# Phonological and morphological variation in Botlikh: Comparing two dictionaries

George Moroz, Chiara Naccarato, Samira Verhees
Linguistic Convergence Laboratory, NRU HSE

24.03.2020

# Botlikh

- Botlikh < Andic < Avar-Andic-Tsezic < East Caucasian
- Spoken by ~5,000-8,000 speakers
- Three villages in the Botlikh district of the Republic of Daghestan: Botlikh, Miarso, and Ashino
- Unwritten and mostly spoken at home; the Cyrillic script of Avar functions as an ad hoc writing system on social media
- Evaluated as "threatened" by Ethnologue (**simonsfenning2018**), but many children still speak the language and attitudes are positive
- Heavy influence from Avar and Russian

# Botlikh

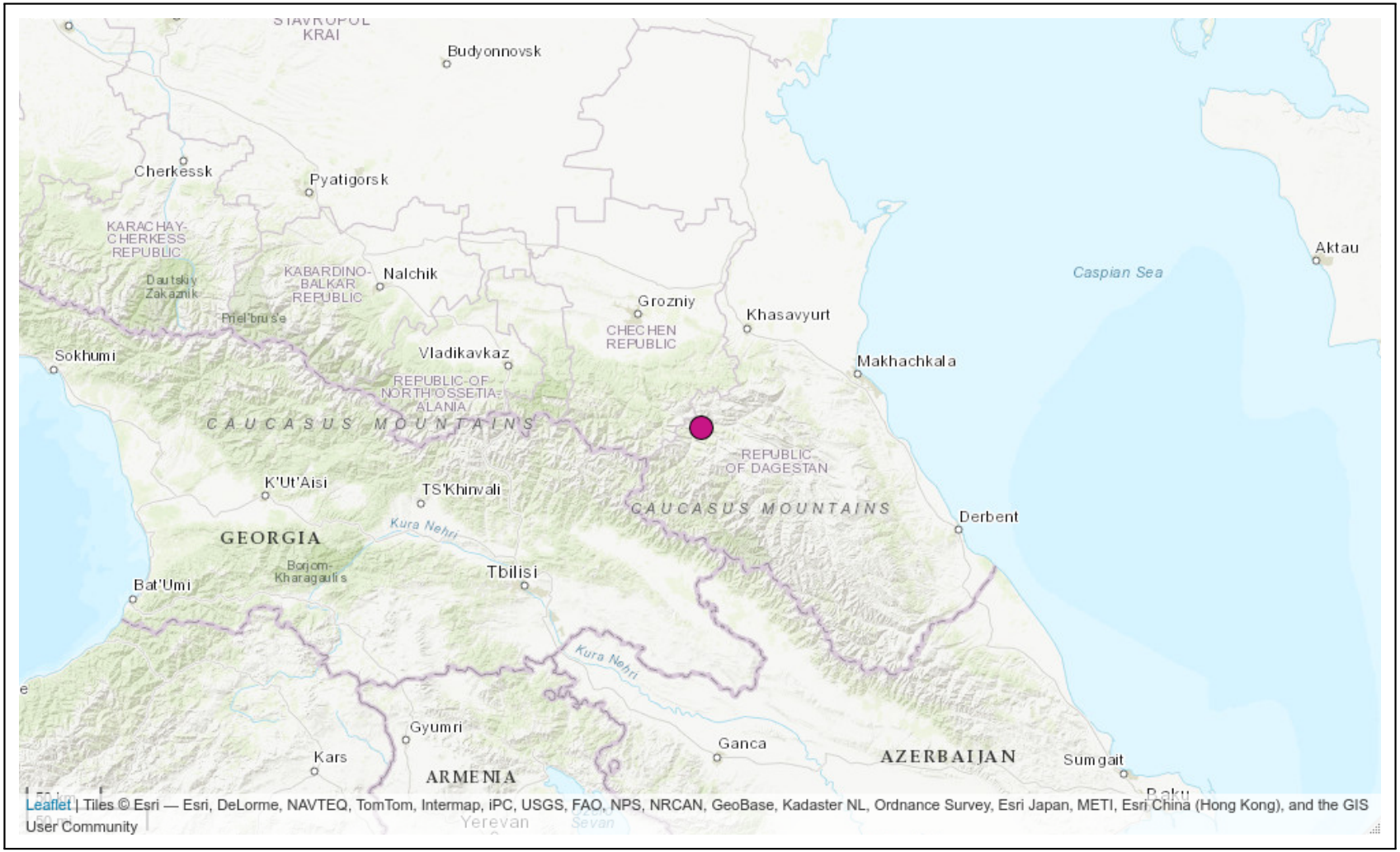



Figure 1: Botlikh on the map

# Botlikh

Figure 2: The old village

# Botlikh literature

- One full reference grammar in Georgian (**gudava1962**)
- Several short sketches mostly based on information contained in the grammar by Togo E. Gudava (**gudava1967**, **azaev2000**, **saidova2001**, **magomedbekova2001**, **xalidova2017**, **alekseevverhees**)
- Several works on the lexicon and word formation (**azaev1975**, **sulejmanova2013**, **alekseev2016**)
- In general poorly described compared to other Andic languages like Godoberi or Bagvalal, BUT two Botlikh-Russian dictionaries are available to date (**saidovaabusov2012**, **alekseev2019**)

