

# 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 M. E. Alekseev,  
Institute of Linguistics RAS, Moscow

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



# Phonological description: data-driven analysis

Traditional analysis	Data-driven analysis
----------------------	----------------------

## Phonological description: data-driven analysis

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

## Phonological description: data-driven analysis

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

## Phonological description: data-driven analysis

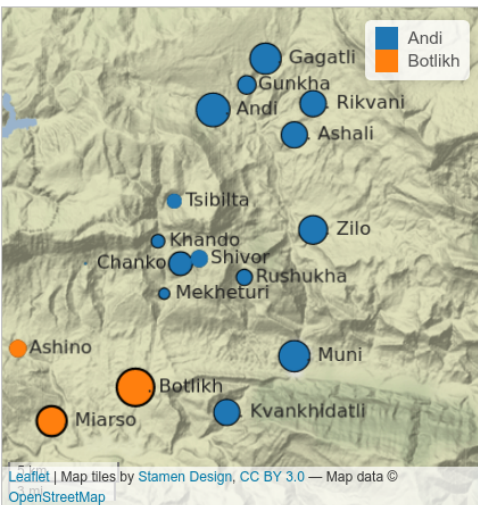
Traditional analysis	Data-driven analysis
1. Done by trained linguist	Evaluated by trained linguist
2. Could be done from scratch	A previous description needed (or at least prior expectations)
3. Doesn't care about amount of data	Care more about amount of data
4. Less reproducible	More reproducible
5.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

# Comparing two Botlikh dictionaries

[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



# Comparing two Botlikh dictionaries

## [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

# Comparing two Botlikh dictionaries

## Summary:

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

А  
АБАБЕ *бот.* пузырник вздутопный (кустарник); *миар.* Гегитур  
АБАДАДИ *миар.* 1) некогда; абалади  
гынган гучи ден биштха некогда в к  
вам не приду 2) ни за что; абалади  
нгьевчи гущуди габ ни за что не  
сделает он этого  
АБАДАЛ-ЦЛАГЫЛ то же, что  
абалади  
АБАЗЕНАВ I (-чуб) абазинец  
АБАЗЕНАЙ II (-лыгъа) абазинка  
АБАЗЕНАЛ (-уб) абазин; абазиналуб  
мичил абазинский язык  
АБАЙ (*миар.*) см. габай  
АБИ (*миар.*) см. габин, габ5  
АБИР (*миар.*) см. габир  
АБИРХА (*миар.*) см. габ5грга  
АБТАБУС III (-алы, -балы)  
автобус; абтабусабалыч истанци  
автобусная станция; абтабусачу гыа ни  
схат на автобус; абтабусабалы гырга  
автобус прибыли  
АБТАМАТ III (-алы, -балы)  
автомат (*род.* автоматического стрелкового  
оружия); абтаматыну бахай/  
абтаматын кыа/гыдай стрелять из  
автомата  
АБТАРЧКА III (-алы, -балы)  
авторучка; абтарчка хварди писать  
авторучкой; абтаручка арси икхи  
платить за авторучку; абтаручкалы  
ишкы бичи в авторучке кончился  
чернила  
АБТЕКА III (-алы, -балы) аптека;  
абтекала дарубалы гыбгъа в аптеку  
поступили лекарства; абтекалыну  
дарубалы ибхи купить лекарства из

А  
аваданабирлы масита буква говорили о  
его жизненности  
АВАДАНЫ III (-алы) веселье,  
веселье; аваданасты; рактваре  
аваданымылы гыбгъа жизнерадостный  
человек (бук. изменяющий в сердце веселость  
человек) аваданы гучиб гыму  
бесрадостная жизнь; рактварчу  
аваданы гучла атам  
нежизнерадостный человек; аваданы  
иты веселиться, устроить веселье  
АВАЛ III (-алы, -де) I) квартал; гыб  
нииб авал ила это наш квартал; нииб  
авале манкка нгу в нашем квартале  
построили мечеть 2) округ; гыб авале  
аудича гущуду зылов гыбгъа в этом  
округе не было такого хорошего человека  
АВАЛ-АХИР III (-алы) начало и  
конец; авал-ахир гучлаб муху работа без  
начала и конца (о запущенной работе, о  
бесконечном деле, безрезультатной  
деятельности)  
АВАНС III (-алы) аванс; авансай  
арси гыбгъа ила получить деньги в  
аванс; аванс икхи давать аванс  
АВАРАГ I (-алы, -балы) *ред.*  
пророк; аварагилы аму вельние  
пророка; аварагилы пилери биксу  
хачи просить именем пророка;  
аварагилы ради пророка; аварагилы  
гыгачалаб мыскы хыди иди по  
пути, указанному пророком;  
АВАРА III (-алы) препятствие;  
затруднение; помеха; гыб муху авара  
нгу этому делу помешали; гыб авара  
ныма мин ара гынгъа? ты пришел,  
чтобы чинить мне препятствия?  
АВАРИ III (-алы) авария; авари  
нгы совершить аварию; аварилы рухки

