## Comparing phonological systems and syllable structure of Botlikh and Zilo Andi: a data-driven analysis

#### George Moroz

Linguistic Convergence Laboratory, NRU HSE, Moscow, Russia

24 October 2019

"Caucasian Languages: Typology and Diachrony" in honor of Mikhail Egorovich Alekseev, Institute of Linguistics RAS, Moscow



Presentation is available here: https://tinyurl.com/y5xoks3l

Traditional analysis Data-driven analysis



	<b>Traditional analysis</b>	Data-driven analysis
1.	Done by trained linguist	Evaluated by trained linguist
	Could be done from scratch	A previous description needed
2.		(or at least prior expectations)
3.	Doesn't care about amount of data	Care more about amount of data
4.	Less reproducible	More reproducible
5.	Could not be automated	Could be automated



Traditional analysis		Data-driven analysis	
1.	Done by trained linguist	Evaluated by trained linguist	
2.	Could be done from scratch	A previous description needed (or at least prior expectations)	
3.	Doesn't care about amount of data	Care more about amount of data	
4.	Less reproducible	More reproducible	
5.	Could not be automated	Could be automated	

Data-driven approach to phonological description and syllable structure analysis:

- was proposed in [Moroz 2018]
- was applied to syllable structure in [Moroz 2019] to Adyghe data
- was applied to syllable structure in [Romanova 2019] to Russian and Macedonian data



Traditional analysis		Data-driven analysis	
1.	Done by trained linguist	Evaluated by trained linguist	
2.	Could be done from scratch	A previous description needed	
		(or at least prior expectations)	
3.	Doesn't care about amount of data	Care more about amount of data	
4.	Less reproducible	More reproducible	
5.	Could not be automated	Could be automated	

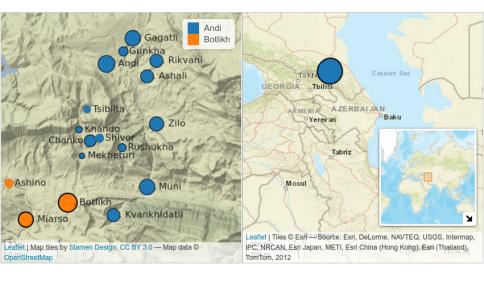
Data-driven approach to phonological description and syllable structure analysis:

- was proposed in [Moroz 2018]
- was applied to syllable structure in [Moroz 2019] to Adyghe data
- was applied to syllable structure in [Romanova 2019] to Russian and Macedonian data

I will present an application of this method to Botlikh and Zilo Andi data



#### Andi and Botlikh villages, created with lingtypology package [Moroz 2017]





- Botlikh < Andic group < EC</li>
- Unwritten (can be written with extended Cyrillic script for Avar)
- ~5,000–8,000 speakers
- Mostly spoken in 3 villages in northwestern Daghestan (Russian Federation): Botlikh, Miarso, Ashino, (Ankho); minor dialectal differences
- One full reference grammar in Georgian [Gudava 1962]
- Two dictionaries:
  [Saidova and Abusov 2012],
  [Alekseev and Azaev 2019]

- Andi < Andic group < EC
- Unwritten (can be written with extended Cyrillic script for Avar)
- ~16,500 speakers
- About 14 villages; There are two main dialect groups: Lower Andi (Muni, Kvankhidatli) and Upper Andi (the rest)
- Several reference grammars
  [Suleymanov 1957] (Rikvani),
  [Salimov 1968] (Gagatli),
  [Tsertsvadze 1965] (Andi)
- No dictionary except [Kibrik and Kodzasov 1988]

#### [Saidova and Abusov 2012]

- Compiled in the 2000s by a native speaker (M. G. Abusov) and an experienced linguist (P. A. Saidova)
- Mostly Botlikh with some notes on Miarso

#### [Saidova and Abusov 2012]

- Compiled in the 2000s by a native speaker (M. G. Abusov) and an experienced linguist (P. A. Saidova)
- Mostly Botlikh with some notes on Miarso