# Two dictionaries

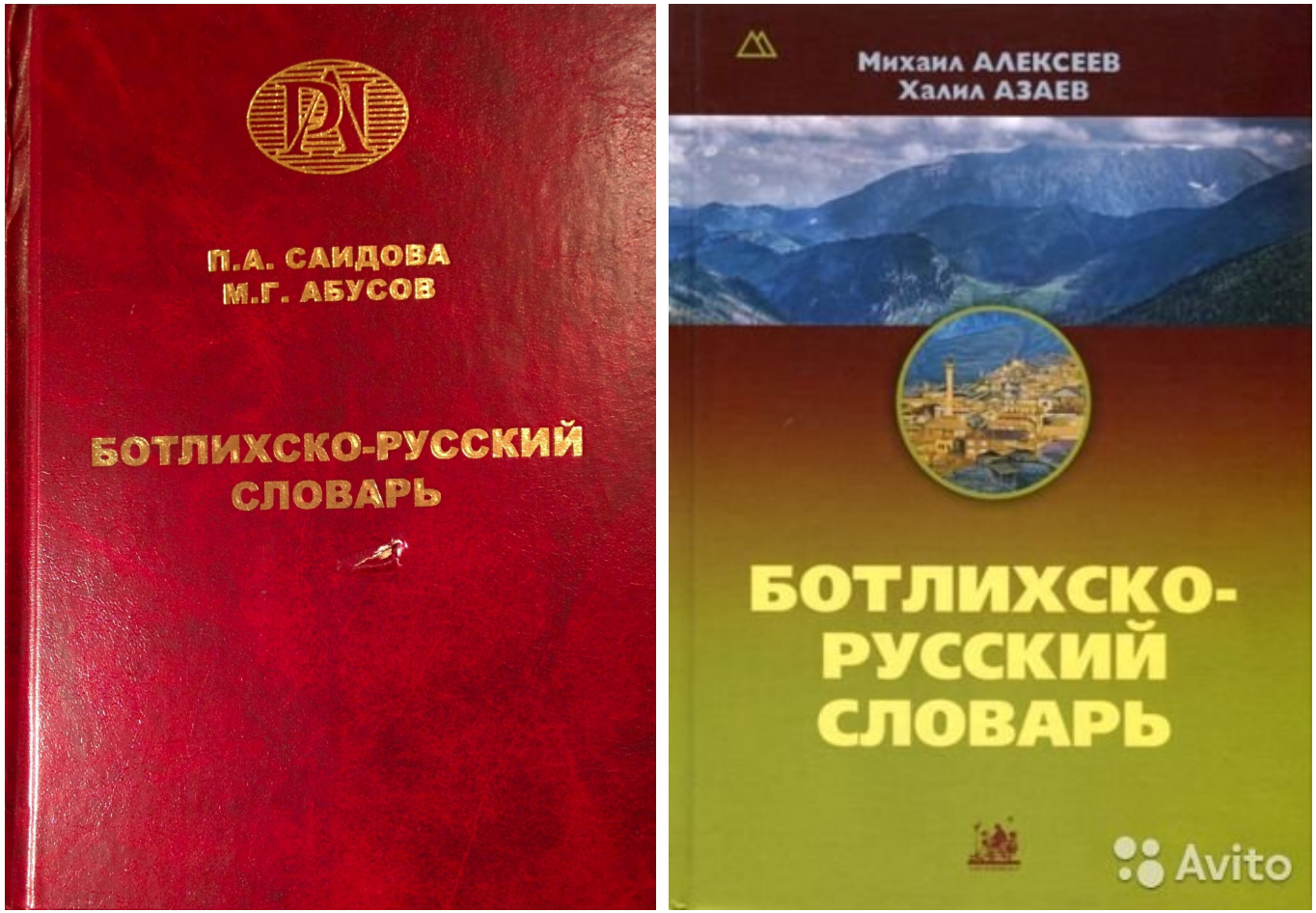



Figure 3: Two Botlikh-Russian dictionaries

# Two dictionaries

- **saidovaabusov2012** compiled in the 2000s by a native speaker of Botlikh (Magomed G. Abusov) and an experienced linguist (Patimat A. Saidova)
- **alekseev2019** compiled in the 1960s/1970s by a native speaker of Botlikh and philologist (Xalil G. Azaev), later (in the 2000s) systematized by an experienced linguist (Mixail E. Alekseev), and published posthumously after the editing by Timur A. Maisak

# Two dictionaries

- Comparable both quantitatively and qualitatively
- ~8,000 headwords for **saidovaabusov2012** vs. ~9,000 words and expressions for **alekseev2019**
- Although the data in **alekseev2019** were collected several decades earlier, Magomed G. Abusov also consulted elderly speakers with the aim of collecting archaic vocabulary
- **saidovaabusov2012** also contains some notes on Miarso; reference to Miarso variants is not explicit in **alekseev2019**, but it seems that such variants are occasionally reported in this dictionary too
- No metadata on the speakers consulted
- At first glance, the two resources seemed to display variation

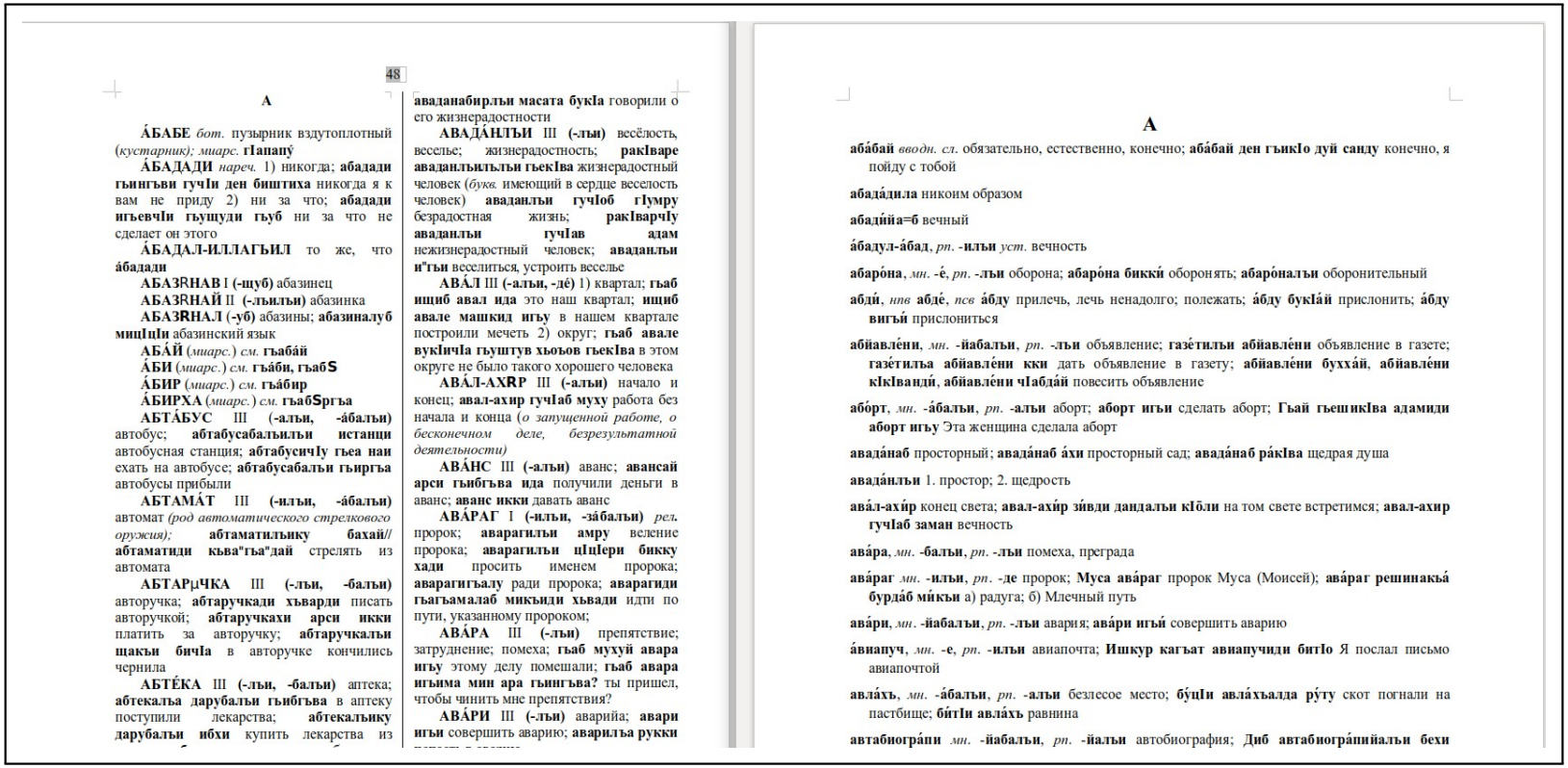
# Our research

- Comparison of the two resources
- A quantitative investigation of an understudied language
- Provide numerical approximations for the impressionistic observations available in the existing literature
- Analysis of both phonological and morphological features
- Detect patterns of systematic variation within these two areas

# Outline

- Data
  - merging
  - extracting grammatical information
  - pairing and annotation
- Analysis
  - phonology (George Moroz)
  - nominal morphology (Chiara Naccarato)
  - verbal morphology (Samira Verhees)
- Results and discussion
- Methodological remarks

# Merging (George Moroz)



48

А

**А́БАБЕ** *бот.* пузырник вздутоплотный (*кустарник*); *миарс.* **гIапапу́**

**А́БАДАДИ** *нареч.* 1) никогда; **абадади гьинго̀ъви гучIи ден биштиха** никогда я к вам не приду 2) ни за что; **абадади игъевчIи гьущуди гьуб** ни за что не сделает он этого

**А́БАДАЛ-ИЛЛАГЬИЛ** то же, что **а́бадади**

**АБАЗ҂НАВ I (-щуб)** абазинец

**АБАЗ҂НАЙ** II **(-лъилъи)** абазинка

**АБАЗ҂НАЛ (-уб)** абазины; **абазиналуб мицIцIи** абазинский язык

**АБА́Й** (*миарс.*) *см.* **гьабáй**

**А́БИ** (*миарс.*) *см.* **гьа́би, гьабS**

**А́БИР** (*миарс.*) *см.* **гьа́бир**

**А́БИРХА** (*миарс.*) *см.* **гьабSргьа**

**АБТА́БУС** III **(-алъи, -а́балъи)** автобус; **абтабусабалъилъи истанци** автобусная станция; **абтабусичIу гьеа наи** ехать на автобусе; **абтабусабалъи гьиргьа** автобусы прибыли

**АБТАМА́Т** III **(-илъи, -а́балъи)** автомат *(род автоматического стрелкового оружия);* **абтаматилъику бахай// абтаматиди кьва"гьа"дай** стрелять из автомата

**АБТАРµЧКА** III **(-лъи, -балъи)** авторучка; **абтаручкади хъварди** писать авторучкой; **абтаручкахи арси икки** платить за авторучку; **абтаручкалъи щакъи бичIа** в авторучке кончились чернила

