inflection: base-number-possessive-case
- » Intriguing contrast between POSS.2PL and LOC, which both look like $-(ə)n$ but differ with the $i\#$ noun $w\theta\lambda i$ 'deer'
- » Structurally lower possessive merges with the final vowel but not case, which is higher in the fseq ([Longobardi 2001](#))

	VC#	əC#	Ca#	Ci#
NOM	xot 'house'	sθm(ə)t 'birch'	λapka 'shop'	wθλi 'deer'
PL	xot-ət	sθmt-ət	λapkaj-(ə)t	wθλe-t
P.2PL	xot-ən	sθmt-ən	λapkaj-(ə)n	wθλe-n
LOC	xot-ən	sθmət-n	λapkaj-(ə)n	wθλij-(ə)n

Is this allomorphy?

Allomorphy solution

- For a similar schwa pattern in a different dialect of Khanty (Tegi), [Kozlov \(2012\)](#) proposes a two-module solution
- Morphophonological module chooses between schwa and schwa-free allomorphs of tense and agreement suffixes
- Phonetic module is responsible for deleting the schwa in connected speech in some contexts, e.g. next to sonorants or between homorganic consonants
- Very similar analysis works for Kazym Khanty

Schwa and no-schwa allomorphs

(7) Verbal base *ir(ə)t* 'turn'

- a. $\sqrt{\text{turn}} \leftrightarrow \text{irət:}$ _NPST[3SG] *irət-λ* 'turn-NPST[3SG]'
- b. $\sqrt{\text{turn}} \leftrightarrow \text{irt:}$ elsewhere *irt-əs* 'turn-PST[3SG]'
- irt-s-ən* 'turn-PST-2SG'

(8) Tense suffix *-(ə)s* 'PST'

- a. PST \leftrightarrow əs: C#_3SG *mǎn-əs* 'walk-PST[3SG]'
- b. PST \leftrightarrow s: elsewhere *mǎn-s-əm̩n* 'walk-PST-1DU'
- ji-s-mən* 'become-PST-1DU'

Agreement suffix depends non-locally on the base

(9) Agreement suffix *-(ə)n* '2SG'

- a. 2SG \leftrightarrow ən: $\sqrt{\text{C\#_}}$ *irt-s-ən* 'turn-PST-2SG'
- b. 2SG \leftrightarrow n: $\sqrt{\text{V\#_}}$ *xunta-s-n* 'walk-PST-1DU'

Phonetic module

- » Not every underlying schwa is pronounced in Kozlov's (2012) analysis
- » Schwas can disappear on the surface in a number of contexts:
 - next to a sonorant
 - between homorganic consonants
 - between a sibilant and a nasal
 - next to /λ/

Adopting the allomorphy solution?

- The schwa vs no-schwa allomorphy is partially phonologically conditioned
- Recall the mixed effects of the base-final consonant and the agreement phi-features on the form of $-(ə)s$ 'PST'

(10) Tense suffix $-(ə)s$ 'PST'

- | | | |
|----|------------------------------------|----------------------------------------------------------------------|
| a. | PST \leftrightarrow əs: C#_3SG | <i>mǎn-əs</i> 'walk-PST[3SG]' |
| b. | PST \leftrightarrow s: elsewhere | <i>mǎn-s-əm̄n</i> 'walk-PST-1DU'
<i>ji-s-mǎn</i> 'become-PST-1DU' |

- This allomorphy seems like a surface-level generalisation that can be analysed further
- The works of the phonetic module need to be tailored to Kazym Khanty and made precise

Are there many schwas?

OT solution

- » Egorov & Tjutjunnikova (2023) suggest postulating 2 schwas: /ə/ and /ə₁/
- » One is stable and one is subject to the DEL(ə₁) constraint
- » Other constraints include:
 - *COMPLEXONS – no complex onsets
 - SSP – Sonority Sequencing Principle; decreasing sonority in codas
 - NUCLEUS – syllable nucleus must be a vowel
 - *CTC – underlying CtC cluster is simplified to CC
- » /λ/ and /n/ can become nuclei

OT solution

- » *COMPLEXONS = SSP > DEL(\emptyset_1) > NUCLEUS > *CTC
- » Phonotactic constraints rank higher than DEL(\emptyset_1), so schwa is deleted only if these constraints are satisfied
- » DEL(\emptyset_1) ranks higher than NUCLEUS because with C# bases, schwa is deleted from the agreement suffix *-əmən* '1DU' even word-finally, where /n/ is the only candidate for the nucleus slot

