

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

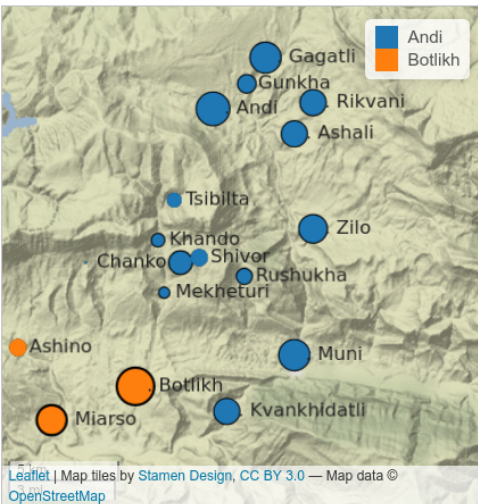
Traditional analysis	Data-driven analysis
1. Done by trained linguist	Evaluated by trained linguist
2. Could be done from scratch	A previous description needed (or at least prior expectations)
3. Doesn't care about amount of data	Care more about amount of data
4. Less reproducible	More reproducible
5. 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*); *мшар.* Гегинг
АБАДАДИ *мшар.* 1) некогда; абадани
гынган гучи ден биштха некогда в к
вам не приду 2) ни за что; абадани
нгьевчи гущуди габ ни за что не
сделает он этого
АБАДАЛ-ЦЛАГЫЛ то же, что
абадани
АБАЗЕНАВ I (-чуб) абазинец
АБАЗЕНАЙ II (-лыгъа) абазинка
АБАЗЕНАЛ (-уб) абазин; абазинадуб
мшарин абазинский язык
АБАЙ (*мшар.*) см. габай
АБИ (*мшар.*) см. габин, габ5
АБИР (*мшар.*) см. габир
АБИРХА (*мшар.*) см. габ5срга
АБАБУС III (-алы, -балы)
автобус; абтабусабылчы истанци
автобусная станция; абтабучулы гыа ни
схат на автобус; абтабусабыл гыргыа
автобус прибыли
АБАМАТ III (-алы, -балы)
автомат (*род.* автоматического стрелкового
оружия); абтаматыну бахай/
абтаматици кыа/гыдай стрелять из
автомата
АБТАРЧКА III (-алы, -балы)
авторучка; абтарчка хварди писать
авторучкой; абтаручкани арси икхи
платить за авторучку; абтаручканы
шакы бичи в авторучке кончился
чернила
АБТЕКА III (-алы, -балы) аптека;
абтекалы дарубалы гыбгыа в аптеку
пошлины лекарства; абтекалыну
дарубалы ибхи купить лекарства из

48
аваданабирлы масита буква говорили о
его жизнелюбности
АВАДАНЫ III (-алы) веселость,
веселье; аваданарасты; рактваре
аваданымылы гыбгыа жизнелюбный
человек (*бук.* изменений в сердце веселость
человек) аваданы гучиб гыму
бесрадостная жизнь; рактварчы
аваданы гучиав атам
нежизнелюбный человек; аваданы
иты веселиться, устроить веселье
АВАЛ III (-алы, -де) 1) квартал; гыб
нииб авал ила это наш квартал; нииб
авале манпка нгу в нашем квартале
построили мечеть 2) округ; гыб авале
аудачи гущуды зылов гыбгыа в этом
округе не было такого хорошего человека
АВАЛ-АХИР III (-алы) начало и
конец; авал-ахир гучиаб муху работа без
начала и конца (*о записной работе, о
бесконечном деле, безрезультатной
деятельности*)
АВАНС III (-алы) аванс; авансай
арси гыбгыа ила получить деньги в
аванс; аванс икхи давать аванс
АВАРАГ I (-алы, -балы) *ред.*
пророк; аварагиты аму вельние
пророка; аварагиты пйери биксу
хачи просить именем пророка;
аварагитыду ради пророка; аварагиты
гыгачалаб мыкхы хыди иди по
пути, указанному пророком;
АВАРА III (-алы) препятствие;
затруднение; помеха; гыб муху авара
нгы этому делу помешали; гыб авара
мычма мни ара гынгыа? ты пришел,
чтобы чинить мне препятствия?
АВАРИ III (-алы) авария; авари
нгы совершить аварию; авариты рухки