**АБТЕ́КА** III **(-лъи, -балъи)** аптека; **абтекалъа дарубалъи гьибгьва** в аптеку поступили лекарства; **абтекалъику дарубалъи ибхи** купить лекарства из

**аваданабирлъи масата букIа** говорили о его жизнерадостности

**АВАДА́НЛЪИ** III **(-лъи)** весёлость, веселье; жизнерадостность; **ракIваре аваданлъилъи гьекIва** жизнерадостный человек (*букв.* имеющий в сердце веселость человек) **аваданлъи гучIоб гIумру** безрадостная жизнь; **ракIварчIу аваданлъи гучIав адам** нежизнерадостный человек; **аваданлъи и"гьи** веселиться, устроить веселье

**АВА́Л** III **(-алъи, -дé)** 1) квартал; **гьаб ищиб авал ида** это наш квартал; **ищиб авале машкид игъу** в нашем квартале построили мечеть 2) округ; **гьаб авале вукIичIа гьуштув хьоъов гьекIва** в этом округе не было такого хорошего человека

**АВА́Л-АХ҂Р** III **(-алъи)** начало и конец; **авал-ахир гучIаб муху** работа без начала и конца (*о запущенной работе, о бесконечном деле, безрезультатной деятельности)*

**АВА́НС** III **(-алъи)** аванс; **авансай арси гьибгьва ида** получили деньги в аванс; **аванс икки** давать аванс

**АВА́РАГ** I **(-илъи, -за́балъи)** *рел.* пророк; **авараги лъи амру** веление пророка; **авараги лъи цIцIери бикку хади** просить именем пророка; **авараги гьагьамалаб микъиди хъвади** идти по пути, указанному пророком;

**АВА́РА** III **(-лъи)** препятствие; затруднение; помеха; **гьаб мухуй авара игъу** этому делу помешали; **гьаб авара игьима мин ара гьингьва?** ты пришел, чтобы чинить мне препятствия?

**АВА́РИ** III **(-лъи)** аварийа; **авари игьи** совершить аварию; **аварилъа рукки**

А

**аба́бай** *вводн. сл.* обязательно, естественно, конечно; **аба́бай ден гьикIо дуй санду** конечно, я пойду с тобой

**абада́дила** никоим образом

**абади́йа=б** вечный

**а́бадул-а́бад**, *рп.* **-илъи** *уст.* вечность

**абаро́на**, *мн.* **-е́**, *рп.* **-лъи** оборона; **абаро́на бикки́** оборонять; **абаро́налъи** оборонительный

**абди́**, *нпв* **абде́**, *псв* **а́бду** прилечь, лечь ненадолго; полежать; **а́бду букIа́й** прислонить; **а́бду вигьи́** прислониться

**абйавле́ни**, *мн.* **-йабалъи**, *рп.* **-лъи** объявление; **газе́тилъи абйавле́ни** объявление в газете; **газе́тилъа абйавле́ни кки** дать объявление в газету; **абйавле́ни буххáй, абйавле́ни кIкIвандú, абйавле́ни чIабда́й** повесить объявление

**або́рт**, *мн.* **-а́балъи**, *рп.* **-алъи** аборт; **аборт игьи** сделать аборт; **Гьай гьешикIва адамиди аборт игьу** Эта женщина сделала аборт

**авада́наб** просторный; **авада́нав а́хи** просторный сад; **авада́нав ра́кIва** щедрая душа

**авада́нлъи** 1. простор; 2. щедрость

**ава́л-ахи́р** конец света; **авал-ахи́р зи́вди дандалъи кIо́ли** на том свете встретимся; **авал-ахир гучIаб заман** вечность

**ава́ра**, *мн.* **-балъи**, *рп.* **-лъи** помеха, преграда

**ава́раг** *мн.* **-илъи**, *рп.* **-де** пророк; **Муса ава́раг** пророк Муса (Моисей); **ава́раг решинакьа́ бурда́б ми́къи** а) радуга; б) Млечный путь

**ава́ри**, *мн.* **-йабалъи**, *рп.* **-лъи** авария; **ава́ри игьи** совершить аварию

**а́виапуч**, *мн.* **-е**, *рп.* **-илъи** авиапочта; **Ишкур кагъат авиапучиди битIо** Я послал письмо авиапочтой

**авла́хъ**, *мн.* **-а́балъи**, *рп.* **-алъи** безлесое место; **бу́цIи авла́хъалда ру́ту** скот погнали на пастбище; **би́тIи авла́хъ** равнина

**автабиогра́пи** *мн.* **-йабалъи**, *рп.* **-йалъи** автобиография; **Диб автабиогра́пийалъи бехи**

Figure 4: Merging the dictionaries

# Merging (George Moroz)

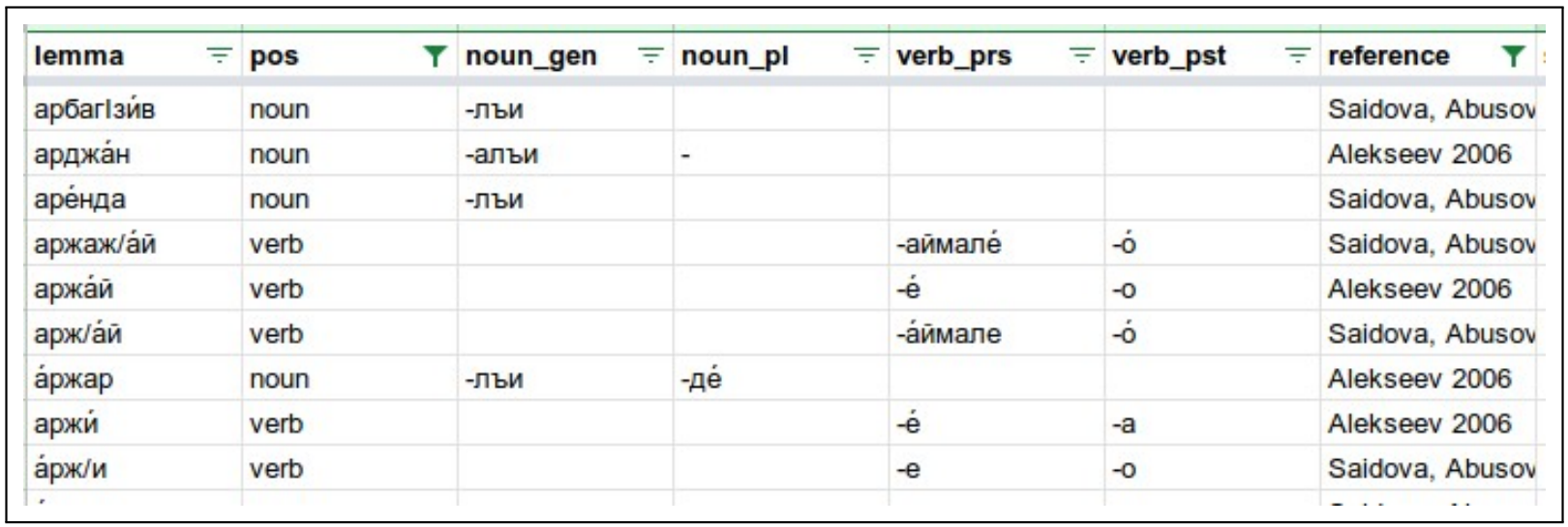
From two .doc files to one .xls:

- .doc preprocessing: convert files to .html, change non-standard symbols, solve several tag problems (I, $^{H}$, etc.)
- extract first bold entrance, parse parenthesis for grammar information

# Extracting grammatical information

- Total number of lexemes extracted: 8,464 from **saidovaabusov2012** and 6,821 from **alekseev2019**
- Nouns: 2,871 from **saidovaabusov2012** and 3,097 from **alekseev2019**
  - grammatical information: genitive and plural
- Verbs: 1,504 from **saidovaabusov2012** and 1,640 from **alekseev2019**
  - grammatical information: habitual and aorist

