Phonology of East Caucasian languages

George Moroz

Linguistic Convergence Laboratory, NRU HSE

21 October 2020
part of the online course on East Caucasian languages

last version of the presentation is available here: tinyurl.com/y2foe09g

Outline of the talk

Introduction

Data

Methods

Inventory size

Frequency analysis

Compare to other databases

Suprasegmental features

About phonology of East Caucasian languages

- phonology
  - [Kibrik and Kodzasov 1990]
  - The indigenous languages of the Caucasus: [Job and Smeets 1994], [Smeets 1994]
  - Yaziki mira: [Alekseev et al. 2001] and predecessor projects
  - [Grawunder 2017]
  - The Oxford Handbook of Languages of the Caucasus: [Beguš 2021], [Boris 2021a], [Boris 2021b]
  - Handbook of Caucasian Languages: [Koryakov and Maisak 202]
  - and phonological part in more general descriptions of East Caucasian languages
- historical-comparative phonetics
  - S. M. Gasanova, G. Kh. Ibragimov, P. T. Magomedova, U. A. Meylanova, B. B. Talibov, S. M. Khaydakov, A. E. Alekseev, Ju. D. Desheriev, B. K. Gigineishvili, D. S. Imnaishvili, J. Nichols, S. L. Nikolayev, S. A. Starostin, A. S. Kasyan, N. A. Ardoteli, O. A. Mudrak

Outline of the talk

Introduction

Data

Methods

Inventory size

Frequency analysis

Compare to other databases

Suprasegmental features

Languages

- Nakh
  - Chechen
  - Ingush
  - Tsova-Tush (Bats, Batsbi)
- Tsezic (Didoic)
  - Tsez
  - Hinuq
  - Bezhta
  - Hunzib
  - Khwarshi-Inkhoqwari
- Avar–Andic
  - Avar
  - Andi
  - Botlikh
  - Godoberi
  - Chamalal
  - Tindi
  - Bagvalal
  - Akhvakh
- Dargwa
- Lak
- Lezgic
  - Archi
  - Tabasaran
  - Aghul
  - Lezgian
  - Udi
  - Budukh
  - Kryz
  - Rutul
  - Tsakhur
- Khinalug

Languages not present in [Kibrik and Kodzasov 1990]

- Nakh
  - Chechen
  - Ingush
  - Tsova-Tush (Bats, Batsbi)
- Tsezic (Didoic)
  - Tsez
  - Hinuq
  - Bezhta
  - Hunzib
  - Khwarshi-Inkhoqwari
- Avar–Andic
  - Avar
  - Andi
  - Botlikh
  - Godoberi
  - Chamalal
  - Tindi
  - Bagvalal
  - Akhvakh
- Dargwa
- Lak
- Lezgic
  - Archi
  - Tabasaran
  - Aghul
  - Lezgian
  - Udi
  - Budukh
  - Kryz
  - Rutul
  - Tsakhur
- Khinalug

Languages not present in [Kibrik and Kodzasov 1990]

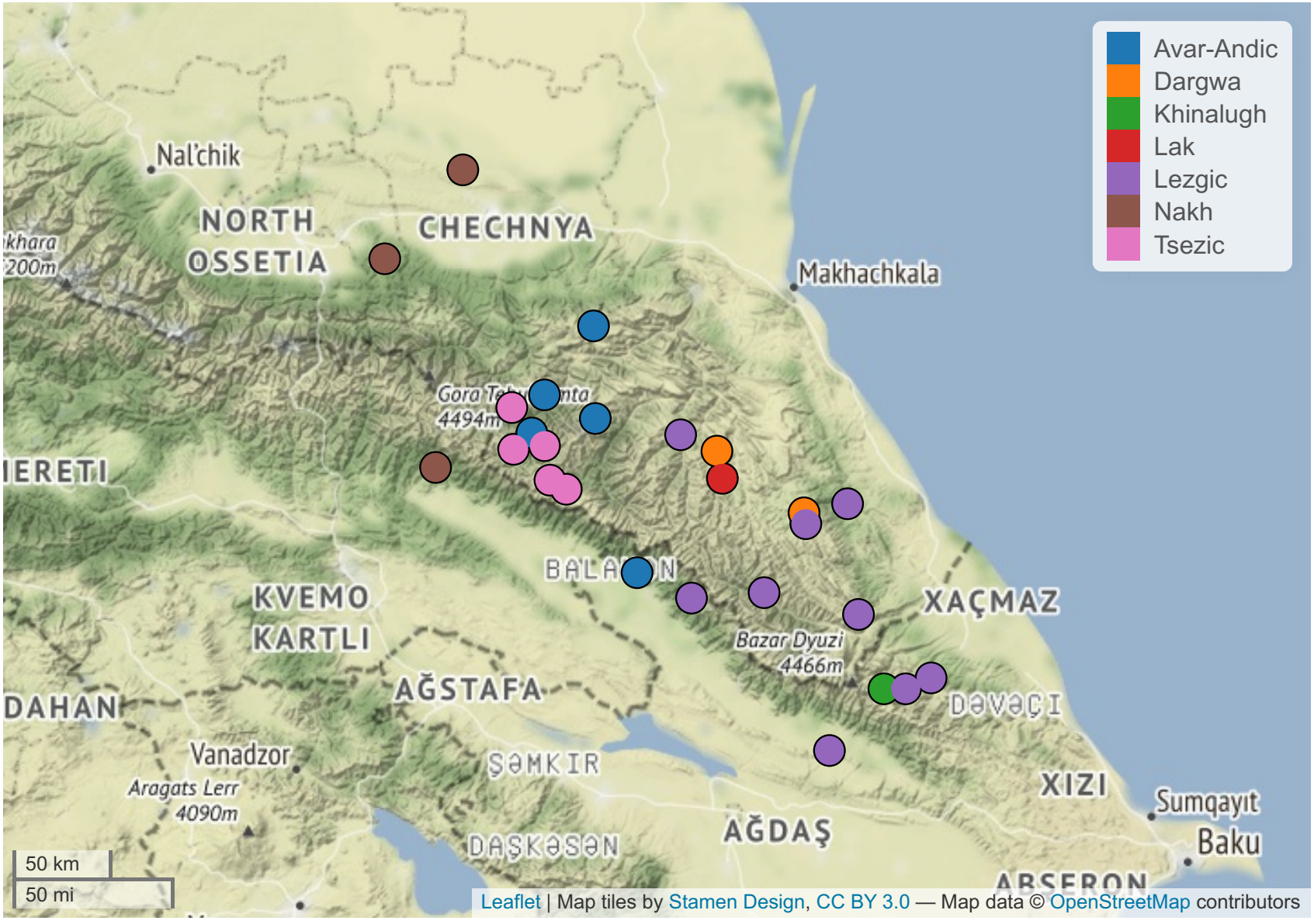
- Nakh, Chechen
- Nakh, Ingush
- Nakh, Tsova-Tush [Holisky and Gagua 1994]
- Dargwa, Mehweb [Moroz 2019]
- Andic, Botlikh
- Andic, Godoberi
- Andic, Bagvalal

Data

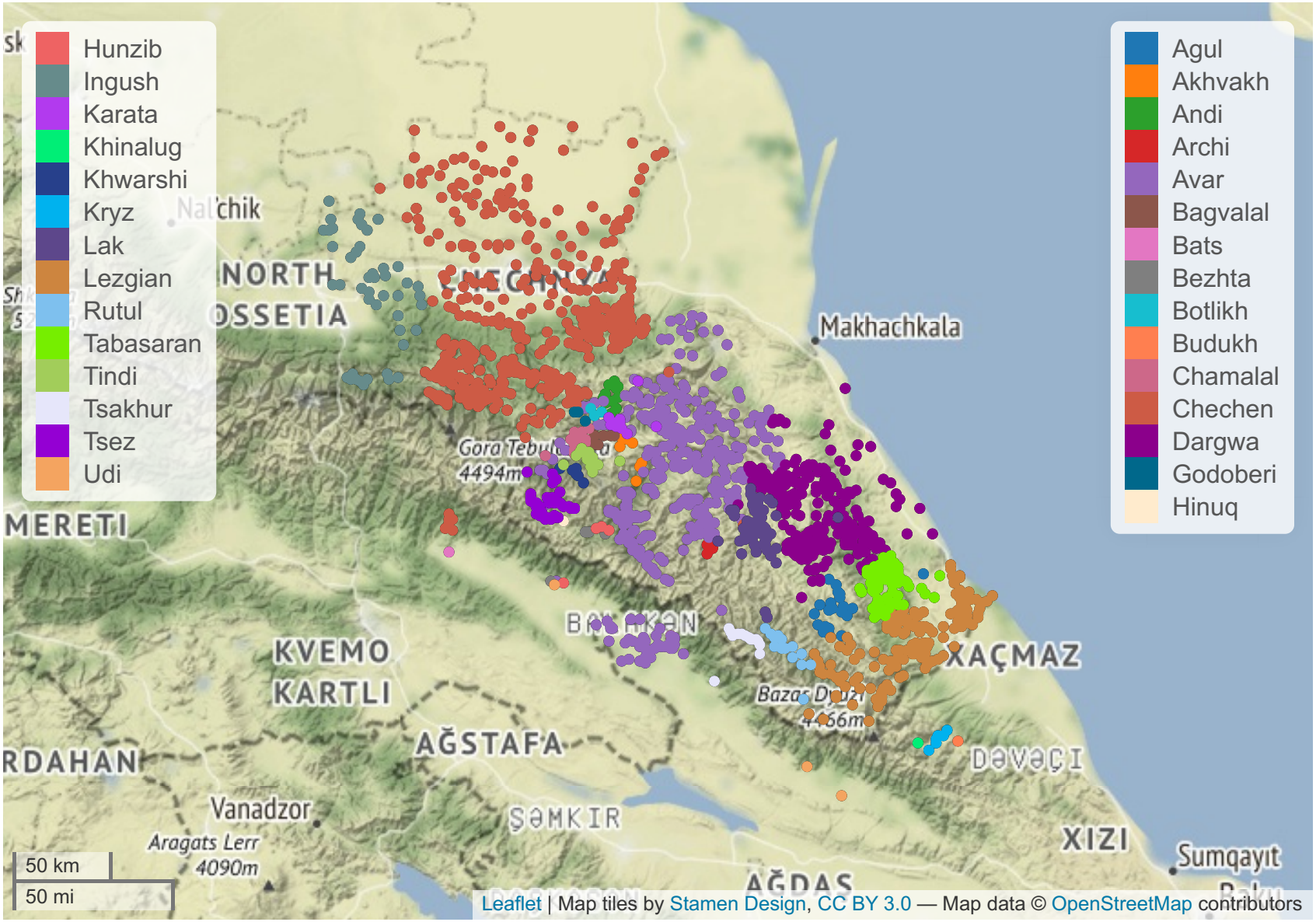
1651 × 9 table with the following columns:

- id
- branch
- language
- segments_IPA
- segments_source
- sound_type
- source
- comments
- glottocode

Data (language view)



Data (village view)



Outline of the talk

Introduction

Data

Methods

Inventory size

Frequency analysis

Compare to other databases

Suprasegmental features

Phonology of East Caucasian languages can be preseneted like this…

	Stops				Fricatives			Affricates				Son.
	Voiced	Voiceless			Voiced	Voiceless		Voiced	Voiceless			
			Gem.	Ej.			Gem.			Gem.	Ej.	
Lab.	b б	p п										w в m м
Den.	d д	t т	t: тт	t’ тI	z з	s с	s: сс		с ц	с: цц	с’ цI	n н
Alv.					ž ж	š ш	š: щ	ǯ дж	č ч	č: чч	č’ чI	r р
Lat.						ɬ лъ	ɬ: лълъ			ƛ лI	ƛ’ кь	l л
Pal.												j й
Vel.	g г	k к	k: кк	k’ кI			x хь					
Uv.		q хъ		q’ къ	ʁ гъ	χ х	χ: хх					
Ph.					ʕ гI	ħ хI						

Phonology of East Caucasian languages can be preseneted like this…

	Stops					Fricatives		Affricates			Son.
			plosives			fricatives		affricates			sonorant
			asp	ej	vcd	asp	vcd	asp	ej	vcd	
Lab.	labial		p		b						w, m
Den.	dental	-gem	t	t’	d	s	z	c	c’		n
		+gem				s:		c:	c’:		
Alv.	alveolar	-gem				š	ž	č	č’	ǯ	r
		+gem				š:		č:	č’:		
Lat.	lateral					ɬ (ɬ:)		ƛ	ƛ’		l
Pal.	palatal					ç, ç^{w}					j
Vel.	velar	-gem	k, k^{w}	k’, k’^{w}	g, g^{w}						
Uv.		+gem	k:, k:^{w}	k’:, k’^{w}:							
Ph.	uvular	-gem		q’, q’^{w}		χ, χ^{w}	ʁ, ʁ^{w}	qχ, qχ^{w}			
		+gem				χ:					

Phonology of East Caucasian languages can be preseneted like this…

	Stops				Fricatives		Affricates			Son.
		plosives			fricatives		affricates			sonorant
		asp	ej	vcd	asp	vcd	asp	ej	vcd	
Lab.	labial									
Den.	dental									
Alv.	alveolar									
Lat.	lateral									
Pal.	palatal									
Vel.	velar									
Uv.	uvular	-gem		q’, q’^{w}		χ, χ^{w}	ʁ, ʁ^{w}	qχ, qχ^{w}		
Ph.		+gem				χ:				

		labial	dental	alveolar	palatal	velar		uvular		pharyngeal		epig
						-lab	+lab	-lab	+lab	-lab	+lab	-lab
plosive	+v	b	d			g	g^{w}					
	-v	p	t			k	k^{w}	q	q^{w}			(ʔ)
	ej	p’	t’			k’	k’^{w}	q’	q’^{w}			
fricative	+v		z	ž		(ɣ)		ʁ	ʁ^{w}			
	-v		s	š		x	x^{w}	χ	χ^{w}	ħ	ħ^{w}	(ʜ)
affricate	+v		(ʒ)	(ǯ)								
	-v		c	č								
	ej		c’	č’								
sonorant		m, w	n, l, r		j							

This way is...

- boring,

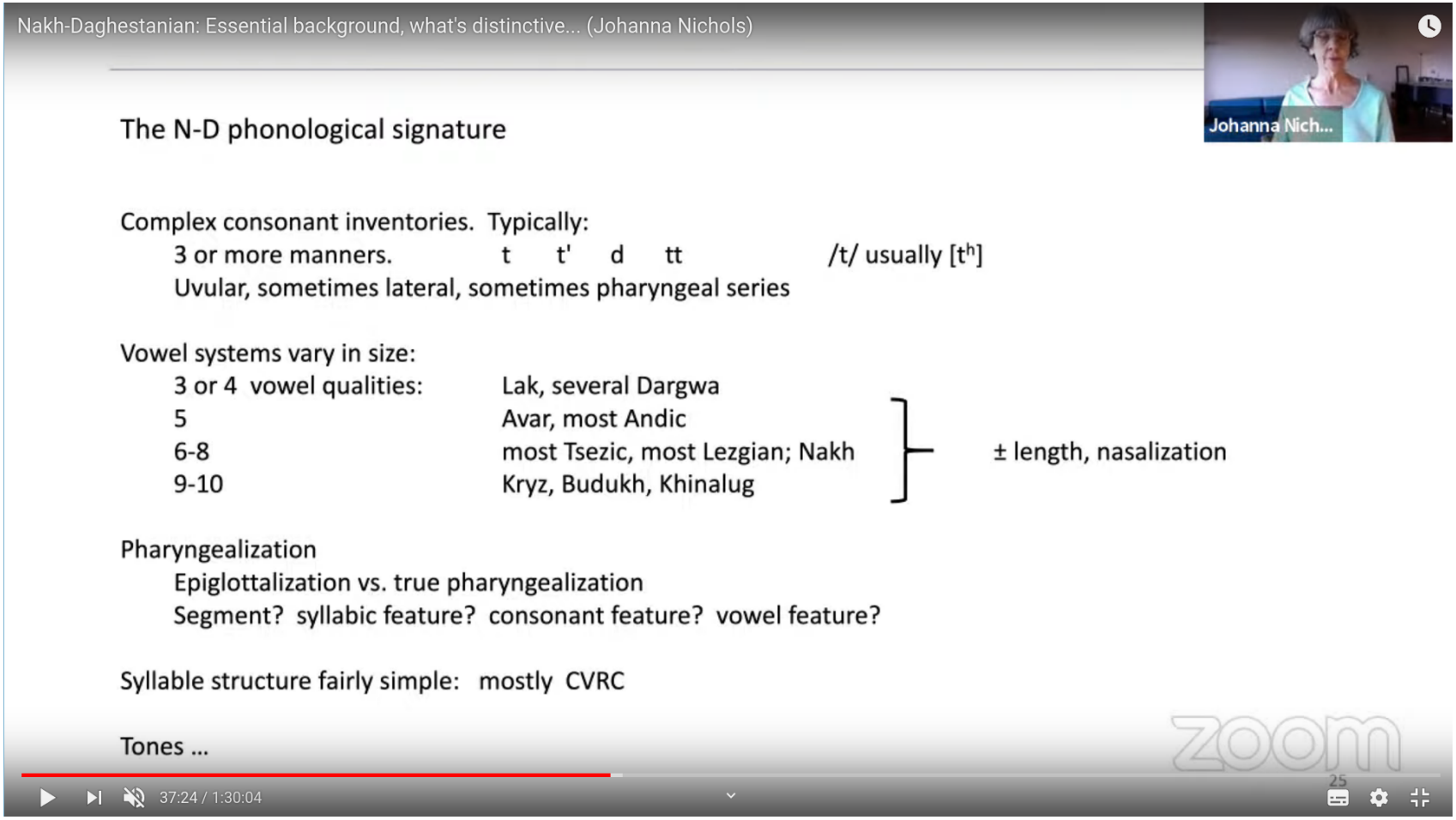
This way is…

- boring,
- mechanical,

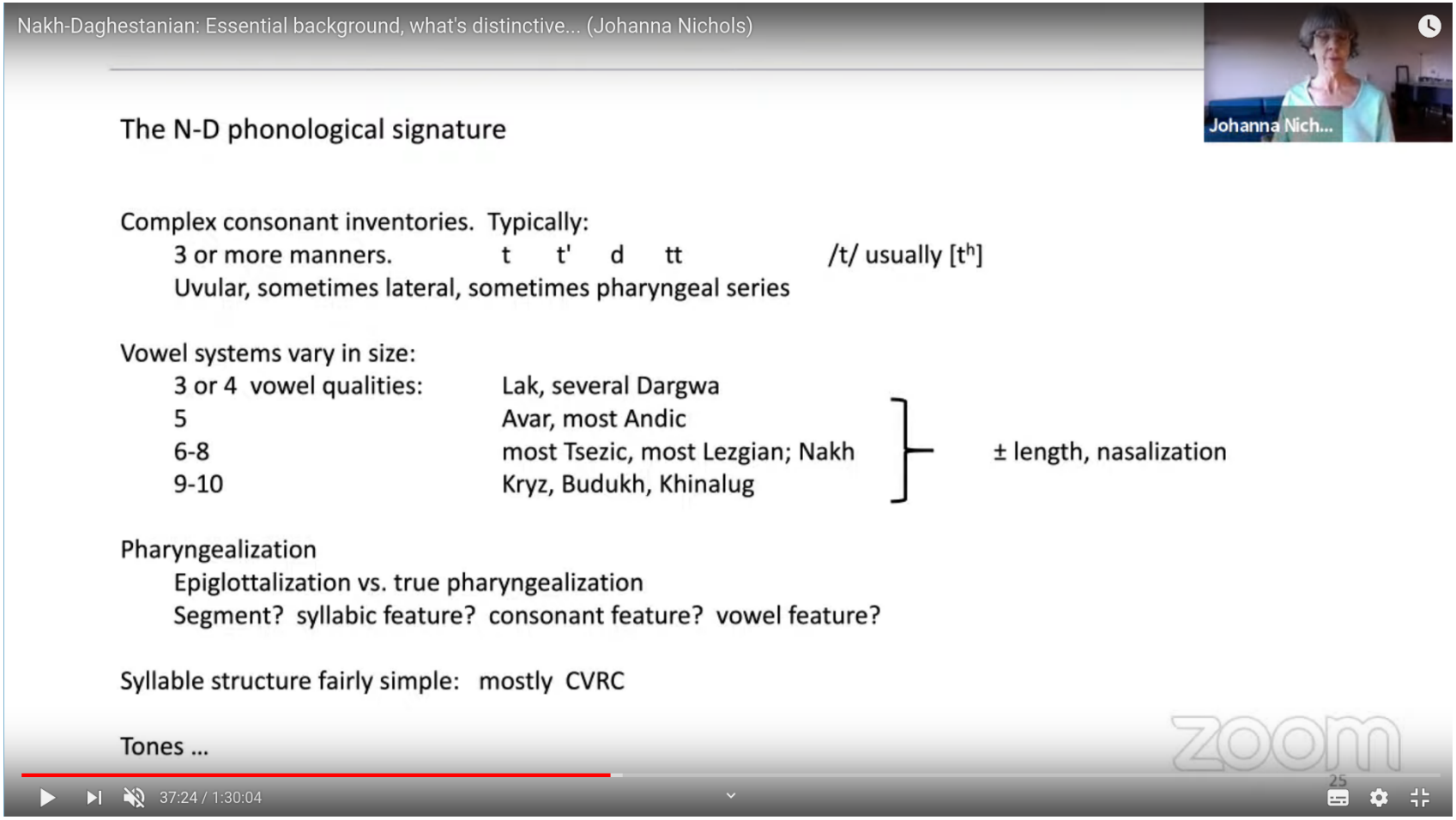
This way is...