А  
абай *аводи.* сл. обязательно, естественно, конечно; абай ден гыкто дуй санду конечно, я  
пойду с тобой  
абаладила ником образом  
абалайа-б вечный  
абалау-абал, *ред.* -алы *уст.* вечность  
абарона, *мин.* -ё, *ред.* -алы оборона; абарона бикси оборонять; абароналы оборонительный  
абай, *мин.* абас, *ис.* абду прилечь, лечь ненадолго; полежать; абду буклай прислонить; абду  
внгы прислониться  
абавлени, *мин.* -йабалы, *ред.* -алы объявление; гаситы абавлени объявление в газете;  
гаситы абавлени кхи дать объявление в газету; абавлени бухай, абавлени  
кхиванди, абавлени чабай повесить объявление  
аборт, *мин.* -йабалы, *ред.* -алы аборт; аборт нгы сделать аборт; Гыб гешикла азамиди  
аборт нгу Эта женщина сделала аборт  
аваданаб просторный; аваданаб ахи просторный сад; аваданаб ракла шедра душа  
аваданы I. простор; 2. шедрость  
авал-ахир конец света; авал-ахир зыди данцалы кбли на том свете встретимся; авал-ахир  
гучлаб заман вечность  
авара, *мин.* -балы, *ред.* -алы помеха, препятствие  
авараг *мин.* -алы, *ред.* -де пророк; Муса авараг пророк Муса (Моисей); авараг решинакка  
бурдаб мыкы а) радуга; б) Млечный путь  
авари, *мин.* -йабалы, *ред.* -алы авария; авари нгы совершить аварию  
аванач, *мин.* -ё, *ред.* -алы аванпост; Ишкур кыгат аванпачиди битю Я послал письмо  
аванпостой  
авлах, *мин.* -йабалы, *ред.* -алы безлесое место; бунги авлахалла руту скот погнали на  
пастбище; битги авлах равнина  
автобиограи *мин.* -йабалы, *ред.* -алы автобиография; Диб автобиограинилы беси

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

А

48

аваданабирлы масата букйа говорили о его жизнерадостности

**АВАДАНТЫЙ** III (-лыи) веселость, веселье; жизнерадостность; **ракваре** авадантагытлы гьекйа жизнерадостный человек (букв. имеющий в сердце веселость человек) **аваданын гучюб** гүмүрү безрадостная жизнь; **ракварчылу** аваданын гучйавам адам нежизнерадостный человек; **аваданын игы**и веселиться, устроить веселье

**АВАЛ** III (-алыи, -лө) 1) квартал; **гыаб ичиб** авал ида это наш квартал; **ичиб** авале машинд игыу в нашем квартале построили мечеть 2) округ; **гыаб авале** **вукичйа** гүштүгү **хыюво** гьекйа в этом округе не было такого хорошего человека

**АВАЛ-АХРР** III (-алыи) начало и конец; **авал-ахир гучйаб** муху работа без начала и конца (о запущенной работе, о бесконечном деле, безрезультатной деятельности)

**АВАНС** III (-алыи) аванс; **авансай** арси гыиыбгьа ида получили деньги в аванс; **аванс** икки давать аванс

**АВАРАГ** I (-илыи, -зйабалыи) *реал* пророк; **аварагилыи** амуру веление пророка; **аварагилыи шйпер**и **биксу** хади просить именем пророка; **аварагигьалу** ради пророка; **аварагиди** гыагьамалаб микьиди хьаиди иди по пути, указанному пророком;

**АВАРА** III (-лыи) препятствие; затруднение; помеха; **гыаб мухуй** **авара игы**у этому делу помешали; **гыаб авара игыи**а **мин ара гыиыгьа**т вы пришел, чтобы чинить мне препятствия?