# Extracting grammatical information



| lemma | pos | noun_gen | noun_pl | verb_prs | verb_pst | reference |
|---|---|---|---|---|---|---|
| арбагIзи́в | noun | -лъи | | | | Saidova, Abusov |
| арджа́н | noun | -алъи | - | | | Alekseev 2006 |
| аре́нда | noun | -лъи | | | | Saidova, Abusov |
| аржаж/а́й | verb | | | -аймале́ | -о́ | Saidova, Abusov |
| аржа́й | verb | | | -е́ | -о | Alekseev 2006 |
| арж/а́й | verb | | | -а́ймале | -о́ | Saidova, Abusov |
| а́ржар | noun | -лъи | -де́ | | | Alekseev 2006 |
| аржи́ | verb | | | -е́ | -а | Alekseev 2006 |
| а́рж/и | verb | | | -е | -о | Saidova, Abusov |

Figure 5: The database

# Pairing

We manually checked for lexemes represented in both dictionaries to carry out phonological and morphological analysis

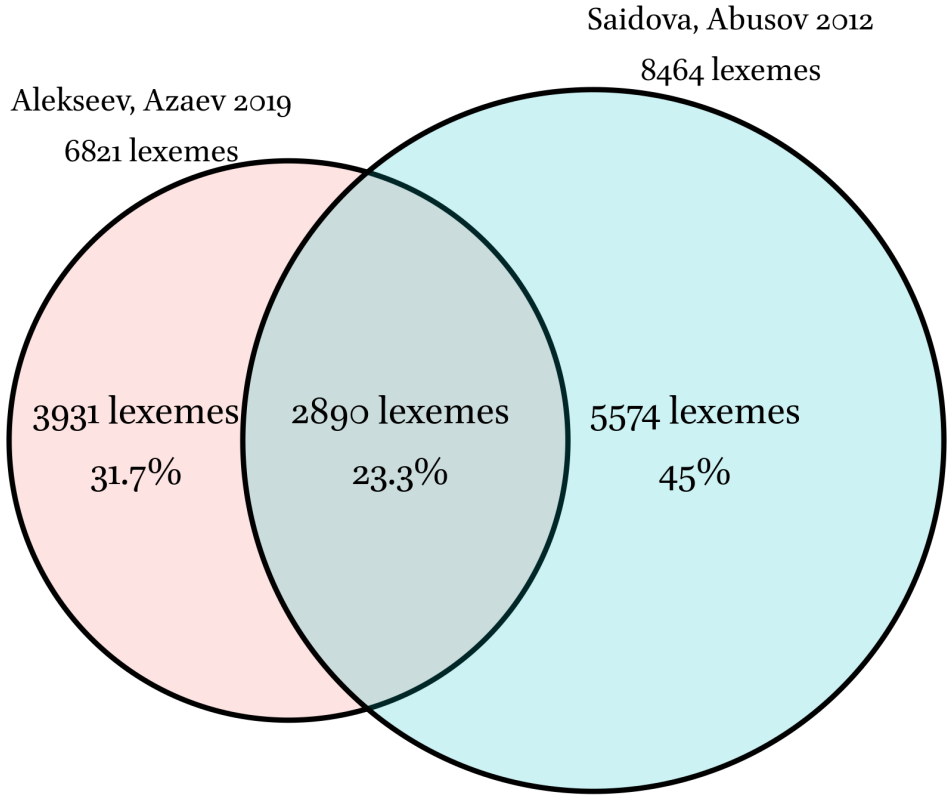



Figure 6: The database

# Pairing

We manually checked for lexemes represented in both dictionaries to carry out phonological and morphological analysis

- Those 2,890 include all POS
- There are multiple entries for the same root
  **alekseev2019:** *besqχe* 'behind', *besqχéku* 'from behind', *besqé-ssu-b* 'back/hind (adjective)', *héči besqérussubla*'in the end'
- There are also some sayings and proverbs (about 670) that constitute separate entries in **alekseev2019**
  *héči besqχérussubla* 'in the end'

# Annotation

- Manual correction of automatically extracted information about grammatical features
- Addition of further annotation (for features that appeared to be potentially relevant for our research):
  - masdars
  - borrowings
- (In progress) Filter out superfluous entries for multi-word expressions and inflected forms

# Phonology

- There are 1,996 lexemes which look phonetically the same, while 909 are different (31%)
- If we remove the stress sign, there are 2,449 lexemes which look phonetically the same, and 456 are different (16%)

⇒ 15% of lexemes have different stress pattern?...

# Phonology

- There are 1,996 lexemes which look phonetically the same, while 909 are different (31%)
- If we remove the stress sign, there are 2,449 lexemes which look phonetically the same, and 456 are different (16%)

⇒ 15% of lexemes have different stress pattern?... Yes, but including 265 (9%) cases where the stress is present in one dictionary and absent in the other

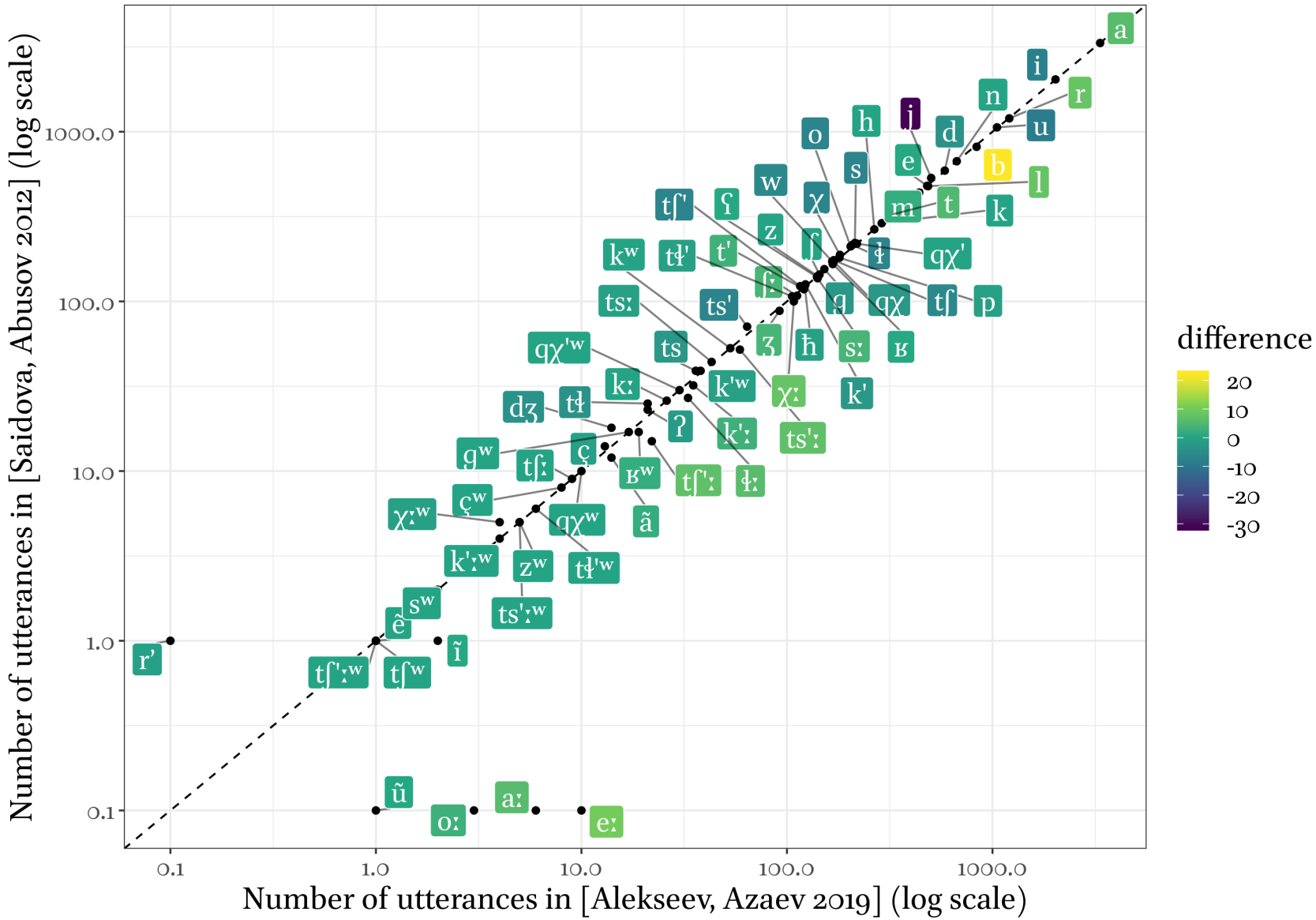
# Phonology

- What causes the difference between the two dictionaries?
  - Stress pattern differences in 188 lexemes (about 6%)
  - Multiple cases where there is a small difference that could be explained either as a typo or in terms of phonological variation: *čuhí* ‘to run’ [aa] vs. *čũhí* [sa], *kusu* ‘cherry plum’ [aa] vs. *kussu* [sa]
  - Multiple cases where Russian borrowings were adopted differently: *a**w**tobus* ‘bus’ [aa] vs. *a**b**tabus* [sa], *bit**o**n* ‘milk can’ [aa] vs. *bit**u**n* [sa], *a**p**teka* ‘pharmacy’ [aa] vs. *a**b**teka* [sa]
  - Morphological preferences: *dinija=**w*** ‘pious’ [aa] vs. *dinija=**b*** [sa]