#### [Alekseev and Azaev 2019]

- Compiled in the 1960s / 1970s by a native speaker / philologist (X. G. Azaev) and later (in the 2000s) systematized by an experienced linguist (M. E. Alekseev)
- Subsequently edited by T. A. Maisak and scheduled for posthumous publication later this year
- Botlikh only



#### **Summary:**

- Dictionaries were compiled independently of each other
- with no metadata on the speakers consulted
- data collection was separated with several decades break

**АБАБЕ** бот. пузырник вздутоплотный (кустарник); миарс. г апапу АБАДАДИ нареч. 1) никогда; абадади гънигъви гуч и ден биштиха никогда я к вам не приду 2) ни за что; абадади

игъевчіи гьущуди гьуб ни за что не сделает он этого АБАДАЛ-ИЛЛАГЬИЛ то же, что

АБАЗКНАВ I (-шуб) абазинец АБАЗВНАЙ II (-льильи) абазинка АБАЗЯНАЛ (-v6) абазины; абазиналуб

миціцін абазинский язык АБАЙ (миапс.) см. гъабай АБИ (миарс.) см. гъаби, гъаб \$

АБИР (миапе.) си. гъзбир. АБИРХА (миарс.) см. гъябЅргъя АБТАБУС III (-алын. -абалын)

автобус: абтабусабальным истании автобусная станция: абтабуенчІу гьез нан схать на автобусе; абтабусабальи гьиргьа автобусы прибыли

АБТАМАТ III (-илъи, -абальи) автомат (под автоматического стпезкового абтаматильнку абтаматиди къва"гъа"дай стрелять из

АБТАРµЧКА III (-лъи, -балъи) авторучка; абтаручкади хъварди писать авторучкой; абтаручкахи арен икки платить за авторучку; абтаручкальи **щакън бич**ја в авторучке кончились чернила

АБТЕКА III (-лън, -бальи) аптека; абтекальа дарубальи гьибгьва в аптеку лекарства; абтекальнку дарубальи ибхи купить декарства из

аваданабирлы масата букіа говорили о его жизнерадостности

АВАДАНЛЪИ III (-лън) весёлость, веселье; жизнерадостность; ракІваре аваданлыктылы гьек Іва жизнерадостный человек (буюк имеющий в сердце весетость человек) аваданлын гучІоб гіумру безрадостная жизнь; ракІварчІу ryvlas нежизнерадостный человек; аваданлын

и"тын веселиться, устроить веселье АВАЛ III (-альи, -де) 1) квартал; гьаб ищиб авал ида это наш квартал; ищиб авале машкид игьу в нашем квартале построили мечеть 2) округ; гъаб авале вукінчія гьуштув хьоьов гьеківа в этом округе не было такого хорошего человека

АВАЛ-АХВР III (-альи) начало и конец: авал-ахир гуч1а6 муху работа без начала и конца (о запушенной работе, о бесконечном деле. безретультатной

демпельности) АВАНС III (-альи) аванс: авансай

арси гъибгъва ила получили леньги в

аванс: аванс икки лавать аванс АВАРАГ I (-ильи, -забальи) ред пророк: аварагилын амру веление пророка: аварагильи піпіери бикку хади просить именем пророка; аварагитьалу ради пророка; аварагиди

гьагьамалаб микъили хьвали илти по пути, указанному пророком: АВАРА III (-лън) препятствие: затрулнение: помеха: гьаб мухуй авара

игьу этому делу помещали; гьаб авара игьима мин ара гьингъва? ты пришел, чтобы чинить мне препятствия?