**АВАРИ** III (-лыи) авария; **авари игы**и совершить аварию; **авариыла** **рукки**

А

**абабай** *вводи*, *сл.* обязательно, естественно, конечно; **абабай** **ден гыиык**о дуй **санду** конечно, я пойду с тобой

**абадйа**дйа **инком**и образом

**абадйа**а=б вечный

**абадул-абад**, *рп.* -илыи *уст.* вечность

**абарона**, *мн.* -ё, *рп.* -лыи оборона; **абарона** **бикки** оборонять; **абароналыи** оборонительный

**абдй**, *мн* **абдё**, *тсв* **абду** прилечь, лечь ненадолго; **по**сезать; **абду** **букйа**и прислонить; **абду** **вигй**и прислониться

**абйавлэни**, *мн.* -йабалыи, *рп.* -лыи объявление; **газэтилыи** **абйавлэни** объявление в газете; **газэтилыи** **абйавлэни** **кки** дать объявление в газету; **абйавлэни** **буххй**, **абйавлэни** **кйкйивандй**, **абйавлэни** **чйабдй** повестить объявление

**аборт**, *мн.* -абалыи, *рп.* -алыи аборт; **аборт игы**и сделать аборт; **Гьай** **гыеникйа** **адамиди** **аборт игы**у Эта женщина сделала аборт

**аваданаб** просторный; **аваданаб** **ахи** просторный сад; **аваданаб** **ракьа**ва щедрая душа

**аваданылыи** 1. простор; 2. щедрость

**авал-ахир** конец света; **авал-ахир** **зйивди** **даидалыи** **кйбди** на том свете встретимся; **авал-ахир** **гучйаб** **заман** вечность

**авара**, *мн.* -балыи, *рп.* -лыи помеха, преграда

**авараг** *мн.* -илыи, *рп.* -де пророк; **Муса** **авараг** пророк Муса (Моисей); **авараг** **решиинакьа** **бурдаб** **мькьи** а) радуга; б) Млечный путь

**авари**, *мн.* -йабалыи, *рп.* -лыи авария; **авари игы**и совершить аварию

**авиануч**, *мн.* -ё, *рп.* -илыи авианочта; **Ишкур** **кагыат** **авианучиди** **битё** Я послал письмо авианочтой

**авлэхх**, *мн.* -абалыи, *рп.* -алыи безлесое место; **бунйи** **авлэххадда** **рүту** скот погнали на пастбище; **битйи** **авлэхх** равнина

**автабиограпи** *мн.* -йабалыи, *рп.* -йалыи автобиография; **Диб** **автабиограпи**иалыи **бехи**

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

А

АБАБЕ *бот.* пузырник вздутоплотный (*кустарник*); *миарс.* Папану

АБАДАДИ *нареч.* 1) никогда; абадали гынгыгын гучи ден биштика никогда я к вам не приду 2) ни за что; абадали итывечи гышуди губ ни за что не делает он этого

АБАДАЛ-ИЛЛАГЫЛ то же, что абадали

АБАЗНАВИ (-шуб) абазинец

АБАЗНАЙ II (-лыты) абазинка

АБАЗНАЛ (-уб) абазин; абазиналуб минци абазинский язык

АБАЙ (*миарс.*) см. гыбай

АБИ (*миарс.*) см. гыби, гыбS

АБИР (*миарс.*) см. гыбир

АБИРХА (*миарс.*) см. гыбSргъа

АБТАБУС III (-алы, -абалы) автобус; абтабусабылты истанци автобусная станция; абтабучу гыа нан ехать на автобус; абтабусабыл гыргъа автобусы прибыли

АБТАМАТ III (-илы, -абалы) автомат (*род автоматического стрелкового оружия*); абтаматырку бахай/абтаматын кыаа гыадай стрелять из автомата