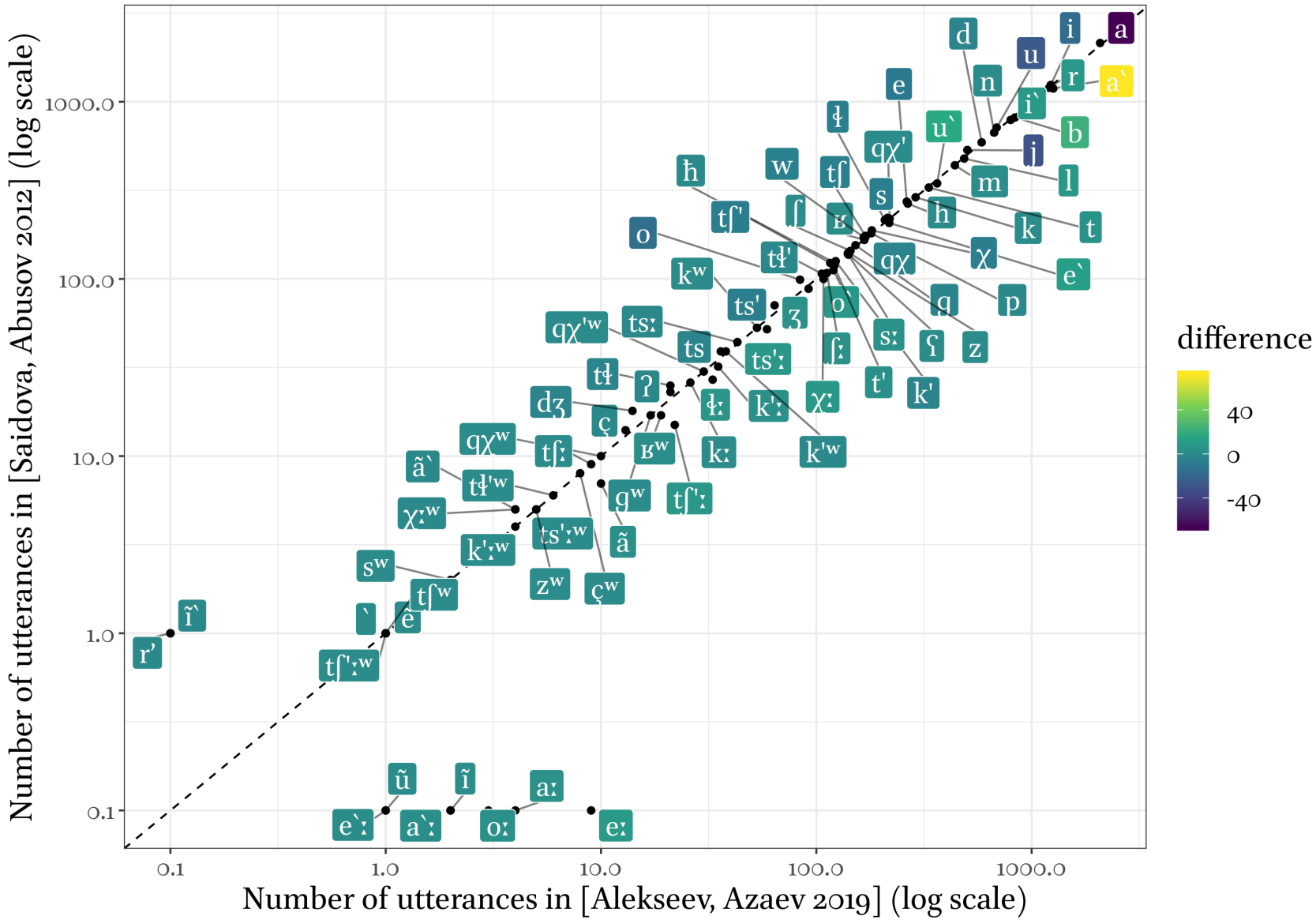
# Phonology: segments

| **alekseev2019** | **saidovaabusov2012** | |
|---|---|---|
| **About 25 cases:** | | |
| *ãhajr* | *ãhar* | 'message' |
| *bezajr* | *bezir* | 'roasting' |
| *mik'kujr* | *mik'ːur* | 'swallowing' |
| *reqχujr* | *reqχʷir* | 'fight' |
| *reʃkujr* | *reʃkur* | 'overnight stay' |
| *rikʷajr* | *rikʷar* | 'lighting' |
| *χwardar* | *χwardir* | 'digging' |
| *miʔar* | *miʔar* | 'nose' |
| ... | ... | ... |
| **About 6 cases:** | | |
| *ʃːalaj* | *ʃːallaj* | 'silt' |
| *inuʕala* | *inuʕalla* | 'everywhere' |
| *ʕila* | *ʕilla* | 'reason' |
| ... | ... | ... |