АВАРИ III (-льи) аварийа; авари игьи совершить аварию; аварилъа рукки

абабай моди, сл. обязательно, естественно, конечно: абабай лен гънкто луй санду конечно, я пойду с тобой

абададила никоим образом

абадийа-б вечный

абадул-абад, pn. -ильи уст. вечность

абарона, мн. -е, рп. -лън оборона; абарона бикки оборонять; абароналън оборонительный абди, или абде, пов абду прилечь, лечь ненадолго; полежать; абду букіай прислонить; абду

вигьй прислониться

абйавлени. ли. -йабалън. ин. -лън объявление: газетилън абйавлени объявление в газете: газётильа абйавлёни кки дать объявление в газету: абйавлёни буххай, абйавлёни кіківанди, абйавлени чіабдай повесить объявление

аборт, лил. - абальи, рил. - альи аборт; аборт игьи сделать аборт; Гьай гьешикІва адамиди аборт игъу Эта женшина следала аборт

аваданаб просторный: аваданаб ахи просторный сад: аваданаб раківа шедрая душа

аваландъи 1. простор: 2. шелрость

авал-ахи́р конец света: авал-ахи́р зи́вли панлалън кТоли на том свете встретимся: авал-ахи́р

гучІаб заман вечность

авара, ли. -балън, рп. -лън помеха, преграда

авараг мм. -илън, рм. -де пророк; Муса авараг пророк Муса (Моисей); авараг решинакъа бурдаб микън а) радуга; б) Млечный путь

авари, ми. -йабалън, пи. -лън авария: авари игъй совершить аварию

авиапуч. лог. -е. оп. -илън авиапочта: Ишкур кагъат авиапучили битю Я послал письмо

авлахь, ми. -абальи, ри. -альи бездесое место; бущи авлахьалда ругу скот погнали на пастбище; битІи авлахь равнина



Automatically merge two .doc file into one unified .xls file, ...

аваданабирльи масата букІа говорили о

АВАЛАНЛЪИ III (-лън) весёлость, веселье; жизнерадостность; ракІваре аваданлыцтылын гьек Іва жизнеоздостный человек (буки имеющий в сердце веселость человек) аваланлын гучІоб гІумру безрадостная жизнь: ракІварчІу аваданлъи гучІав нежизнерадостный человек; аваданлый

и"гы веселиться, устроить веселье АВАЛ III (-альи, -де) 1) квартал; гьаб ищиб авал ида это наш квартал; ищиб авале машкид игьу в нашем квартале построили мечеть 2) округ; гьаб авале вукінчія гьуштув хьоьов гьеківа в этом

округе не было такого хорошего человека АВАЛ-АХВР III (-альи) начало и конец; авал-ахир гучІаб муху работа без начала и конца (о запущенной работе, о бесконечном деле безрезультатной деятельности

АВАНС III (-алъи) аванс; авансай арси гъибгъва ила получили леньги в аванс: аванс икки лавать аванс

АВАРАГ I (-илъи, -забальи) рел. пророк; аварагилън амру веление пророка: аварагилъи пПпІери бикку хади просить именем пророка: аварагигъалу ради пророка; аварагиди гьагьамалаб микъиди хьвади идти по пути, указанному пророком;

АВАРА III (-льи) препятствие; затруднение; помеха; гьаб мухуй авара игьу этому делу помешали; гьаб авара игъима мин ара гъингъва? ты пришел, чтобы чинить мне препятствия?

АВАРИ III (-лън) аварийа: авари игьи совершить аварию: аварильа рукки

его жизнерадостности

**АБАБЕ** бот. пузырник взлутоплотный (кустарник): миарс. г апапу АБАЛАЛИ нареч. 1) никогла: абалали гъингъви гучТи лен биштиха никогла я к вам не прилу 2) ни за что: абалали игьевчій гьушуди гьуб ни за что не сделает он этого

АБАДАЛ-ИЛЛАГЬИЛ то же, что АБАЗЯНАВ І (-щуб) абазинец

АБАЗКНАЙ II (-лъилъи) абазинка АБАЗЯНАЛ (-v6) абазины; абазиналуб мин ГиТи абазинский язык АБАЙ (миарс.) см. гьабай

АБИ (миарс.) см. гъаби, гъаб S АБИР (миарс.) см. гъабир АБИРХА (миарс.) см. гьабЅргьа АБТАБУС III (-альи, -абальи) автобус; абтабусабальильи истанци автобусная станция; абтабусичІу гьеа наи

ехать на автобусе; абтабусабальи гьиргьа автобусы прибыли АБТАМАТ III (-илъи, -абальи) автомат (род автоматического стрелкового абтаматильику бахай//