- boring,
- mechanical,
- it is hard to make generalisations (at least, on the fly)

Phonology of East Caucasian languages can be preseneted like this…



Phonology of East Caucasian languages can be preseneted like this…



(If you have not seen it, go and wath our previous lecture!)

Those ways are…

- To list all inventories
  - boring,
  - mechanical,
  - hard to make generalisations (at least, on the fly)
- To present generalisations
  - restricted to peculiarities,

Those ways are…

- To list all inventories
  - boring,
  - mechanical,
  - hard to make generalisations (at least, on the fly)
- To present generalisations
  - restricted to peculiarities,
  - you need to be a specialist

Those ways are...

- To list all inventories
  - boring,
  - mechanical,
  - hard to make generalisations (at least, on the fly)
- To present generalisations
  - restricted to peculiarities,
  - you need to be a specialist
- May be those conclusion could be data driven?

Those ways are...

- To list all inventories
  - boring,
  - mechanical,
  - hard to make generalisations (at least, on the fly)
- To present generalisations
  - restricted to peculiarities,
  - you need to be a specialist
- May be those conclusion could be data driven?
  - some of them can be visualised with a map
  - some of them can be gahtered via combinatoric analysis

Outline of the talk

Introduction

Data

Methods

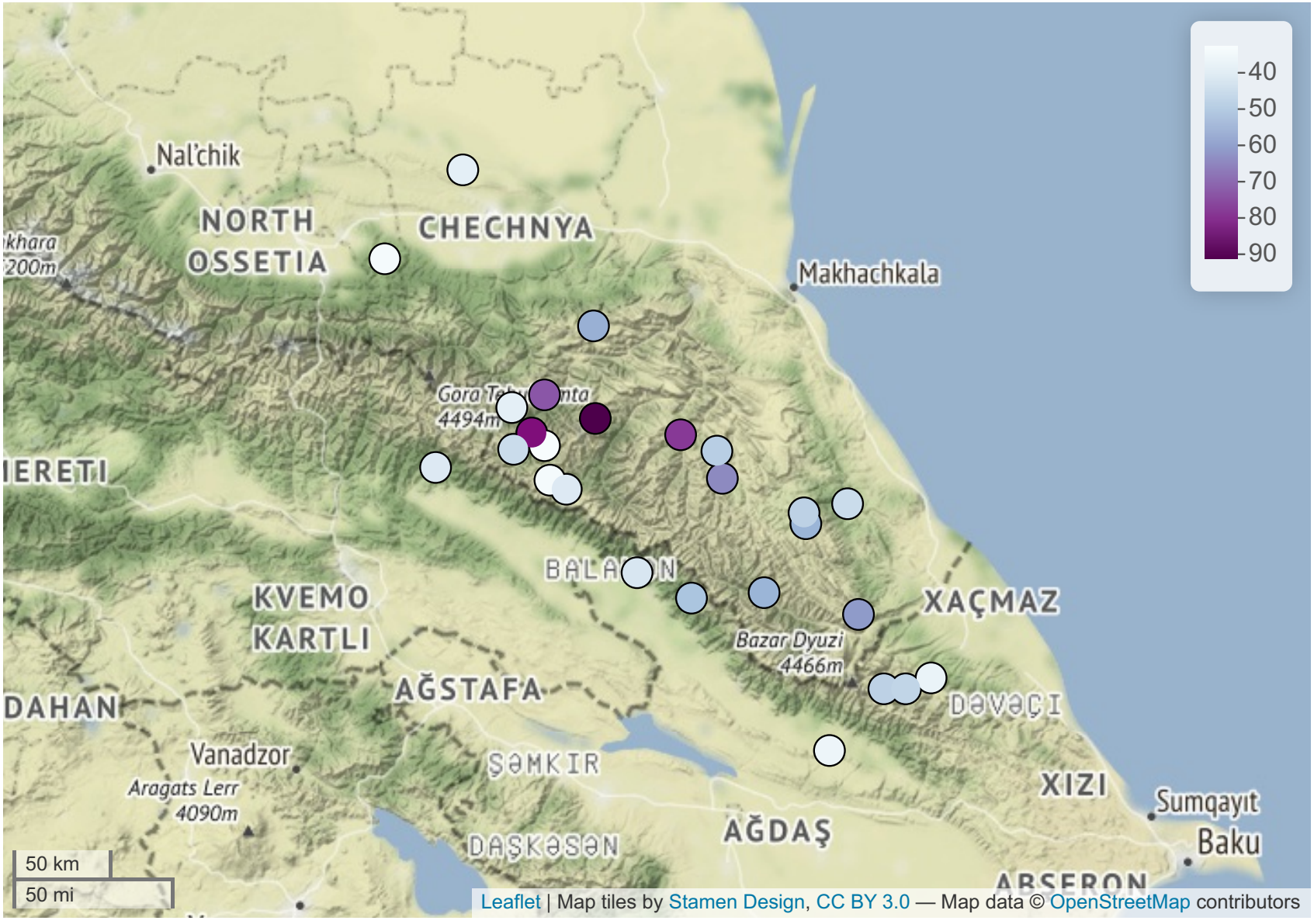
Inventory size

Frequency analysis

Compare to other databases

Suprasegmental features

Consonants


40
50
60
70
80
90
Nal'chik
NORTH OSSETIA
CHECHNYA
Makhachkala
Gora Tebulosmta
4494m
IERETI
BALAKƏN
KVEMO KARTLI
XAÇMAZ
Bazar Dyuzi
4466m
AĞSTAFA
DƏVƏÇI
DAHAN
Vanadzor
ŞƏMKIR
XIZI
Aragats Lerr
4090m
Sumqayıt
AĞDAŞ
Baku
DAŞKƏSƏN
ABŞERON
50 km
50 mi
Leaflet | Map tiles by Stamen Design, CC BY 3.0 — Map data © OpenStreetMap contributors

Consonants

Nal'chik
NORTH OSSETIA
CHECHNYA
Makhachkala
39
34
57
73
39
91
78
49
64
33
41
41
45
56
42
52
56
61
XAÇMAZ
KVEMO KARTLI
Bazar Dyuzi 4466m
37
AĞSTAFA
DAHAN
DƏVƏÇI
36
Vanadzor
ŞƏMKIR
Aragats Lerr 4090m
XIZI
Sumqayıt
AĞDAŞ
Baku
DAŞKƏSƏN
50 km
50 mi
Leaflet | Map tiles by Stamen Design, CC BY 3.0 — Map data © OpenStreetMap contributors

Consonants

- Akhwakh (Andic) – 91
- Tindi (Andic) – 84
- Chamalal (Andic) – 73
- Archi (Lezgic) – 78
- …
- Ingush (Nakh) – 34
- Bezhta (Tsezic) – 34
- Inkhoqwari (Tsezic) – 33

Consonants: Akwakh

b	d			dʒ				g					
	d^w			$dʒ^w$				g^w					
p^h	t^h	ts	tsː	tʃ	tʃː	tɬ	tɬː	k	kː	q	qː		ʔ
	t^{hw}	ts^w	$tsː^w$	$tʃ^w$	$tʃː^w$	$tɬ^w$	$tɬː^w$	k^w	$kː^w$	q^w	$qː^w$		$ʔ^w$
p’	t’	ts’	tsː’	tʃ’	tʃ’ː	tɬ’	tɬ’ː	k’	k’ː	q’	q’ː		
	$t’^w$	$ts’^w$	$tsː’^w$	$tʃ’^w$	$tʃ’ː^w$	$tɬ’^w$	$tɬ’ː^w$	$k’^w$	$k’ː^w$	$q’^w$	$q’ː^w$		
		z		ʒ						ʁ		ʕ	
		z^w		$ʒ^w$						$ʁ^w$			
		s	sː	ʃ	ʃː	ɬ	ɬː	x	xː	χ	χː	ʜ	h
		s^w	$sː^w$	$ʃ^w$	$ʃː^w$	$ɬ^w$	$ɬː^w$	x^w	$xː^w$	$χ^w$	$χː^w$		h^w
m	n												
w	r					l		j					