# Phonology: compare Botlikh segments



# Phonology: compare Botlikh segments

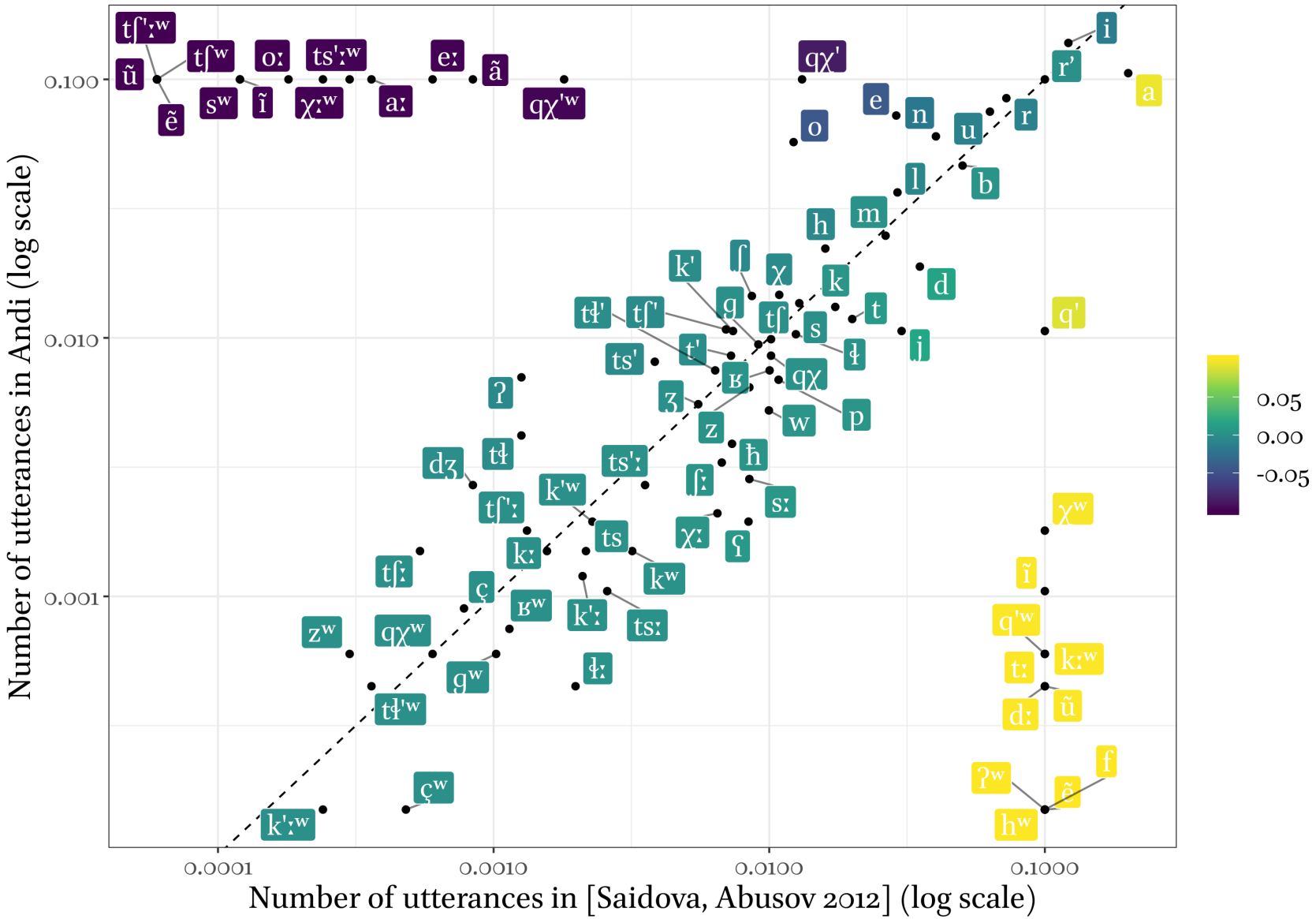


# Zilo Andi data

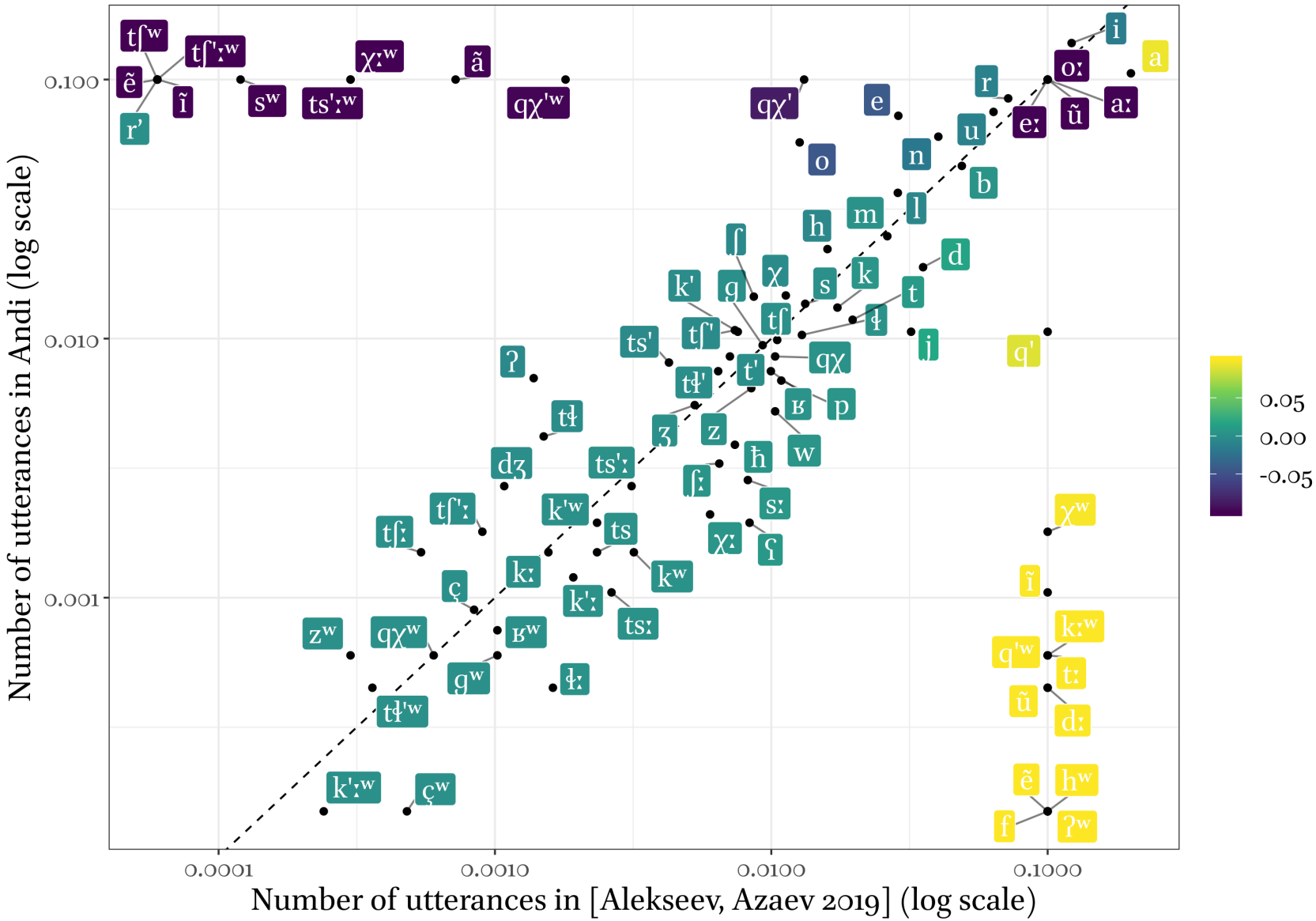
Dictionary data for Zilo were collected during fieldtrips to Zilo in 2016–2019 with N. Rochant, S. Verhees, A. Martynova and A. Zakirova who contributed to the same FieldWorks project

- Contain morphological affixes
- Do not contain additional affixes in a lemma form
- Contain different stems of the same lexeme (e.g. SG.ABS, SG.OBL, PL.ABS, PL.OBL, PST, NPST); those forms were removed during the analysis
- No information about stress

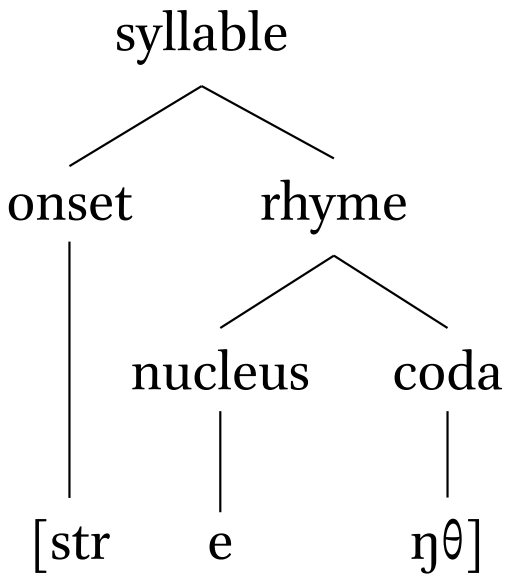
# Phonology: compare Botlikh and Zilo segments