абтаматили къва"гъа"лай стрелять из АБТАРЦЧКА III (-льи, -бальи)

авторучка; абтаручкади хъварди писать авторучкой: абтаручкахи арси икки платить за авторучку; абтаручкальи шакън бичја в авторучке кончились

АБТЕКА III (-льи, -бальи) аптека: абтекальа дарубальи гьибгьва в аптеку поступили лекарства; абтекальику дарубальи ибхи купить лекарства из

абабай вводи. сл. обязательно, естественно, конечно; абабай ден гънкію дуй санду конечно, я пойлу с тобой

абададила никоим образом

абалийа=б всчный

абадул-абад, pn. -ильи уст. вечность

абарона, мп. -é, pn. -лън оборона; абарона бикки оборонять; абароналън оборонительный

абди, нав абде, асв абду прилечь, лечь ненадолго; полежать; абду букіай прислонить; абду вигьй прислониться абйавлени, ми. - йабалън, рп. -лън объявление; газетилън абйавлени объявление в газете;

газетильа абйавлени кки дать объявление в газету; абйавлени буххай, абйавлени ы к Іваний абйавлени и Іаблай повесить объявление аборт, ми. -абалън, рп. -алън аборт; аборт игън сделать аборт; Гъай гъешик ва адамиди

аборт игьу Эта женшина следала аборт

аваданаб просторный; аваданаб ахи просторный сад; аваданаб раківа щедрая душа

аваданлъи 1. простор; 2. шедрость

авал-ахир конец света; авал-ахир зивди дандальи кІбли на том свете встретимся; авал-ахир гучлаб заман вечность

авара, мн. -балъи, рп. -лъи помеха, преграда

авараг мн. - илън, рп. - де пророк; Муса авараг пророк Муса (Моисей); авараг решинакъа бурдаб микън а) радуга; б) Млечный путь

авари, ми. -йабалъи, рп. -лъи авария: авари игъй совершить аварию

авиапуч, мн. -e, pn. -илъи авиапочта; Ишкур кагъат авиапучиди битю Я послал письмо

авлахъ, мн. -абалъи, рп. -альи безпесое место; буцІи авлахъалда руту скот погнали на пастбище; битІи авлахь равнина

автабнограпи ми. -йабальи, рп. -йальи автобнография; Диб автабнограпийальи бехи



- Automatically merge two .doc file into one unified .xls file, ...
- Manually check for similarities (S. Verhees, C. Naccarato and me)

аваданабирльи масата букІа говорили о

АВАЛАНЛЪИ III (-лън) весёлость, веселье; жизнерадостность; раківаре аваданлыцтылын гьек Іва жизнеоздостный человек (буки имеющий в сердце веселость человек) аваланлын гучІоб гІумру безрадостная жизнь: ракІварчІу аваданлъи гучІав нежизнерадостный человек; аваданлый

и"гы веселиться, устроить веселье АВАЛ III (-альи, -де) 1) квартал; гьаб ищиб авал ида это наш квартал; ищиб авале машкид игьу в нашем квартале построили мечеть 2) округ; гьаб авале вукінчія гьуштув хьоьов гьеківа в этом

округе не было такого хорошего человека АВАЛ-АХВР III (-альи) начало и конец; авал-ахир гучІаб муху работа без начала и конца (о запущенной работе, о бесконечном деле безрезультатной деятельности

АВАНС III (-алъи) аванс; авансай арси гъибгъва ила получили леньги в аванс: аванс икки лавать аванс

АВАРАГ I (-илъи, -забальи) рел. пророк; аварагилън амру веление пророка: аварагилъи пПпІери бикку хади просить именем аварагигъалу ради пророка; аварагиди гьагьамалаб микъиди хьвади идти по пути, указанному пророком;

АВАРА III (-льи) препятствие; затруднение; помеха; гьаб мухуй авара игьу этому делу помешали; гьаб авара игъима мин ара гъингъва? ты пришел, чтобы чинить мне препятствия?