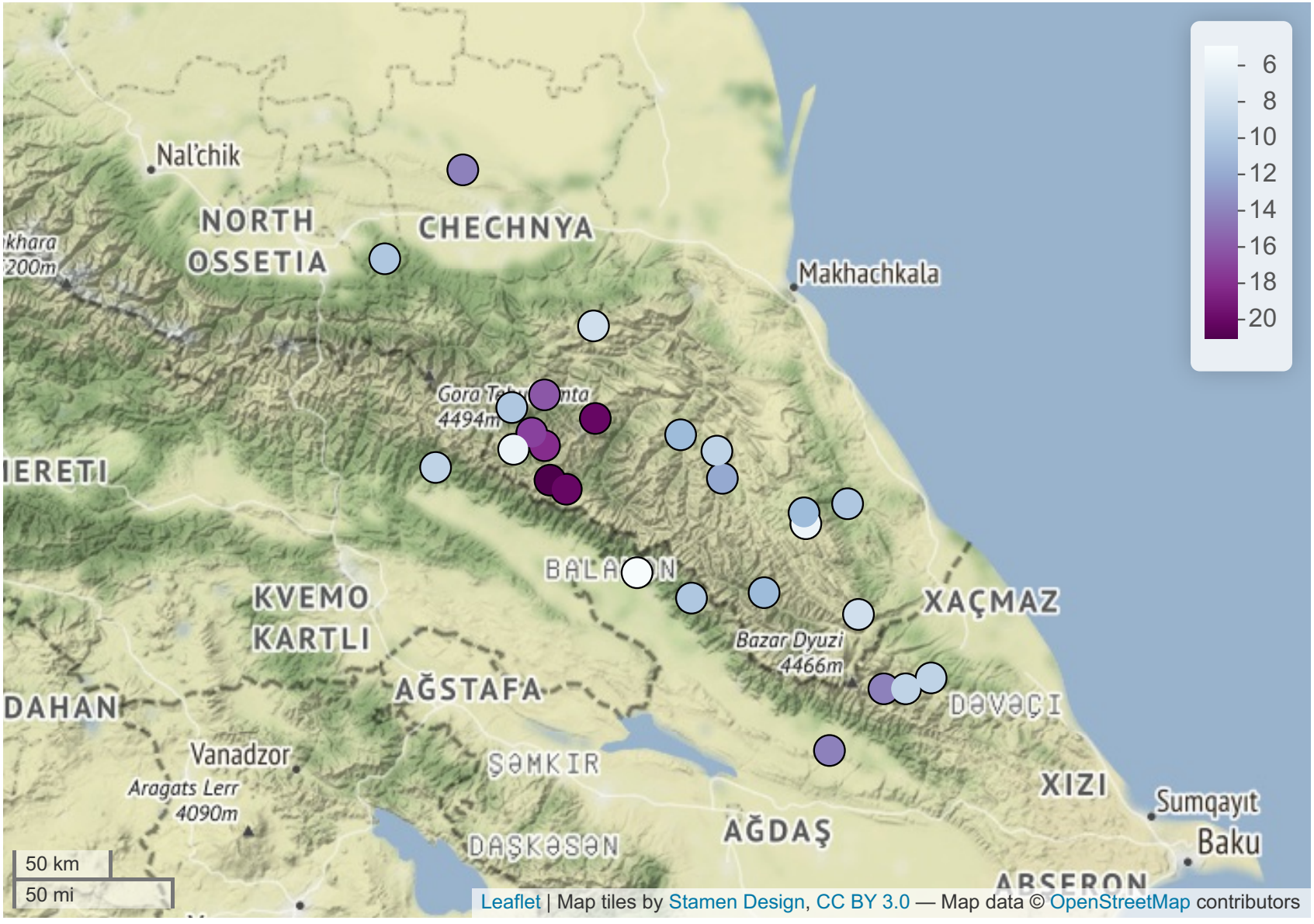
- gemenation
- labialisation

Consonants: geminate ejectives

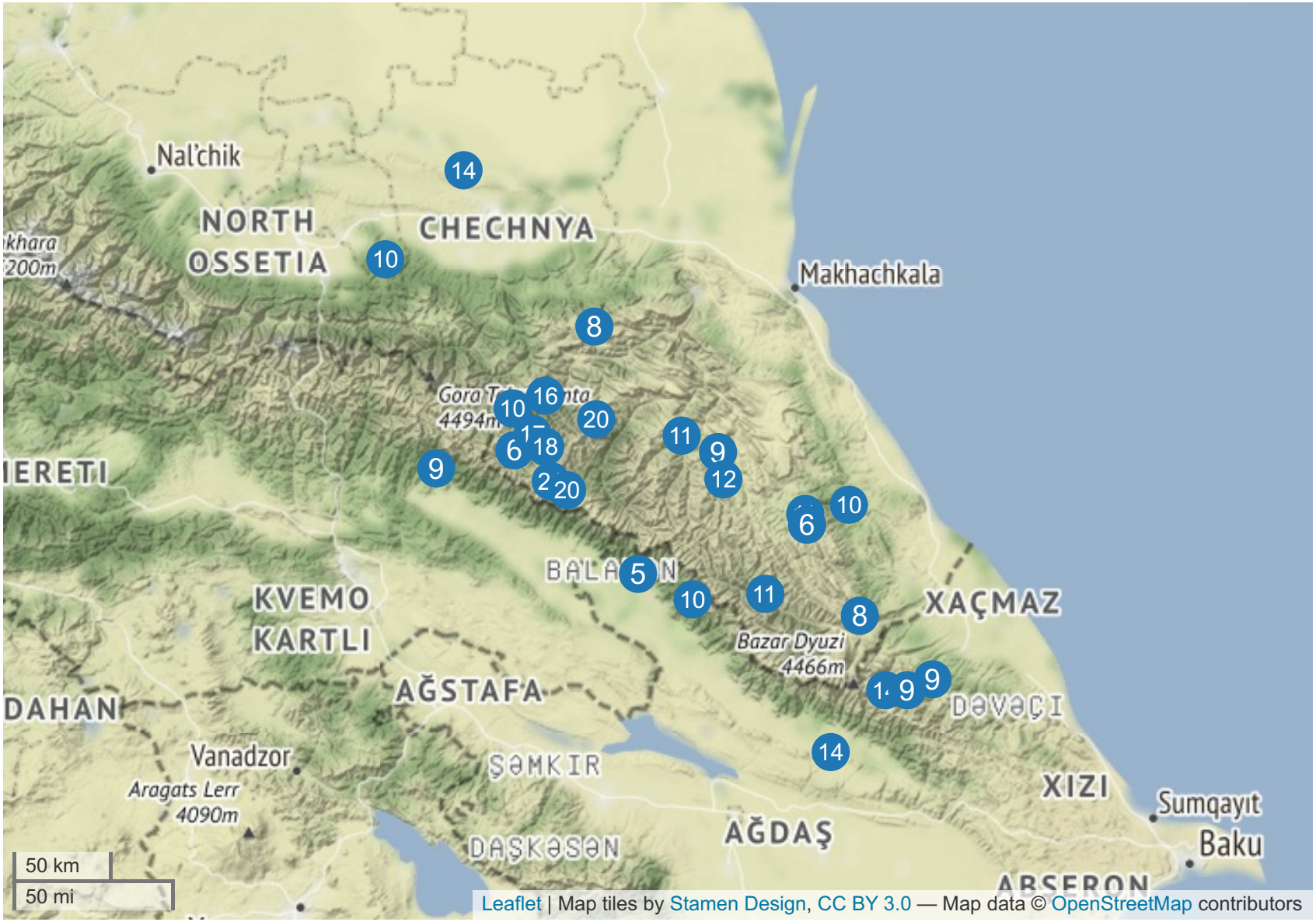
Consonants: Inkhoqwari

b	d		dʒ		g			
pʰ	tʰ	ts	tʃ	tɬ	k	q		ʔ
pʼ	tʼ	tsʼ	tʃʼ	tɬʼ	kʼ	qʼ		
		z	ʒ			ʁ	ʕ	
		s	ʃ	ɬ		χ	ʜ	h
m	n							
w	r			l	j			

Vowels


6
8
10
12
14
16
18
20
Nal'chik
NORTH OSSETIA
CHECHNYA
Makhachkala
4494m
BALAKƏN
KVEMO KARTLI
AĞSTAFA
XAÇMAZ
Bazar Dyuzi 4466m
DƏVƏÇI
Vanadzor
ŞƏMKIR
Aragats Lerr 4090m
XIZI
Sumqayıt
AĞDAŞ
Baku
DAŞKƏSƏN
50 km
50 mi
Leaflet | Map tiles by Stamen Design, CC BY 3.0 — Map data © OpenStreetMap contributors

Vowels


Nal'chik
NORTH OSSETIA
CHECHNYA
Makhachkala
Gora Tebulosmta 4494m
IERETI
KVEMO KARTLI
AĞSTAFA
DAHAN
Vanadzor
Aragats Lerr 4090m
ŞƏMKIR
DAŞKƏSƏN
AĞDAŞ
BALAKƏN
XAÇMAZ
Bazar Dyuzi 4466m
DƏVƏÇI
XIZI
Sumqayıt
Baku
ABŞERON
50 km
50 mi
Leaflet | Map tiles by Stamen Design, CC BY 3.0 — Map data © OpenStreetMap contributors

Vowels

- Bezhta (Tsezic) – 21
- Akhvakh (Andic) – 20
- Hunzib (Tsezic) – 20
- …
- Aghul (Lezgic) – 6
- Tsez (Tsezic) – 6
- Avar – 5