А
абабай *аводи.* сл. обязательно, естественно, конечно; абабай ден гыкто дуй санду конечно, я
пойду с тобой
абаданила ником образом
абадийа-б вечный
абадул-абад, *рт.* -алы *уст.* вечность
абарона, *мш.* -ё, *рт.* -алы оборона; абарона бикси оборонять; абаронаты оборонительный
абай, *мш.* абас, *ис.* абду прилечь, лечь ненадолго; полежать; абду буклай прислонить; абду
внгы прислониться
абавлени, *мш.* -йабалы, *рт.* -алы объявление; гаситы абавлени объявление в газете;
гаситы абавлени кхи дать объявление в газету; абавлени бухай, абавлени
кйиванди, абавлени чйабай повесить объявление
аборт, *мш.* -абалы, *рт.* -алы аборт; аборт нгы сделать аборт; Гыб гешиква адамци
аборт нгу Эта женщина сделала аборт
аваданаб просторный; аваданаб ахи просторный сад; аваданаб ракта шедра душа
аваданы 1. простор; 2. шедрость
авал-ахир конец света; авал-ахир зыди данцаты кйли на том свете встретимся; авал-ахир
гучиаб заман вечность
авара, *мш.* -балы, *рт.* -алы помеха, препятствие
авараг *мш.* -алы, *рт.* -де пророк; Муса авараг пророк Муса (Моисей); авараг решинакка
бурдаб мыкы а) радуга; б) Млечный путь
авари, *мш.* -йабалы, *рт.* -алы авария; авари нгы совершить аварию
аванач, *мш.* -ё, *рт.* -алы аванпост; Ишкур кыгат аванпачи битю Я послал письмо
аванпостой
авлах, *мш.* -абалы, *рт.* -алы безлесое место; бунги авлахалла руту скот погнани на
пастбище; битги авлах равнина
автобиограи *мш.* -йабалы, *рт.* -алы автобиография; Диб автобиограиниылы 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 (букя, имеющий в сердце веселость человек) **аваданын гучюб** гьумру безрадостная жизнь; **ракварчу** аваданын гучйавам адам нежизнерадостный человек; **аваданын игы**и веселиться, устроить веселье

АВАЛ III (-алыи, -лэ) 1) квартал; **гыб ишиб** авал ида это наш квартал; **ишиб** авале машинд игыу в нашем квартале построили мечеть 2) округ; **гыб авале** **вукичйа** гьунтув **хьотов** 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 гыбгыва ида получили деньги в аванс; **аванс** икши давать аванс

АВАРАГ I (-илыи, -зйабалыи) *реж.* пророк; **аварагилы** амру веселение пророка; **аварагилы шйпер**и **биску** хади просить именем пророка; **аварагигьалу** ради пророка; **аварагиди** гьагьамалаб **мисхиди** хьавади иди по пути, указанному пророком;

АВАРА III (-лыи) препятствие; затруднение; помеха; **гыб мухуй** **авара игы**у этому делу помешали; **гыб авара игы**и **мина** **ара** гьингыва? ты пришел, чтобы чинить мне препятствия?

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

А

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

абадйадла **инком**и образом

абадйа=б вечный

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

абарона, *мн.* -э, *рп.* -лыи оборона; **абарона** **бикки** оборонять; **абароналы** оборонительный

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

абйавлэни, *мн.* -йабалыи, *рп.* -лыи объявление; **газэтилы** **абйавлэни** объявление в газете; **газэтилы** **абйавлэни** **кки** дать объявление в газету; **абйавлэни** **буххйа**, **абйавлэни** **кйкйавидй**, **абйавлэни** **чйабдай** 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** **ракйа**ва шедрая душа

аваданылы 1. простор; 2. шедрость

авал-ахир конец света; **авал-ахир** **зйивди** **дацдалы** **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** **реши**и **накыа** **бурдаб** **мйкыи** а) радуга; б) Млечный путь