(11) *irt.λə.mn*, **i.rətλ.mən* 'turn-PST-1DU'

- » DEL(\emptyset_1) ranks higher than *CTC because clusters are not simplified where a deleted schwa used to break them up; otherwise, schwa would just remain to save the cluster

(12) *orλəmn* 'divide-PST-1DU' vs *irtλəmn* 'turn-PST-1DU'

OT fails

- » Tableau from [Egorov & Tjutjunnikova \(2023\)](#), modified to include more candidates
- » *ir.λn* would in fact win over *ir.λən*, which the authors suggest as a winner
- » Both are not the actual output – *irt.λən*
- » Ranking $\text{DEL}(\emptyset_1)$ over everything and simultaneously widening the range of possible nuclei to include /λ/ and /n/ does not work

$\text{ir}\emptyset_1\text{t} + \lambda + \emptyset_1\text{n}$	*COMPLEXONS	SSP	$\text{DEL}(\emptyset_1)$	NUCLEUS	*CTC
☞ <i>ir.λn</i>				*	
<i>ir.λən</i>			*		
<i>irt.λən</i>			*		*
<i>i.rət.λn</i>			*	**!	
<i>i.rət.λən</i>			**!		

OT gets back up

- » In order to get to the correct optimal output, Harmonic Serialism (McCarthy 2010) is assumed
- » GEN only generates candidates that are similar to the input
- » Candidate ranking is repeated until just one candidate remains

Bottom line: constraints are difficult to rank; we need at least two different schwas

Search for the simplest analysis

- [Kozlov \(2012\)](#): schwa is stable, alternations explained by allomorphy
- [Egorov & Tjutjunnikova \(2023\)](#): there are two schwas (stable and alternating), alternations come from an interplay of $\text{DEL}(\text{ə}_1)$ with other constraints; there's also allomorphy

Can we dispense with both allomorphy and two distinct schwas?

The right representations

The association line

- » Strict CV (Kaye, Lowenstamm & Vergnaud 1990, Scheer 2004) supposes that phonological representations consist of a syllabic tier and a melodic tier with association lines in between
- » A piece of melody can take three logically possible underlying forms:
 - Associated
 - Floating
 - Empty (empty slots are filled when ungoverned)

(13) Associated

C V
|
ə

(14) Floating

C V

ə

(15) Empty slot

C V

The association line

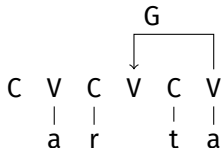
» Vowel-zero alternations are restricted by the Empty Category principle:

(16) Empty Category principle

A position may be uninterpreted phonetically if it is properly governed (Kaye, Lowenstamm & Vergnaud 1990: p. 219)

» where proper government is a relation between a filled V-slot and an empty slot on its left

(17) Empty nucleus governed in the cluster in [arta]



The association line

- » Associated schwa does not alternate with zero
- » Floating schwa is expected to alternate with zero
- » Empty V is only filled when phonotactic requirements are not met, i.e. in prohibited clusters

What kinds of schwa do we find in Kazym Khanty?

All three

Schwas in the verbal paradigm

What kinds of schwa do we find:

- » In verbal bases of two types (alternating and stable)?
- » Before tense endings $-\lambda$ 'NPST' and $-s$ 'PST'?
- » In agreement endings $-\partial n$ 'NPST' and $-\partial mn$ 'PST'?

And why are they behaving like they do?

Proviso

» Together with rules dictated by association and government, there is a constraint against two schwas in a row

(18) No Two SCHWAS IN A ROW

If two schwas occur one after the other on the surface, the latter is deleted.

» Not sure yet if I can formalise this constraint well

» If I could, I would have avoided postulating it

Empty V-slots

- » I assume that schwa can appear as an epenthetic vowel that fills empty V-slots to rescue prohibited clusters
- » In Khanty, initial clusters are prohibited; schwa epenthesis results

(19) Rescuing illicit clusters in loanwords

- a. *kinška* from Russian *knižka* 'book'
- b. *aškola* from Russian *škola* 'school'

- » Also, coda clusters must have decreasing sonority ([Egorov & Tjutjunnikova 2023](#))

Empty V-slots

- -s 'PST' appears with a schwa on the left only in 3SG after (ə)C# bases (with an alternating schwa)
- -λ 'NPST' never occurs with a schwa

- (20) a. *irətλ* 'turn.NPST'
b. *irtəs* 'turn.PST'