АВАРИ III (-лън) аварийа: авари игьи совершить аварию: аварильа рукки

его жизнерадостности **АБАБЕ** бот. пузырник взлутоплотный

АБАЛАЛИ нареч. 1) никогла: абалали гъингъви гучТи лен биштиха никогла я к вам не прилу 2) ни за что: абалали игьевчій гьушуди гьуб ни за что не сделает он этого АБАДАЛ-ИЛЛАГЬИЛ то же, что

(кустарник): миарс. г апапу

АБАЗЯНАВ І (-щуб) абазинец АБАЗКНАЙ II (-лъилъи) абазинка АБАЗЯНАЛ (-v6) абазины; абазиналуб

мин ГиТи абазинский язык АБАЙ (миарс.) см. гьабай АБИ (миарс.) см. гъаби, гъаб S АБИР (миарс.) см. гъабир

АБИРХА (миарс.) см. гьабЅргьа АБТАБУС III (-альи, -абальи) автобус; абтабусабальильи истанци автобусная станция; абтабусичІу гьеа наи ехать на автобусе; абтабусабальи гьиргьа

автобусы прибыли АБТАМАТ III (-илъи, -абальи) автомат (род автоматического стрелкового абтаматильику бахай// абтаматили къва"гъа"лай стрелять из

АБТАРЦЧКА III (-льи, -бальи) авторучка; абтаручкади хъварди писать авторучкой: абтаручкахи арси икки платить за авторучку; абтаручкальи шакън бичја в авторучке кончились

АБТЕКА III (-льи, -бальи) аптека: абтекальа дарубальи гьибгьва в аптеку поступили лекарства; абтекальику дарубальи ибхи купить лекарства из

абабай вводи. сл. обязательно, естественно, конечно; абабай ден гънкію дуй санду конечно, я пойлу с тобой

абададила никоим образом

абалийа=б всчный

абадул-абад, pn. -ильи уст. вечность

абарона, мп. -é, pn. -лън оборона; абарона бикки оборонять; абароналън оборонительный

абди, нав абде, асв абду прилечь, лечь ненадолго; полежать; абду букіай прислонить; абду вигьй прислониться

абйавлени, ми. - йабалън, рп. -лън объявление; газетилън абйавлени объявление в газете; газетильа абйавлени кки дать объявление в газету; абйавлени буххай, абйавлени ы к Іваний абйавлени и Іаблай повесить объявление

аборт, ми. -абалън, рп. -алън аборт; аборт игън сделать аборт; Гъай гъешик ва адамиди аборт игьу Эта женшина следала аборт

аваданаб просторный; аваданаб ахи просторный сад; аваданаб раківа щедрая душа

аваданлъи 1. простор; 2. шедрость

авал-ахир конец света; авал-ахир зивди дандальи кІбли на том свете встретимся; авал-ахир гучлаб заман вечность

авара, мн. -бальи, рп. -льи помеха, преграда

авараг мн. - илън, рп. - де пророк; Муса авараг пророк Муса (Моисей); авараг решинакъа бурдаб микън а) радуга; б) Млечный путь

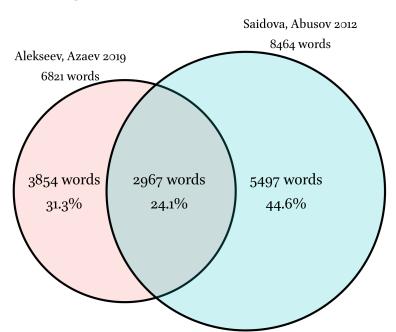
авари, ми. -йабалъи, рп. -лъи авария: авари игъй совершить аварию

авиапуч, мн. -е, pn. -ильи авиапочта; Ишкур кагьат авиапучиди битю Я послал письмо

авлахъ, мн. -абалъи, рп. -альи безпесое место; буцІи авлахъалда руту скот погнали на пастбище; битІи авлахь равнина

автабнограпи ми. -йабальи, рп. -йальи автобнография; Диб автабнограпийальи бехи





- If we remove a stress sign then there is 2492 same words and 398 different words (14%)
- If we don't remove a stress sign then there is 2024 same words and 866 different words (30%)
- ⇒ There are 16% of words have different stress pattern?