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

авйапуч, *мн.* -е, *рп.* -илыи авиационца; **Ишкур** **кагыат** **авйапучиди** **битйо** Я послал письмо авиациончой

авлэхх, *мн.* -абалыи, *рп.* -алыи безлесое место; **буни**и **авлэххадла** **рйту** скот погнали на пастбище; **битйи** **авлэхх** равнина

автабиограпи *мн.* -йабалыи, *рп.* -йалыи автобиография; **Диб** **автабиограпи**и **налыи** **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 итьевчи 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; абтабусалыты истанци автобусная станция; абтабусичу гыа нан ехать на автобус; абтабусабалы гыргъа автобусы прибыли

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

АБТАРЧКА III (-лы, -балы) авторучка; абтарчкати хъварди писать авторучкой; абтарчкахи аргси иски платит за авторучку; абтарчкаты шыкы бичи в авторучке кончились чернила

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

аваданабирлы масата букла говорили о его жизнерадостности

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

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

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

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

АВАРАГ I (-лы, -абалы) *рел.* пророк; аварагилы амуру веление пророка; аварагилы шлперу биску хади просить именем пророка; аварагигылу ради пророка; аварагиди гыгымалаб миксиди хъвади 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 гыешикwa адаминди аборт игы Эта женщина сделала аборт

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

аваданлы 1. простор; 2. шедкость

авал-ахир конец света; авал-ахир зивди дадалы кйбли на том свете встретимся; авал-ахир гучиаб 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

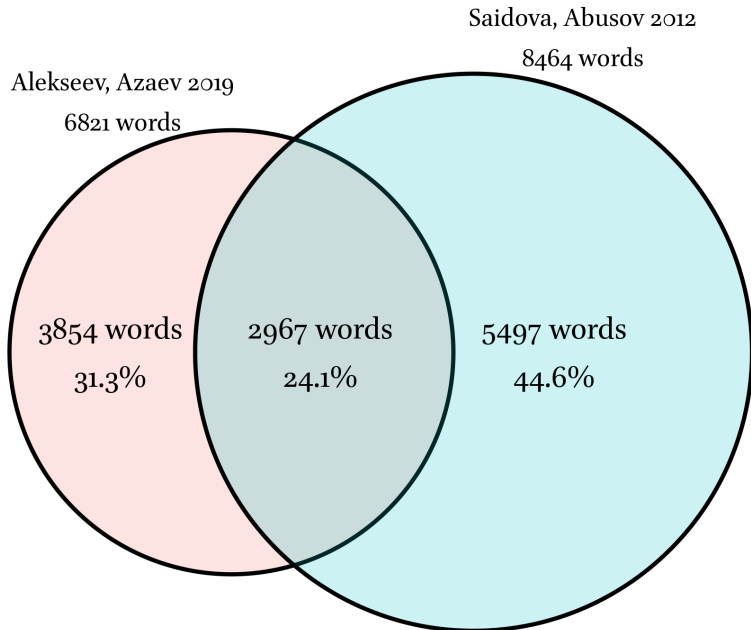
авари, *ми.* -йабалы, *рп.* -лы авария; авари игы совершить аварию

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. *ku: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^wajr

χwardar

miʔar

...

[[Saidova and Abusov 2012](#)]

āhar

bezir

mik'ur

reqχ^wir

reʃkur

rik^war

χwardir

miʔar

...

'message'

'roasting'

'swallowing'

'fight'

'overnight stay'

'lighting'

'digging'

'nose'

...

About 6 cases:

ʃ:alaj

inuʃala

ʃila

...

ʃ:allaj

inuʃalla

ʃilla

...