АБТАРЧКА III (-лы, -балы) авторучка; абтарчкани хыарди писать авторучкой; абтарчкани ари кски платит за авторучку; абтарчканы шакы бичи в авторучке кончились чернила

АБТЕКА III (-лы, -балы) аптека; абтекалы дарубалы гынгыгва в аптеку поступили лекарства; абтекалыку дарубалы нхи купить лекарства из

аваданабырлы масата букла говорили о его жизнерадостности

АВАДАНТЫ III (-лы) веселость, веселье; жизнерадостность; ракваре авадаптылы гыкwa жизнерадостный человек (*бук.* имеющий в сердце веселость человек) авадапты гучиоб гымру безрадостная жизнь; ракварчу авадапты гучиав адам нежизнерадостный человек; авадапты итгы веселиться, устроить веселье

АВАЛ III (-алы, -дэ) 1) квартал; гыаб ишиб авал ида это наш квартал; ишиб авале машид игы в нашем квартале построили мечеть 2) округ; гыаб авале пукичи гыштыв хьотон гыкwa в этом округе не было такого хорошего человека

АВАЛ-АХРР III (-алы) начало и конец; авал-ахир гучиав муху работа без начала и конца (*о записанной работе, о бесконечном деле, безрезультатной деятельности*)

АВАНС III (-алы) аванс; авансай арен гынгыгва ида получили деньги в аванс; авански давать аванс

АВАРАГ I (-лы, -абалы) *рел.* пророк; аварагилы амуру веселение пророка; аварагилы шпери биску хади просить именем пророка; аварагигылу ради пророка; аварагиди гыгымалаб миксиди хыади иди по пути, указанному пророком;

АВАРА III (-лы) препятствие; затруднение; помеха; гыаб муху авара игы этому делу помешали; гыаб авара игыма мин ара гынгыгва? ты пришел, чтобы чинить мне препятствия?

АВАРИ III (-лы) авария; авари игы совершить аварию; аварилы рукки

А

абабай *вводи. сл.* обязательно, естественно, конечно; абабай ден гыкlo дуй санду конечно, я пойду с тобой

абададила никном образом

абадина=б вечный

абадул-абад, *рп.* -илы *уст.* вечность

абарона, *ми.* -э, *рп.* -лы оборона; абарона бикки оборонять; абароналы оборонительный

абди, *ми.* абдэ, *тсв.* абду прилечь, лечь ненадолго; полезать; абду буклай прислонить; абду вигий прислониться

абнавлэни, *ми.* -йабалы, *рп.* -лы объявление; газэтилы абнавлэни объявление в газете; газэтилы абнавлэни кки дать объявление в газету; абнавлэни буххай, абнавлэни кикwанди, абнавлэни члабдай повестить объявление

аборт, *ми.* -абалы, *рп.* -алы аборт; аборт игы сделать аборт; Гыай гыеникwa адаминди аборт игы Эта женщина сделала аборт

аваданаб просторный; авадаанаб ахи просторный сад; аваданаб ракwa щедрая душа

аваданылы 1. простор; 2. шедкость

авал-ахир конец света; авал-ахир зивди дадалы кйбли на том свете встретимся; авал-ахир гучиав заман вечность

авара, *ми.* -балы, *рп.* -лы помеха, преграда

авараг *ми.* -илы, *рп.* -де пророк; Муса авараг пророк Муса (Моисей); авараг решинакы бурдаб миксы а) радуга; б) Млечный путь