# Phonology: compare Botlikh and Zilo segments

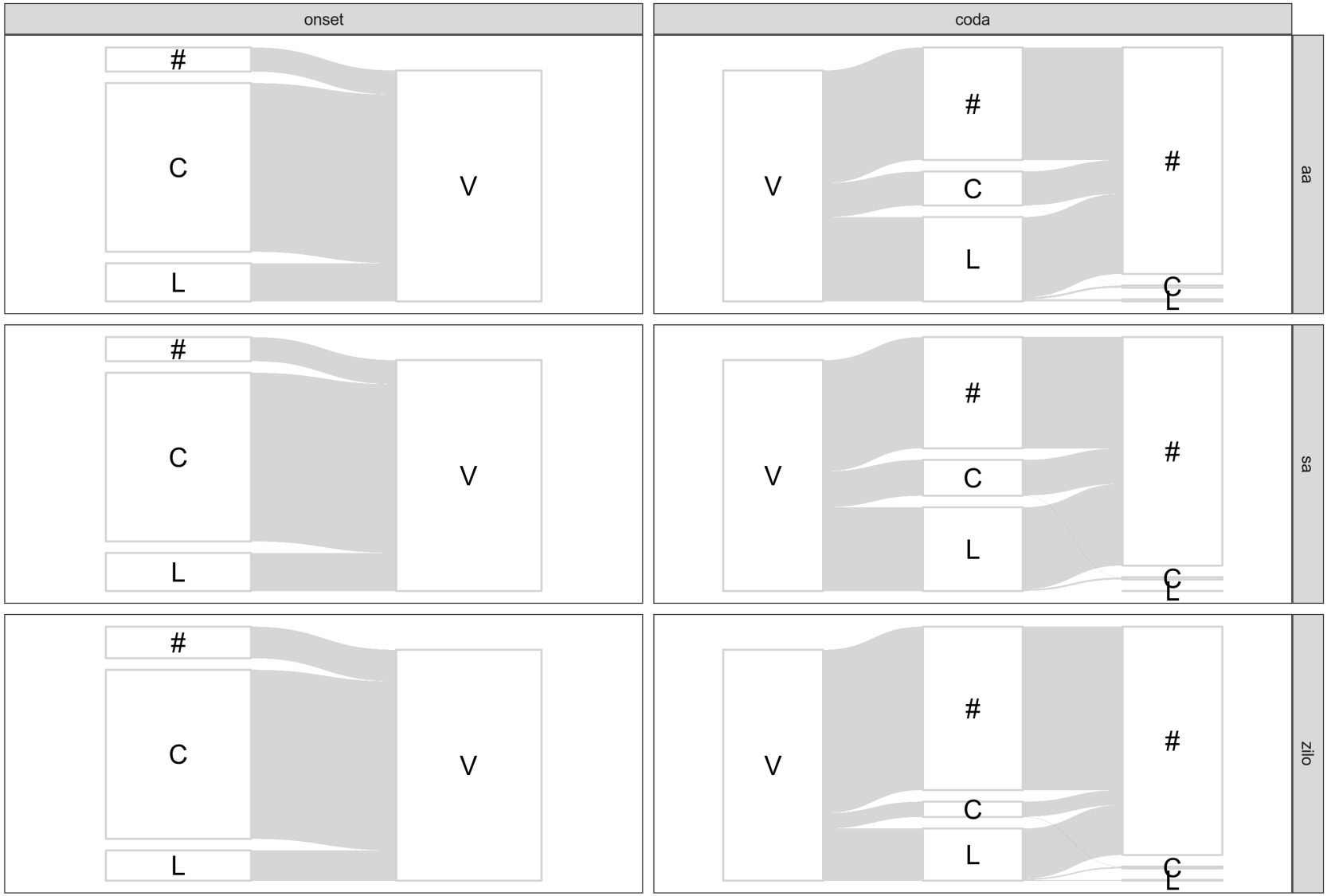


# Phonology: data–driven analysis of syllables




- Analyze all onsets of initial syllables in the corpus
- Analyze all codas of final syllables in the corpus
- Generalize obtained initials and codas into a syllable model
- Check, whether this model describes all intervocal consonant clusters

# Phonology: data–driven analysis of syllables




C — obstruent, L — sonorant, V — vowel, # — syllable boundary

# Nominal morphology

Two topics investigated:

- Formation of the plural
  - to check the productivity of different suffixes
- Formation of the genitive
  - to study alternations in the formation of oblique stems

Comparison of the two resources to look for possible variation in such areas of nominal morphology
(based on 1,072 pairs retrieved during the first annotation round)

# Plural formation in Botlikh

- A suffix is attached to the absolutive stem:
  *na* ‘thing’ < *na-**bałi*** ‘things’
- With stems ending in a consonant, the vowel *-a-* is often inserted before the suffix:
  *majmalak* ‘monkey’ < *majmalak-**a-bałi*** ‘monkeys’
- With stems ending in a vowel, alternation can occur:
  *ruš**a*** ‘tree’ < *ruš**i-bałi*** ‘trees’,
  *sal**u*** ‘tooth’ < *sal**a-bałi*** ‘teeth’,
  *burał**i*** ‘pitcher’ < *burał**a-bałi*** ‘pitchers’

# Plural formation in Botlikh

Among the most common suffixes are:

- *-bałi* and allomorphs (*-małi* for stems ending in a nasal, *-wabałi* for stems ending in *-u*, etc.), the variant *-zabałi* (mostly with borrowings)
  *apicer* 'officer' < *apicer-**zabałi*** 'officers'
- *-de* (mostly for stems ending in a sonorant)
  *ambur* 'roof' < *ambur-**de*** 'roofs'
- *-e* and its variant *-we* (for stems ending in *-u*)
  *čan* 'deer' < *čan-**e*** 'deers'

Other, less common, suffixes are: *-(b)dałi*, *-(b)diłi*, *-(a)l*, *-rdi*, *-bala(l)*

# Plural formation in Botlikh

Can our dictionary data help us be more precise about the distribution/frequency/productivity of plural suffixes in Botlikh?

…

Do the two dictionaries show any variation in these respects?

# Plural suffixes in the dictionaries

- Plural suffixes are not reported for all nouns, cf. *singularia tantum* and plural entries (nationalities, *pluralia tantum*)
- Quite often more than one variant is reported

Figure 7: Plural suffixes in the dictionaries

| | **Saidova & Abusov (2012)** | **Alekseev & Azaev (2019)** |
|---|---|---|
| **-(*x*)*baɬi*** | 292 | 298 |
| **-*de*** | 128 | 354 |
| **-(*w*)*e*** | 141 | 239 |
| **other** | 24 | 21 |
| **no plural** | 499 | 193 |

# Plural suffixes in the dictionaries

Figure 8: Plural suffixes in the dictionaries


($\chi^2$ = 59.368, df = 2, p-value = 1.283e-13)

# Plural suffixes in the dictionaries

- Preference for -(*x*)*bałi* over -*de* in **saidovaabusov2012** vs. the opposite trend in **alekseev2019**
- The higher frequency of -*de* in **alekseev2019** is partly due to masdars
  - **saidovaabusov2012** almost never report the plural form for such nouns, whereas **alekseev2019** consistently report -*de*
- Variation often involves (but is not restricted to) borrowings
  - *birgadir* 'foreman' < *birgadir-**zabałi*** vs. *birgadir-**de***
  - *kassir* 'cashier' < *kassir-**zabałi*** vs. *kassir-**de***
- The frequent mentioning of more than one variant might suggest idiosyncratic variation

# Case declension in Botlikh

Two declension types

- I type — the stem does not change when a suffix is attached (mostly stems ending in a vowel and masdars)
  *babu* 'mom' < *babu-**ɬi*** (genitive)
  *masir* 'measurement' < *masir-**ɬi*** (genitive)
- II type — case suffixes are attached to the oblique stem of the noun (mostly stems ending in a consonant, sometimes stems ending in a vowel)
  *askar* 'army' < *askar-**a**-**ɬi*** (genitive)
  *din* 'religion' < *din-**i**-**ɬi*** (genitive)
  *ima* 'father' < *imu-**ɬi*** (genitive)

# Case declension in Botlikh

Can our dictionary data help us be more precise about the patterns of case declension (oblique stem formation) and their frequencies?

…

Do the two dictionaries show any variation in these respects?