'silt'

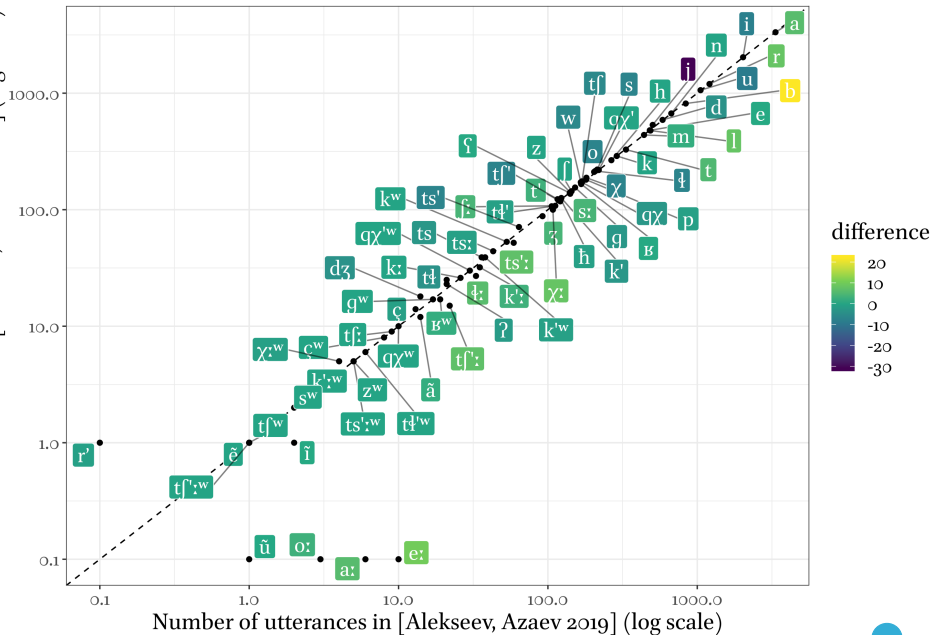
'everywhere'

'reason'

...

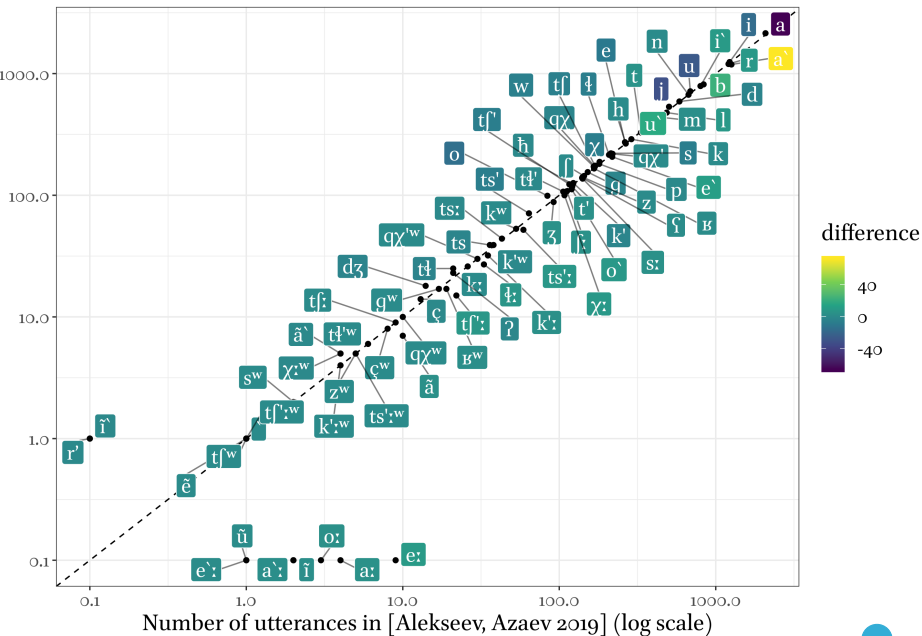
Comparing two Botlikh dictionaries (without stress patterns)

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)



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
- Model computed using row frequencies should be extended with Markov Chains and vector models (like in a Distributional semantics)

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
- Model computed using row frequencies should be extended with Markov Chains and vector models (like in a Distributional semantics): only in this way it will be possible to compare **paradigmatic** and **syntagmatic** relations within the phonological systems and across languages

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 Jazykoznanija AN SSSR.
- Tsertsvadze, I. I. (1965). *Andiuri Ena*. Tbilisi: Metsniereba.