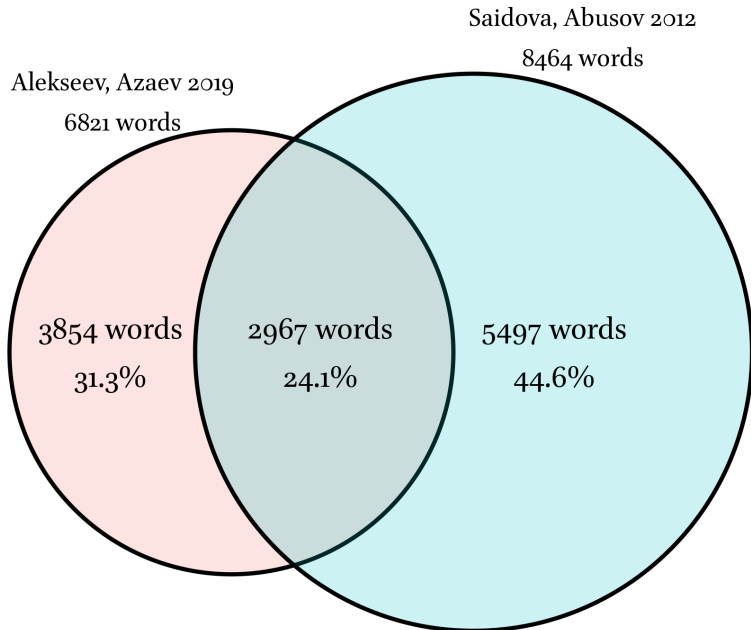
авари, *ми.* -йабалы, *рп.* -лы авария; авари игы совершить аварию

авиначу, *ми.* -е, *рп.* -илы авиачота; Ишкур кагыат авиачуиди битю Я послал письмо авиачотой

авлахъ, *ми.* -абалы, *рп.* -алы безлесое место; буңи авлахъада руту скот погнали на пастбище; бити авлахъ равнина

автабиограпи *ми.* -йабалы, *рп.* -йалы автобиография; Диб автабиограпинылы бехи

# Comparing two Botlikh dictionaries



## Comparing two Botlikh dictionaries

- If we remove a stress sign then there is 2495 same words and 395 different words (14%)
  - If we don't remove a stress sign then there is 2027 same words and 863 different words (30%)
- ⇒ There are 16% of words have different stress pattern?

## Comparing two Botlikh dictionaries

- If we remove a stress sign then there is 2495 same words and 395 different words (14%)
  - If we don't remove a stress sign then there is 2027 same words and 863 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.

# Comparing two Botlikh dictionaries

- If we remove a stress sign then there is 2495 same words and 395 different words (14%)
  - If we don't remove a stress sign then there is 2027 same words and 863 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' [Alekseev and Azaev 2019] vs. *čūhí* [Saidova and Abusov 2012],  
*kusu* 'cherry plum' [Alekseev and Azaev 2019] vs. *kus:u* [Saidova and Abusov 2012]

- Multiple cases where Russian borrowings were adopted differently  
*awtobus* 'bus' [Alekseev and Azaev 2019] vs. *abtabus* [Saidova and Abusov 2012],  
*biton* 'milk can' [Alekseev and Azaev 2019] vs. *bitun* [Saidova and Abusov 2012],  
*apteka* 'pharmacy' [Alekseev and Azaev 2019] vs. *abteka* [Saidova and Abusov 2012]

- Morphological preferences

*dinija=w* 'pious' [Alekseev and Azaev 2019] vs. *dinija=b* [Saidova and Abusov 2012]



# Comparing two Botlikh dictionaries

About 25 cases:

[[Alekseev and Azaev 2019](#)]

*āhajr*

*bezajr*

*mik'kujr*

*reqχujr*

*reʃkujr*

*rik<sup>w</sup>ajr*

---

*χwardar*

*miʔar*

...

[[Saidova and Abusov 2012](#)]

*āhar*

*bezir*

*mik'ur*

*reqχ<sup>w</sup>ir*

*reʃkur*

*rik<sup>w</sup>ar*

---

*χwardir*

*miʔar*

...

'message'

'roasting'

'swallowing'

'fight'

'overnight stay'

'lighting'

---

'digging'

'nose'

...

About 6 cases:

*ʃ:alaj*

*inuʃala*

*ʃila*

...

*ʃ:allaj*

*inuʃalla*

*ʃilla*

...