- /rts/ and /rtλ/ clusters are possible word-internally but not word-finally

- (21) a. *irt-λ-ən* 'turn-NPST-2SG'
b. *irt-s-ən* 'turn-PST-2SG'

- /rts/ and /rtλ/ are broken up word-finally but in different ways
- /λ/ apparently can become nucleus while /s/ can't

Deriving epenthetic schwas

(22) *iratλ* 'turn.NPST'

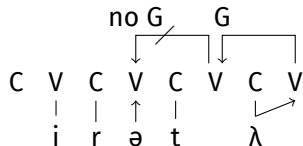
C	V	C	V	C	V	+	C	V
	i	r		t			λ	

(23) *irtəs* 'turn.PST'

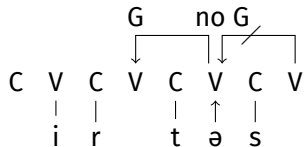
C	V	C	V	C	V	+	C	V
	i	r		t			s	

Deriving epenthetic schwas

(24) *iratλ* 'turn.NPST'



(25) *irtəs* 'turn.PST'



Stable schwa

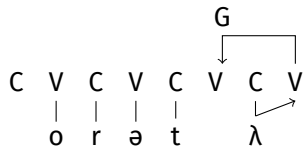
- » Stable schwa is found in non-alternating verbal bases like *orət-* 'drag'
- » The vowel is associated in the lexical representation and therefore acts like other full vowels
- » The schwa-deleting rule removes the second schwa that should appear in *orəts* 'drag-PST[3SG]'

(26) *orət-* 'drag'

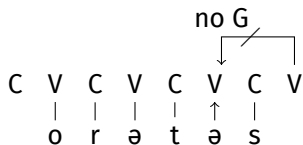
C	V	C	V	C	V
	o	r	ə	t	

Deriving stable schwa

(27) *orətλ* 'drag.NPST'



(28) *orəts* 'drag.PST'



→ *orəts* (second schwa in a row deleted)

Floating schwas

- » Floating schwas are found in agreement endings
- » *-əmən* '1DU' contains a second schwa because it can show up in some cases: *ji-s-mən* 'become-PST-1DU'

(29) *-ən* '2SG'

	C	V
	ə	n

(30) *-əmən* '1DU'

	C	V	C	V
	ə	m	ə	n

- » The effect of C# vs V# bases

(31) a. *xunta-s-mən*

'run-PST-1DU'

b. *irt-s-əmən*

'turn-PST-1DU'

Agreement endings

- » Agreement endings contain alternating schwas
- » After V# bases, the floating schwa is deleted

- (32) a. *xunta-s-n* 'run-PST-3SG'
b. *xunta-s-mən* 'run-PST-1DU'

- » After C# bases, the floating schwa remains

- (33) a. *irt-s-ən* 'turn-PST-3SG'
b. *irt-s-əmn* 'turn-PST-1DU'

Non-local phonology

- Is it possible that agreement suffix influences the base over the tense suffix?
- No apparent reason to remove the initial schwa of suffixes after V# bases
 - Schwas are not deleted after full vowels elsewhere

- (34) a. *xot-ət* 'house-PL'
b. *irətλ* 'turn.NPST'
- Stress pattern or syllable count do not matter

- (35) Two bases, different rhythm, both C#
- a. *(o.rət)(sə.mn)* 'drag-PST-1DU'
b. *(mǎn.sə)mn* 'go-PST-1DU'

- Khanty stress is trochaic, feet parsing goes from left to right

Non-local action

- » The alternation in the agreement endings is similar to a vowel coalescence escape strategy

(36) Vowel deletion to avoid coalescence (Moksha, Uralic)

- a. *kud + ən'* → *kudən'* 'house-GEN' /dən'/ → /dən'/'
- b. *ava + ən'* → *avan'* 'woman-GEN' /aən'/ → /an'/'

- » Schwa cannot surface after another vowel
- » The only difference is that in Khanty verbal inflection, this process is non-local: it occurs over another segment
- » Non-local phonological effects have been observed in other languages ([Myler 2017](#))

Deriving alternating schwas

- » Why is the /rts/ cluster licit word-internally and broken up word-finally?
- » I tentatively suggest that the agreement suffix is inserted prior to tense (cf. parallel case from Nyakusa; Hyman 2000, Myler 2017)

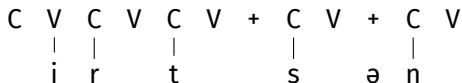
(37) Variable blocking of spirantization by reciprocal -an (Hyman 2000: 9, his 44)

- a. *sob-* 'get lost (intr.)'
- b. *sof-j* 'to lose (tr.)'
- c. ***sob-an-j*** 'get each other lost' (causativized reciprocal)
- d. ***sof-an-j*** 'to lose each other' (reciprocalized causative)

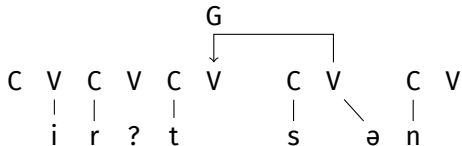
Deriving alternating schwas

- Assume perfectly linear composition
- The final empty nucleus of the base is governed by the vowel of *-ən* '2SG'
- What about the internal empty slot of the base? Why is it not filled?