Vowels

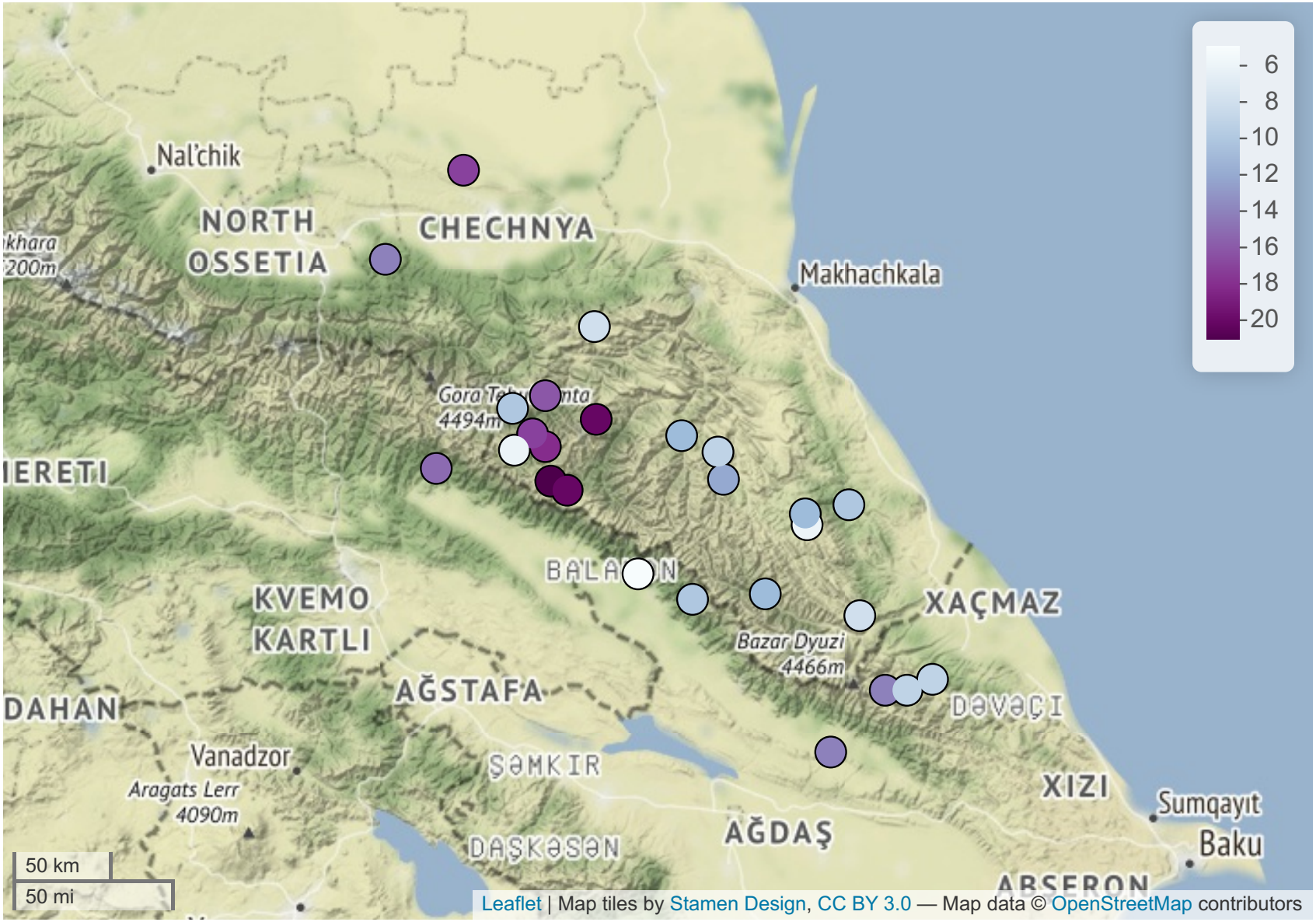
- Bezhta

i	iː	iⁿ	y	yⁿ				u	uː	uⁿ
e		eⁿ	ø					o	oː	oⁿ
æ	æː	æⁿ								
					a	aː	aⁿ			

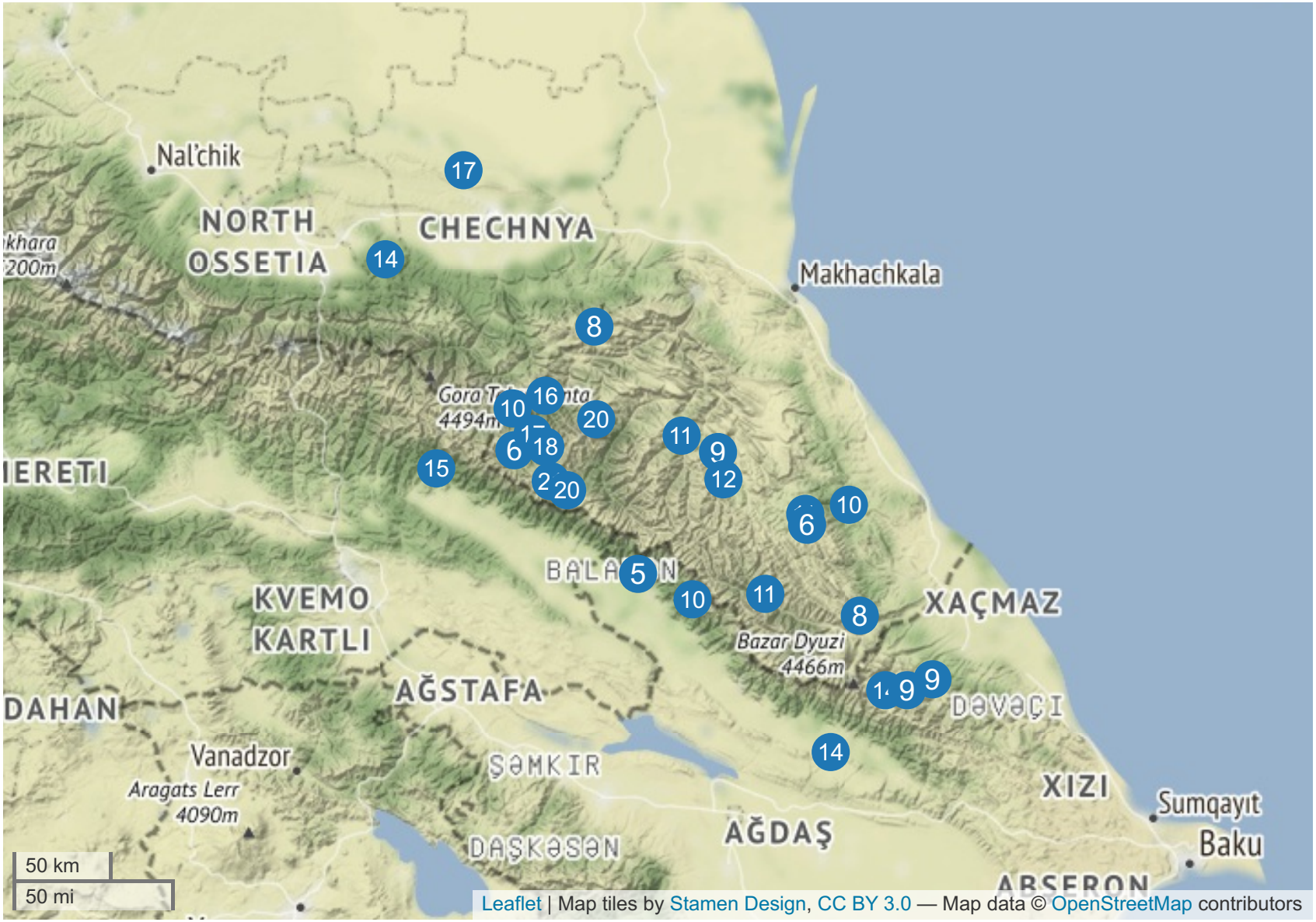
- Avar

i		u
e		o
	a	

Vowels + diphthongs



Vowels + diphthongs



Outline of the talk

Introduction

Data

Methods

Inventory size

Frequency analysis

Compare to other databases

Suprasegmental features

Segments present in all languages

b	d					g	
p^h	t^h						ʔ
		z	ʒ				
		s	ʃ				h
m	n						
	r			l	j		

- Where are the famouse ejectives?