'silt'

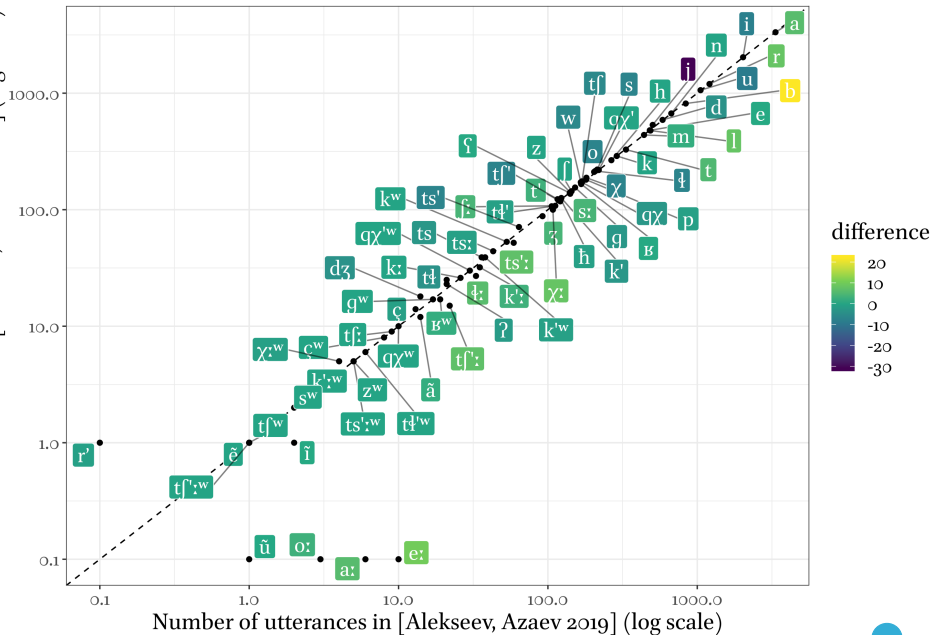
'everywhere'

'reason'

...

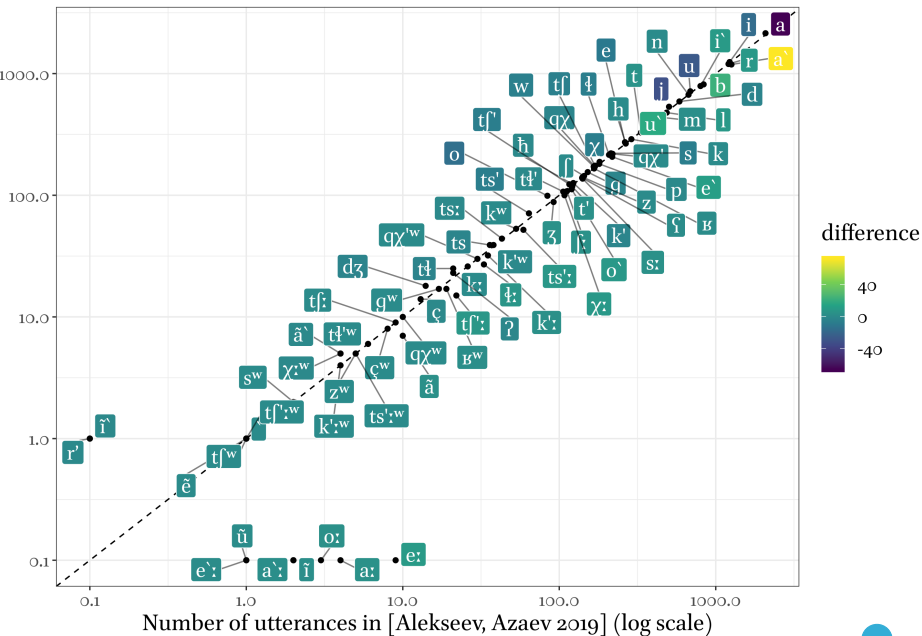
# Comparing two Botlikh dictionaries (without stress patterns)

Number of utterances in [Saidova, Abusov 2012] (log scale)



# Comparing two Botlikh dictionaries (with stress patterns)

Number of utterances in [Saidova, Abusov 2012] (log scale)



## Conclusions:

- Data-driven approach

## Discussion:

- Botlikh dictionaries were specially selected to share the meaning, the same procedure for Andic were not done
- Botlikh dictionaries contain a lot of borrowings, it is not true for the Andic dictionary
- 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](mailto:agricolamzgmail.com)

Presentation is available here:  
[tinyurl.com/y3wtkcbq](https://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 Jazykoznanija AN SSSR.
- Tsertsvadze, I. I. (1965). *Andiuri Ena*. Tbilisi: Metsniereba.