(38) *irt-s-ən* 'turn-PST-2SG'



(39) *irt-s-ən* 'turn-PST-2SG'



Deriving alternating schwas

- Now assume that the agreement suffix is attached first
- The internal empty V of the base is governed

Step 1

(40) *irt + ən* 'turn + 2SG'

C	V	C	V	C	V	+	C	V
	i	r		t			ə	n

(41) *irt + ən* 'turn + 2SG'

			G					
			↓	┌	└			
C	V	C	V	C	V	+	C	V
	i	r		t			ə	n

Deriving alternating schwas

- The tense suffix is inserted in between the base and the agreement suffix
- Now the base-final empty V is governed

Step 2

(42) *irt<s>ən* 'turn-<PST>-2SG'

C	V	C	V	C	V	+	C	V	+	C	V
	i	r		t			s			ə	n

(43) *irt<s>ən* 'turn-<PST>-2SG'

					G						
					↓	┌──────────┐					
C	V	C	V	C	V	+	C	V	+	C	V
	i	r		t			s			ə	n

Other cases of non-locality in Khanty verbs

» Recall the floating glide-containing base *ari<j>*- ‘to sing’

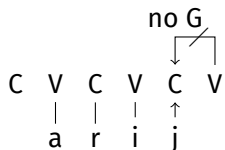
- (45) a. *arij-s* ‘sing-PST[3SG]’
b. *ari-s-ə̃n* ‘sing-PST-2SG’

- » Wrt. schwa in the agreement suffix, it behaves like a C# base, which confirms the presence of a floating glide
- » [Egorov & Tjutjunnikova \(2023\)](#) explain the distribution of the glide by introducing allomorphy
- » In the tense-as-infix model, the glide is base-final only
- » When an overt agreement suffix is attached, it disappears

Floating glide in 3SG

Step 1: glide associates

(46) *ari<j>* 'sing'



Step 2 (why is there no schwa – **arijəs*?)

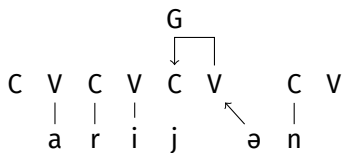
(47) *arij-s* 'sing'



Floating glide in 2SG

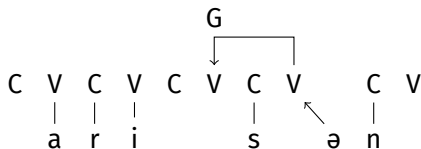
Step 1: glide cannot associate

(48) *ari<j> + ən* 'sing + 2SG'



Step 2

(49) *ari-s-ən* 'sing'



Nominal paradigm

Nominal inflection

Base-number-possession-case

	VC#	əC#	Ca#	Ci#
NOM	xot 'house'	sʌm(ə)t 'birch'	λapkɑ 'shop'	wʌli 'deer'
PL	xot-ət	sʌmt-ət	λapkaj-(ə)t	wʌle-t
P.2PL	xot-ən	sʌmt-ən	λapkaj-(ə)n	wʌle-n
LOC	xot-ən	sʌmət-n	λapkaj-(ə)n	wʌlij-(ə)n

- » Number and possessive markers show similarities with tense
- » Case, on the other hand, behaves like the agreement endings
- » Evidence: glide insertion patterns and schwa-zero alternations

LOC	NPST	POSS.2PL	2SG
wʌlij-(ə)n	arij-λ		N/A
sʌmət-n	irət-λ	sʌmt-ən	irt-λ-ən

Phase boundary

- » I claim that the difference between $-(ə)n$ 'POSS.2PL' and $-(ə)n$ 'LOC' is in the phase boundary
- » Case is higher than possession in the functional sequence (Longobardi 2001)
- » $K > Poss > Num > N$
- » The similarity to the verbal paradigm is that in 3SG verbal forms, the processing of the base shuts down before the tense suffix appears
- » Same with case, but there is a phase boundary in between the suffix and the base