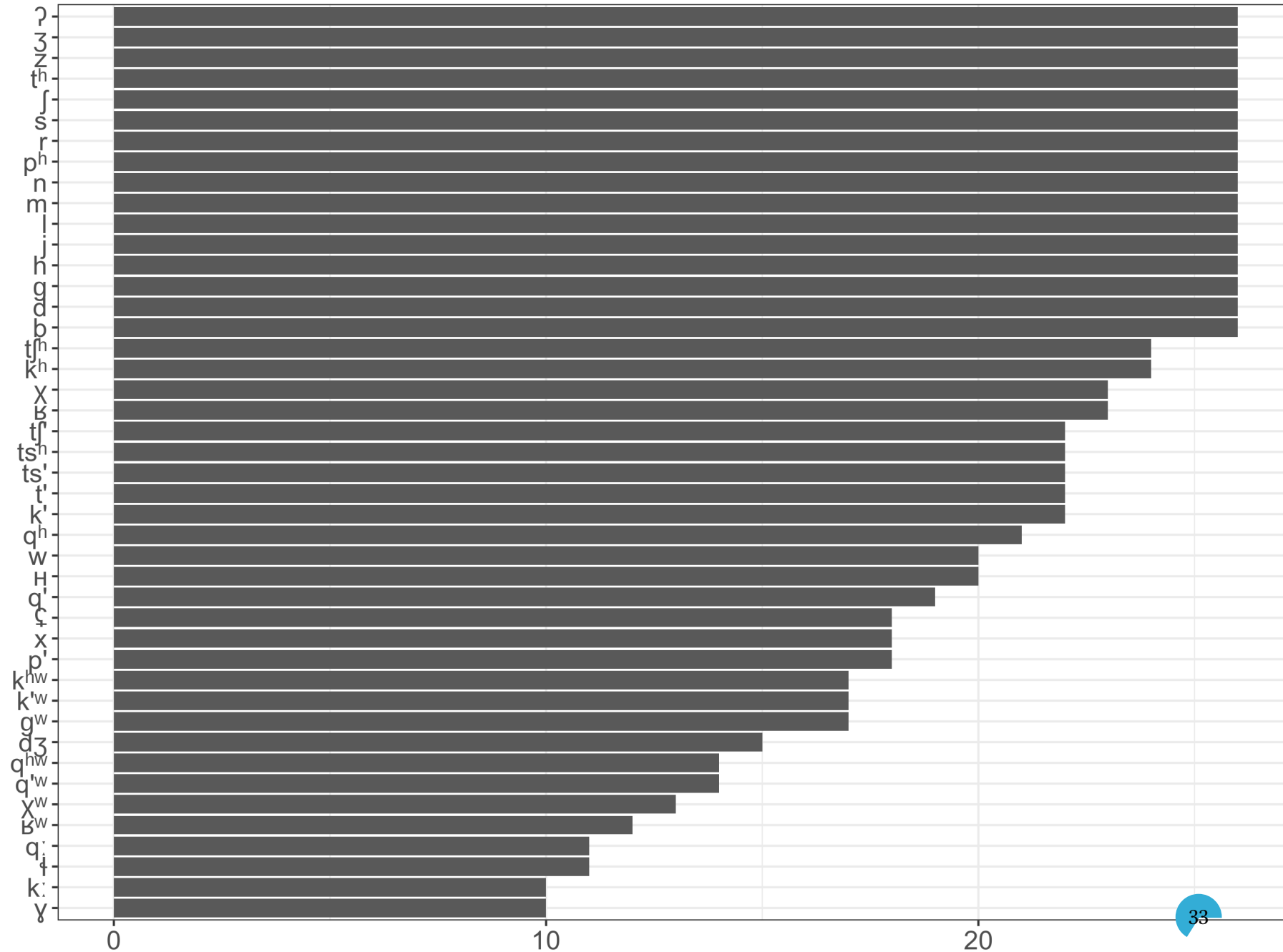
Segments present in all languages

b	d					g	
p^h	t^h						ʔ
		z	ʒ				
		s	ʃ				h
m	n						
	r			l	j		

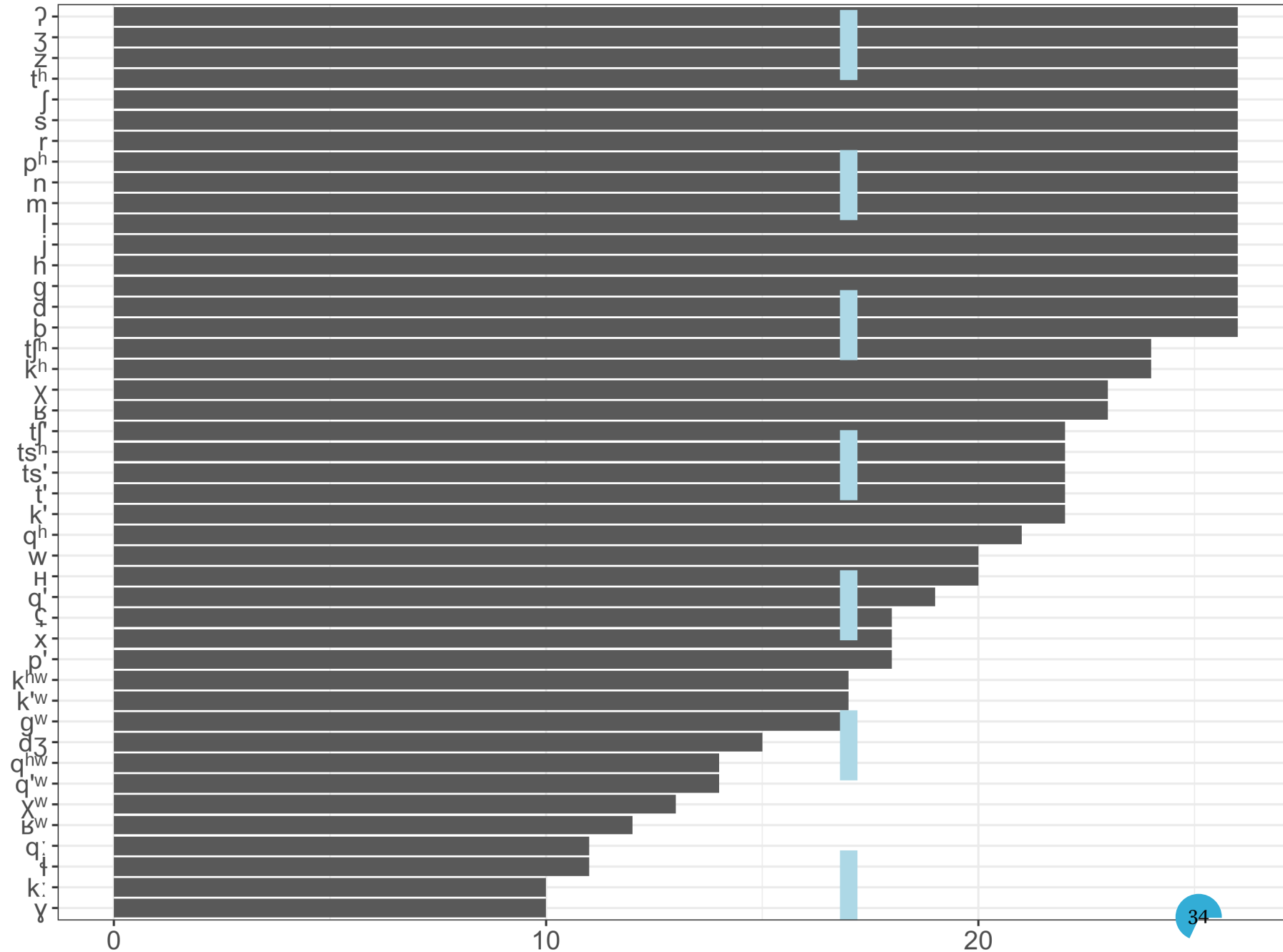
- Where are the famouse ejectives? Udi:

b	d	dz	dʒ			g		
p^h	t^h	ts^h	$tʃ^h$			k^h	q^h	
p	t	ts	tʃ			k	q	ʔ
v		z	ʒ				ʁ	
f		s	ʃ				χ	h
m	n							
w	r			l	j			

Consonants: typical inventory



Consonants: typical inventory



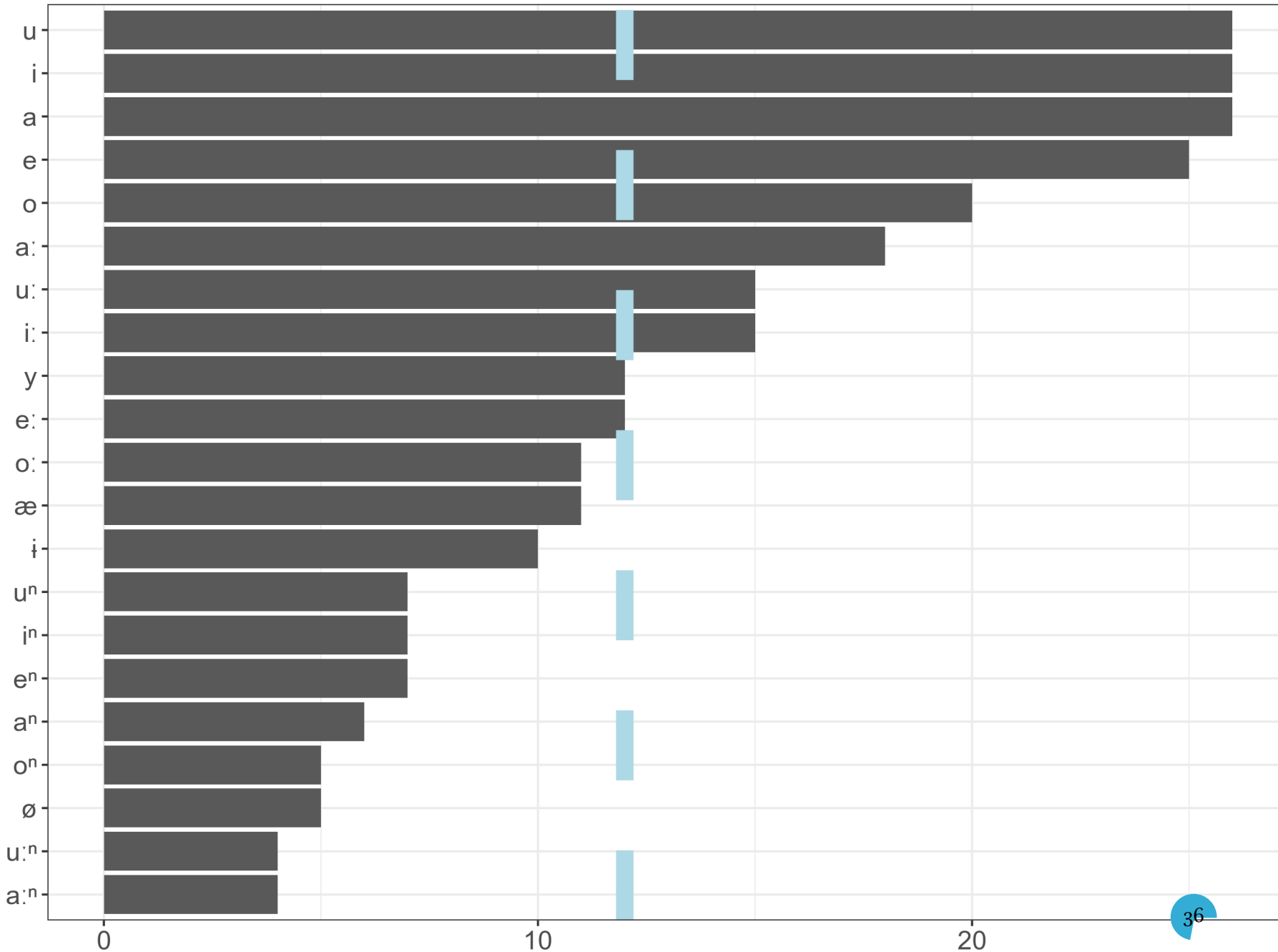
Consonants: typical inventory

b	d					g			
p^h	t^h	ts	$tʃ^h$			k^h	q^h		ʔ
p’	t’	ts’	tʃ’			k’	q’		
		z	ʒ				ʁ	ʢ	
w		s	ʃ			x	χ	ʜ	h
m	n								
	r			l	j				

For more consonants:

- labialisation
- gemination
- laterals
- pharyngealisation

Vowels: typical inventory



Vowels: typical inventory

i	iː			u	uː
e				o	
		a	aː		

For more vowels:

- length
- nasalizsation

Outline of the talk

Introduction

Data

Methods

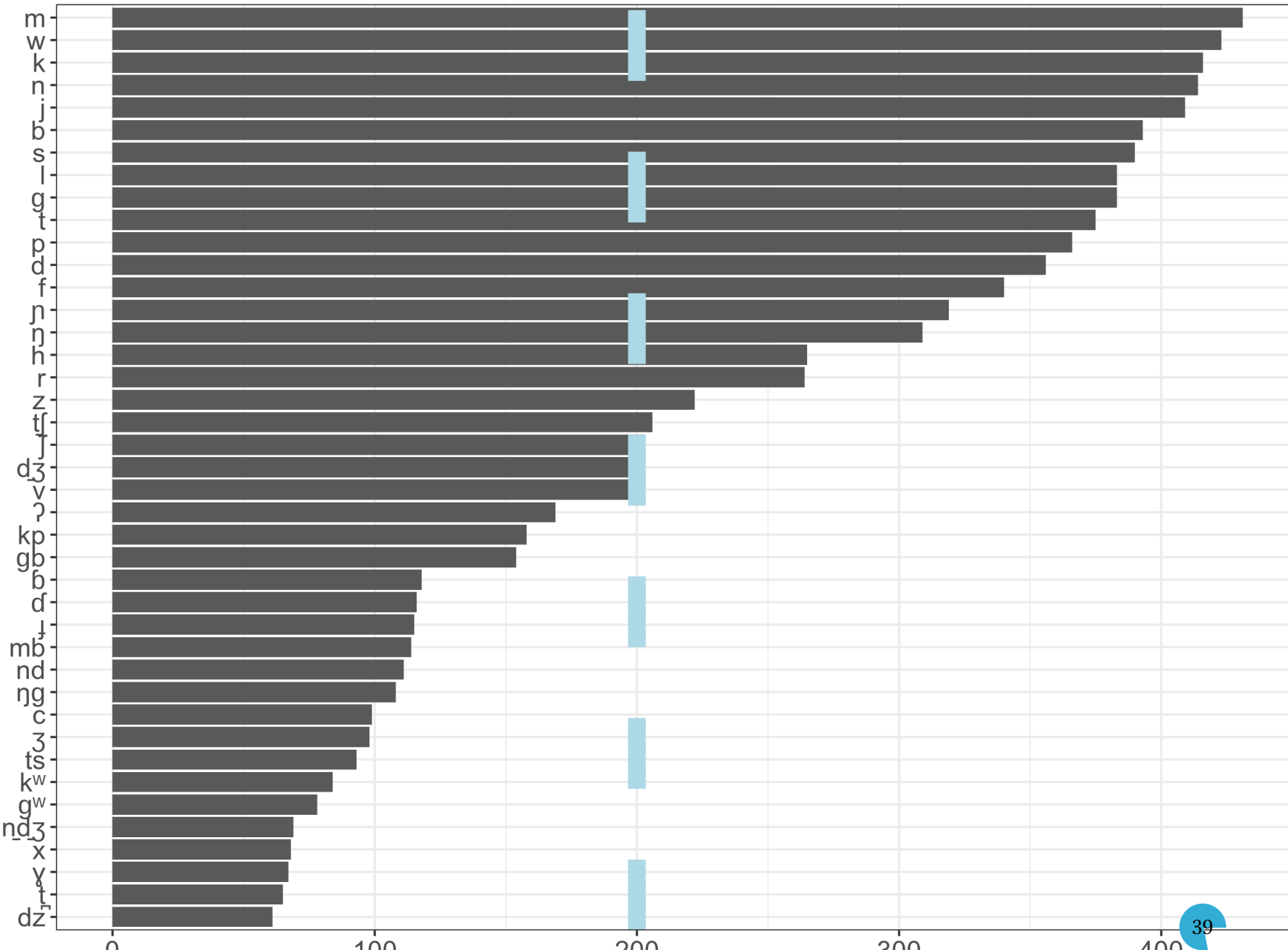
Inventory size

Frequency analysis

Compare to other databases

Suprasegmental features

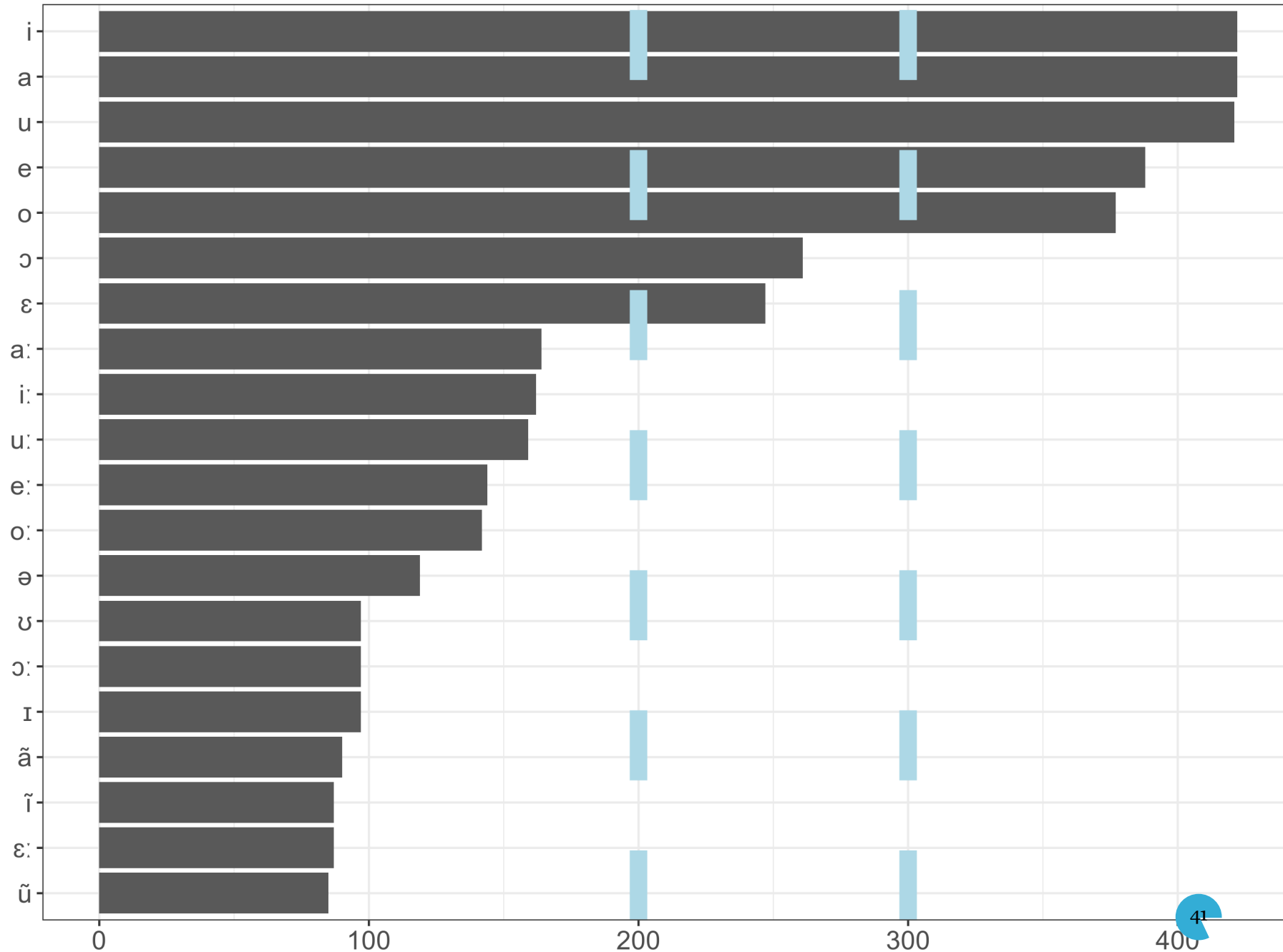
PHOIBLE [Moran and McCloy 2019]: consonants



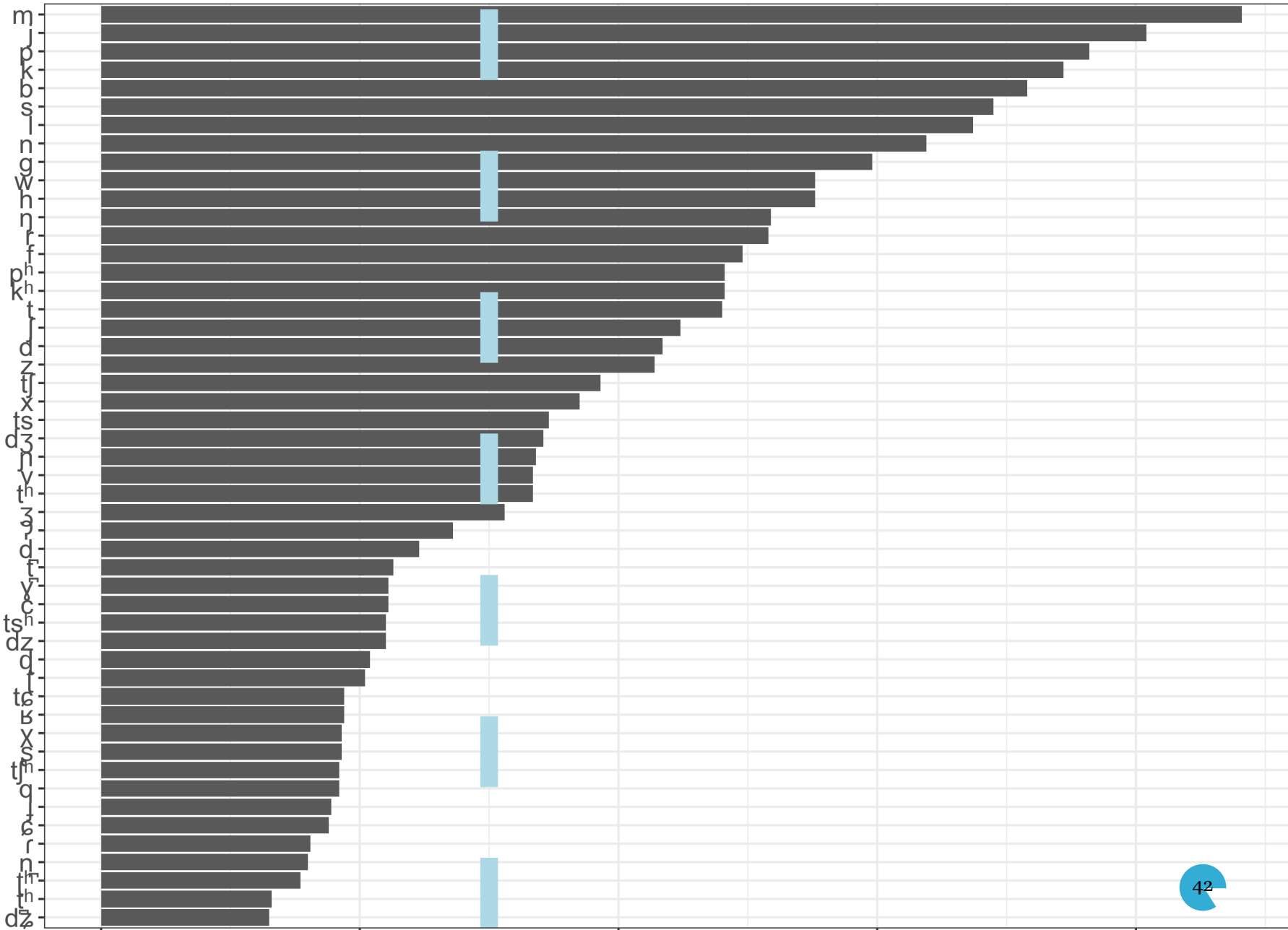
PHOIBLE [Moran and McCloy 2019]: consonants