- If we remove a stress sign then there is 2492 same words and 398 different words (14%)
- If we don't remove a stress sign then there is 2024 same words and 866 different words (30%)
- ⇒ There are 16% of words have different stress pattern? Including cases when stress is present in one dictionary and absent in the other.



- If we remove a stress sign then there is 2492 same words and 398 different words (14%)
- If we don't remove a stress sign then there is 2024 same words and 866 different words (30%)
- ⇒ There are 16% of words have different stress pattern? Including cases when stress is present in one dictionary and absent in the other.
  - What cause the difference between dictionaries?
    - Stress pattern differences in 317 words (about 11%)
    - Multiple cases where there is a small difference that could be explained either as a typo or in terms phonological variation (čuhí 'to run' (A&A) vs. čũhí (S&A), kusu 'cherry plum' (A&A) vs. kus:u (S&A))
    - Multiple cases where Russian borrowings were adopted differently (pojiz 'train' (A&A) vs. poezd (S&A), biton 'milk can' (A&A) vs. bitun (S&A))



#### About 25 cases:

[Alekseev and Azaev 2019]	[Saidova and Abusov 2012]	
ãha <b>j</b> r	ãhar	'message'
beza <b>j</b> r	bezir	'roasting'
mik'ku <b>j</b> r	mik':ur	'swallowing'
reqxu <b>j</b> r	reqx <sup>w</sup> ir	'fight'
reſku <b>j</b> r	refkur	'overnight stay'
rik <sup>w</sup> a <b>j</b> r	rik <sup>w</sup> ar	'lighting'
χwardar	χwardir	'digging'
mi?ar	mi?ar	'nose'
	•••	

#### About 6 cases

Galai

juug	3110
inuSalla	'everywhere'
Silla	'reason'
	inuSalla

'cilt'

Gallai

·· ···

#### Conclusions:

Data-driven approach

#### Discussion:

- Lemmata are not the same as wordforms, so gathered model should be checked with the wordform material
- Lemmata could contain some affix that will shift all frequencies (e. g. INF or PL) for some types of phonological units
- It is nice to compare obtained models with the models built on corpora data

# Send me a letter! agricolamzgmail.com

Presentation is available here: tinyurl.com/y3wtkcbq



#### References

- Alekseev, M. and X. Azaev (2019). *Botlixsko-russkij slovar'* [*Botlikh-Russian dictionary*]. Moscow: Academia.
- Gudava, Togo, E. (1962). Botlixuri ena [The Botlikh language]. Tbilisi: Mecniereba.
- Kibrik, A. E. and S. V. Kodzasov (1988). Sopostavitelnoye izucheniye dagestanskikh yazykov [Comparative study of Daghestanian languages]. Moscow: Moscow State University.
- Moroz, G., A. (2018). lingphonology: automatic phonological description. R package draft.
- Moroz, G., A. (2019). Slogovaya struktura adygeyskogo yazika: ot dannyx k obosheniyam [Adyghe syllable structure: From empirical data to generalizations]. *Voprosy Jazykoznanija* 2, 82–95.
- $Moroz, G.\ (2017).\ lingtypology: easy mapping for Linguistic \ Typology.$
- Romanova, K., I. (2019). Automatic Syllable Structure Extracting From Dictionaries: Slavic Data. Term paper.
- Saidova, P. A. and M. G. Abusov (2012). Botlixsko-russkij slovar' [Botlikh-Russian dictionary]. Makhachkala: IJaLI.
- Salimov, X. S. (2010 (1968)). *Gagatlinskij govor andijskogo jazyka* [The Gagatli dialect of the Andi language]. Makhachkala.
- Suleymanov, J. G. (1957). Grammatičeskij očerk andijskogo jazyka (po dannim govora s. Rikvani) [Grammar sketch of the Andi language (based on material from the dialect of the village Rikvani)]. Ph. D. thesis, Institut Jazykoznania AN SSSR.
- Tsertsvadze, I. I. (1965). Andiuri Ena. Tbilisi: Metsniereba.