(50) Non-local interaction – tense similar to locative

- Tense: $irt \# ən \rightarrow irt-s-ən$ 'turn-PST-2SG'
- Locative: $sʌmət \# \rightarrow sʌmət-n$ 'birch-LOC'

Derivation: C#

With C# bases without alternating schwa, both POSS.2PL and LOC look the same

(51) *xot-ən* 'house-POSS.2PL'

C	V	C	V	C	V
			←		
x	o	t	ə	n	

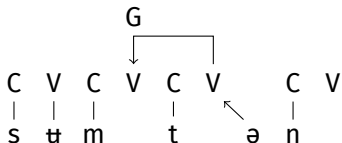
(52) *xot-ən* 'house-LOC'

C	V	C	V	#	C	V
x	o	t		#	ə	n

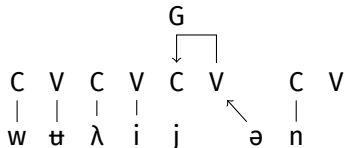
In the locative, schwa could be epenthetic

Derivation: possessive

(53) *sʌmt-ən* 'birch-POSS.2PL'



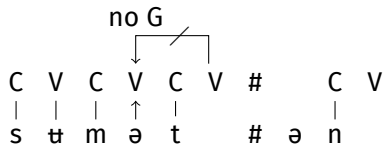
(54) *wʌλe-n* 'reindeer-POSS.2PL'



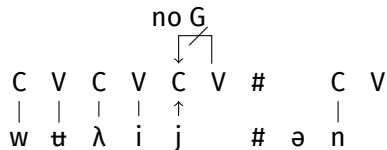
Glide association fails, so vowels coalesce: *wʌλi-ən* → *wʌλe-n*

Derivation: case

(55) *sʈmæt-n* 'birch-Loc'



(56) *wʈlij-(ə)n* 'reindeer-Loc'








Conclusions

- » Two-tiered autosegmental phonology allows for three schwas with distinct patterns of behaviour
- » No allomorphy or distinct phonemes for schwa postulated yet
- » The No Two SCHWAS IN A ROW rule seems like a theoretical dead end
- » Early attachment of agreement markers has morphosyntactic implications, yet to be explored





Credits for data and discussion

- ♥ Ilia Egorov
- ♥ Varvara Tjutjunnikova
- ♥ Anna Moskalëva
- ♥ Daniil Burov
- ♥ Kazym Khanty fieldwork project

References I

-  Egorov, Il'ja & Varvara Tjutjunnikova. 2023. /ə/ v kazymskom xantyjskom: meždu fonetikoju i morfonologijej [Schwa in Kazym Khanty: between phonetics and morphophonology]. Talk at Minor Languages in Big Linguistics.
-  Hyman, Larry M. 2000. Bantu suffix ordering and its phonological consequences. Paper presented at the University of California, Berkeley.
-  Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud. 1990. Constituent structure and government in phonology. *Phonology* 7(1). 193–231.
-  Kozlov, Alexei. 2012. Nejtral'nyj glasnyj šva v teginskom xantyjskom: morfonologija+fonologija [The neutral vowel schwa in Tegi Khanty: morphophonology+phonology]. Handout for a fieldwork group seminar.
-  Longobardi, Giuseppe. 2001. The structure of DPs: some principles, parameters, and problems. In Mark Baltin & Chris Collins (eds.), *The Handbook of Contemporary Syntactic Theory*, 562–603. New Jersey: John Wiley & Sons, Ltd.

References II

-  McCarthy, John. 2010. An introduction to harmonic serialism. *Language and Linguistics Compass* 4(10). 1001–1018.
-  Myler, Neil. 2017. Exceptions to the mirror principle and morphophonological “action at a distance”: The role of “word”-internal phrasal movement and spell out. In Heather Newell et al. (eds.), *The structure of words at the interfaces*, 100–125. Oxford: Oxford University Press.
-  Scheer, Tobias. 2004. *A Lateral Theory of Phonology*. Berlin: De Gruyter Mouton.
-  Solovar, Valentina Nikolaevna. 2014. *Xantyjsko-russkij slovar' (kazymskij dialekt) [Khanty-Russian dictionary (Kazym dialect)]*. Tjumen': OOO "FORMAT".

Glossing abbreviations

1 first person

2 second person

3 third person

DU dual

GEN genitive

LOC locative

NOM nominative

NPST non-past

P possessive

PL plural

POSS possessive

PST past

SG singular