b	d		dʒ			g	
p	t		tʃ			k	
		z	ʃ				
f		s					h
m	n				ɲ	ŋ	
w	r			l	j		

PHOIBLE [Moran and McCloy 2019]: vowels



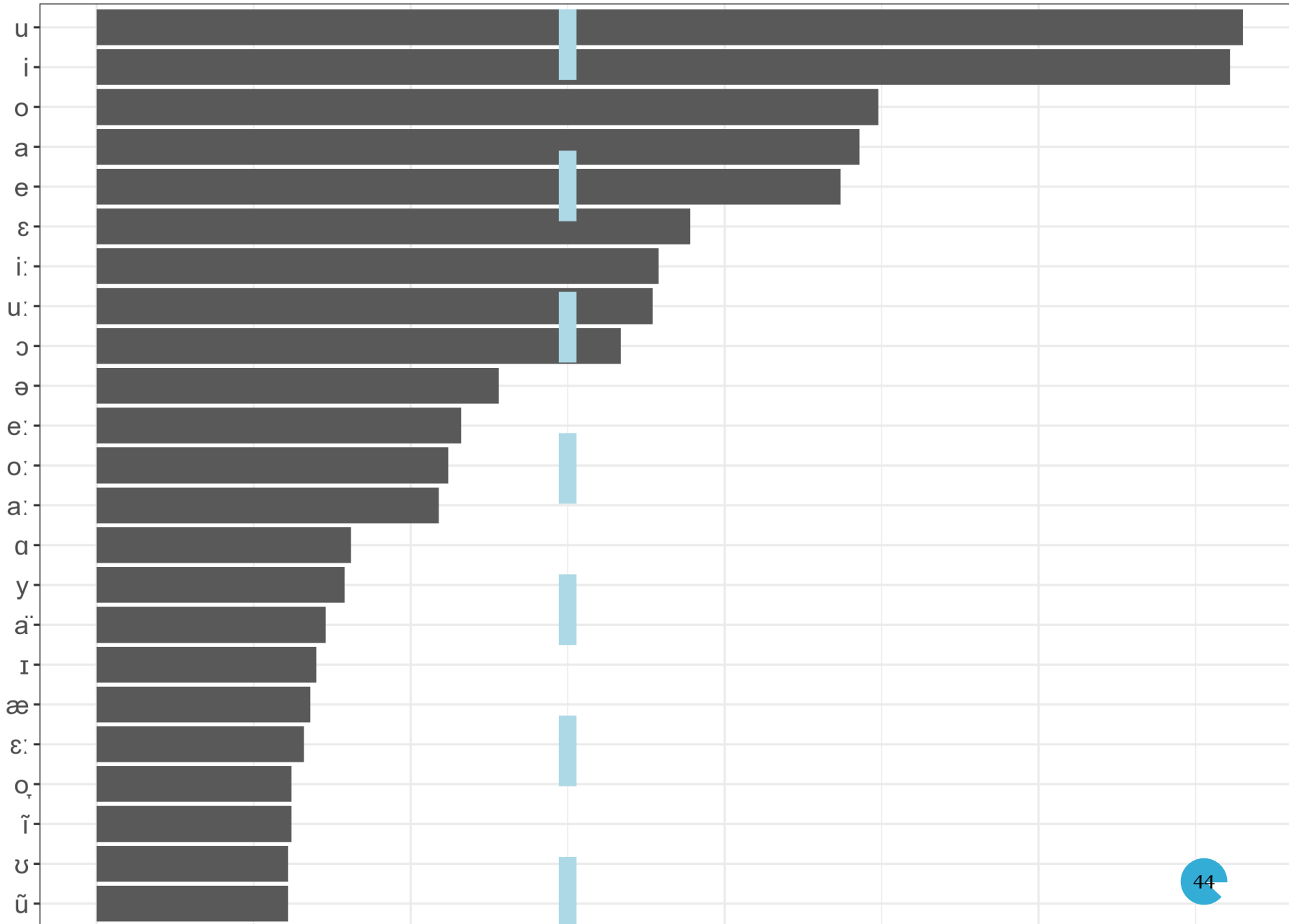
Eurasian phonological inventories [Nikolaev et al. 2015]: consonants



Eurasian phonological inventories [Nikolaev et al. 2015]: consonants

b	d		dʒ			g	
p	t	ts	tʃ			k	
p^h	t^h					k^h	
		z	ʒ				
f	v	s	ʃ			x	h
m	n				ɲ	ŋ	
w	r			l	j		

Eurasian phonological inventories [Nikolaev et al. 2015]: vowels



Outline of the talk

Introduction

Data

Methods

Inventory size

Frequency analysis

Compare to other databases

Suprasegmental features

Syllable structure and suprasegmental features

Syllable structure CVCC, sometimes CVL (L – sonorant or *b*).

A lot of East Caucasian languages has stress:

- some languages has more o less fixed stress (e. g. Mehweb)
- some languages has a tendency to have stress on first three syllables (Andic languages)
- a lot of languages distinguish morphological meanings via stress (e. g. plural marker in Dargic languages)

pharyngealization – mystirius feature (absent only in Andic branch).

d-irʔ-an	d-irɬ-aˤn /d-irʔˤ-an/
npl-gather:ipfv-hab	npl-gather:ipfv-hab
‘gathers them’	‘gathers them’
ħa-d-irʔ-an	нaˤ-d-irɬ-aˤn
neg-npl-gather:ipfv-hab	neg-npl-freeze:ipfv-hab
‘does not gather them’	‘they are not freezing’

Thanks

- to reasercher teams who collected all data
- to Viola Ivanova for helping with digitalisation of [Kibrik and Kodzasov 1990]
- to Linguistic Convergence Laboratory for organising this course

Thanks

- to reasercher teams who collected all data
- to Viola Ivanova for helping with digitalisation of [Kibrik and Kodzasov 1990]
- to Linguistic Convergence Laboratory for organising this course
- and YOU for your attention

Thanks

- to reasercher teams who collected all data
- to Viola Ivanova for helping with digitalisation of [Kibrik and Kodzasov 1990]
- to Linguistic Convergence Laboratory for organising this course
- and YOU for your attention

- if you like presented maps, see `lingtypology` [Moroz 2017]
- this presentation made with `rmarkdown` [Xie et al. 2018]; source code and data for this presentation are availible online: https://tinyurl.com/y2oh9sg8

References

Alekseev, M. E., Starostin, S. A., Klimov, G. A., and Testelets, Y. G. (2001). *Yazyki mira. Kavkazskiye yaziki*. Academia.

Beguš, G. (2021). Segmental phonetics and phonology. In Polinsky, M., editor, *The Oxford Handbook of Languages of the Caucasus*, pages 1–40. Language Science Press.

Boris, L. (2021a). Tone and intonation in languages of the caucasus. In Polinsky, M., editor, *The Oxford Handbook of Languages of the Caucasus*, pages 1–30. Oxford University Press.

Boris, L. (2021b). Word stress in the languages of the caucasus. In Polinsky, M., editor, *The Oxford Handbook of Languages of the Caucasus*, pages 1–30. Language Science Press.

References

Grawunder, S. (2017). The caucasus. In Hickey, R., editor, *The Oxford Handbook of Languages of the Caucasus*, pages 356–395. Cambridge University Press.

Holisky, D. A. and Gagua, R. (1994). Tsova-Tush (Batsbi). *The indigenous languages of the Caucasus*, 4(part 2):147–212.

Job, D. M. and Smeets, R., editors (1994). *The indigenous languages of the Caucasus*, volume 3. Caravan Books.

Kibrik, A. E. and Kodzasov, S. V. (1990). *Sopostavitelnoye izucheniye dagestanskix yazykov: Imya. Fonetika*, volume 2. Moskovskij Gosudarstvennyj Universitet.

Koryakov, Y. and Maisak, T., editors (202_). *Handbook of Caucasian Languages*, volume 2. De Gruyter Mouton, Berlin/New York.

References

Moran, S. and McCloy, D., editors (2019). *PHOIBLE 2.0*. Max Planck Institute for the Science of Human History, Jena.

Moroz, G. (2017). *lingtypology: easy mapping for Linguistic Typology*.

Moroz, G. (2019). Phonology of mehweb. In Daniel, M., Dobrushina, N., and Ganenkov, D., editors, *The Mehweb language: Essays on phonology, morphology and syntax*, pages 17–39. Language Science Press.

Nikolaev, D., Nikulin, A., and Kukhto, A., editors (2015). *The database of Eurasian phonological inventories*.

Smeets, R., editor (1994). *The indigenous languages of the Caucasus*, volume 4. Caravan Books.

Xie, Y., Allaire, J., and Grolemund, G. (2018). *R Markdown: The Definitive Guide*. Chapman and Hall/CRC, Boca Raton, Florida. ISBN 9781138359338.