# Oblique stem formation in the dictionaries

We used the grammatical information included in the dictionaries (genitive suffix) to investigate oblique stem formation in Botlikh

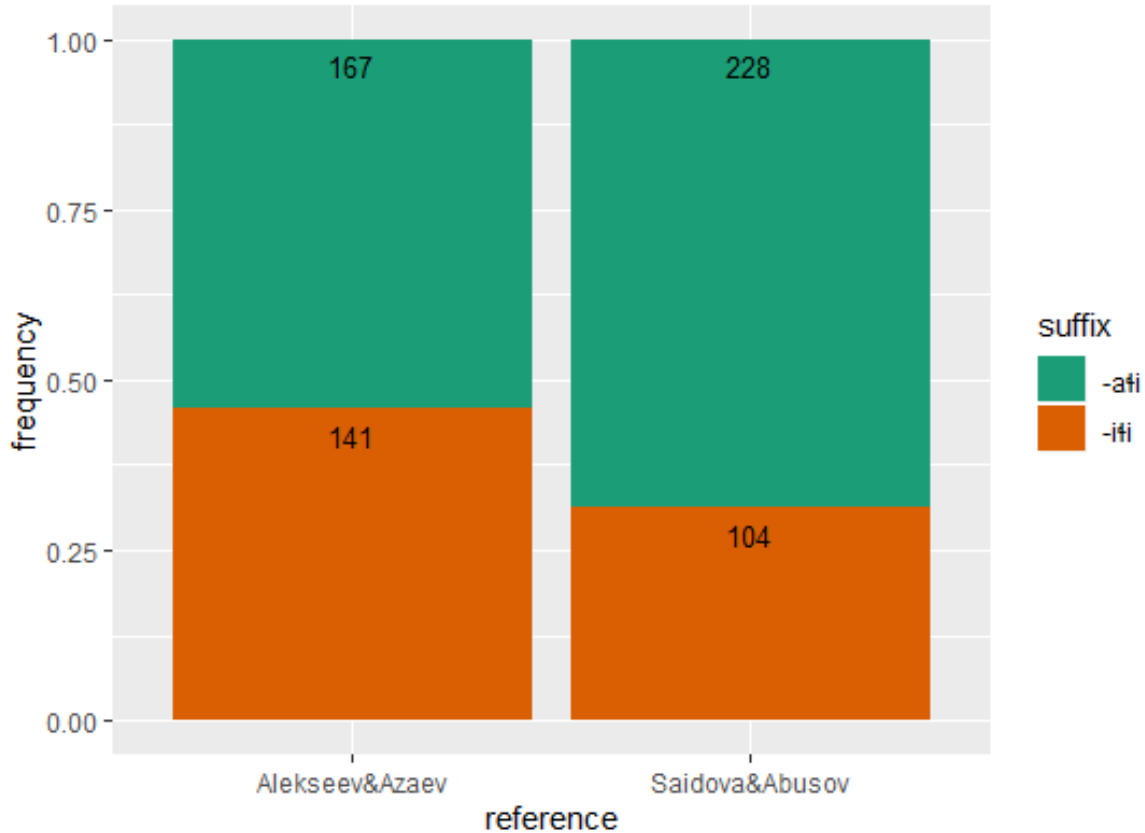
Figure 9: Oblique stem formation in the dictionaries

| | *-ɬi* | | *-a-ɬi* | | *-i-ɬi* | | *-u-ɬi* | |
|---|---|---|---|---|---|---|---|---|
| consonant | 232 | 266 | 228 | 167 | 104 | 141 | - | 1 |
| *-a* | 182 | 167 | - | - | 3 | 13 | 15 | 14 |
| *-i* | 143 | 151 | 10 | 4 | - | - | 3 | 2 |
| *-u* | 81 | 78 | 1 | - | - | 3 | - | - |
| *-e* | 6 | 6 | - | - | - | - | - | - |
| *-o* | 7 | 7 | - | - | - | - | - | - |

Saidova & Abusov (2012) vs. Alekseev & Azaev (2019)

# Oblique stem formation in the dictionaries

Figure 10: Oblique stem formation for stems ending in a consonant




($\chi^2$ = 13.523, df = 1, p-value = 0.0002357)

# Oblique stem formation in the dictionaries

- Significant variation between the two dictionaries in the formation of oblique stems for nouns ending in a consonant
- This again involves (but is not restricted to) borrowings (a general preference for *-a-* over *-i-* in **saidovaabusov2012**)
  - *dakument* 'document' < *dakument-**a**-ɬi* vs. *dakument-**i**-ɬi*
  - *kassir* 'cashier' < *kassir-**a**-ɬi* vs. *kassir-**i**-ɬi*
  - *adijal* 'blanket' < *adijal-**a**-ɬi* vs. *adijal-**i**-ɬi*
- Different variants for one and the same noun are reported far less frequently as compared to plural suffixes

# Verbal morphology

Formation of present (habitualis) and past (aorist) forms of:

- Basic verbs (infinitive in -*i*)
- Derived verbs (infinitive in -*ɬi*)
- Causative verbs (infinitive in -*a-j*)

Comparison of the two resources to look for possible variation in such areas of verbal morphology
(based on 554 pairs retrieved during the first annotation round)

# Basic verbs

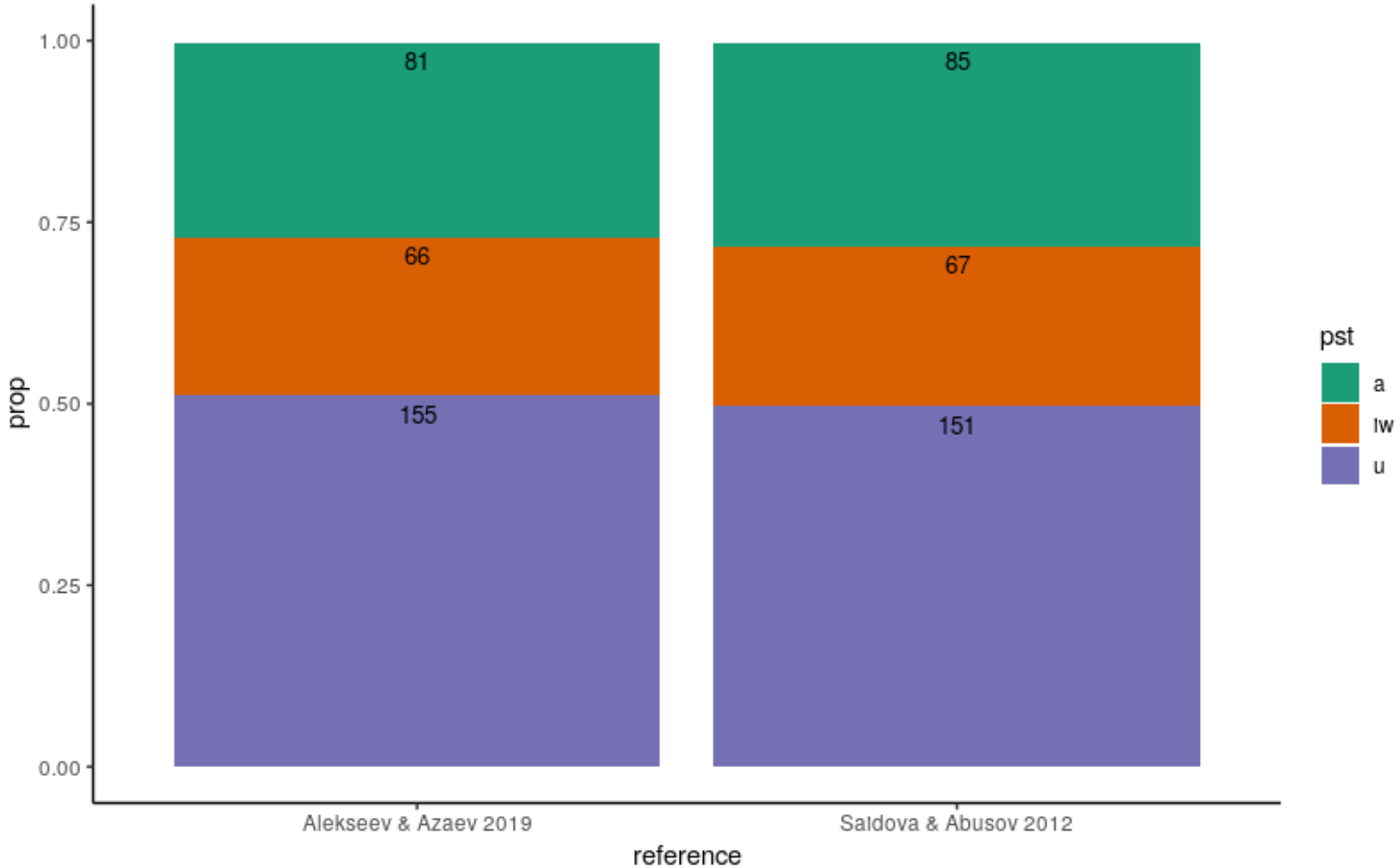
- Habitualis: *-e*
- Aorist: *-a* / *-u* / *-iw*

Figure 11: Basic verbs: inflection

| | Infinitive | Habitualis | Aorist |
|---|---|---|---|
| see | *нав-i* | *нав-e* | *нав-**a*** |
| do | *ih-i* | *ih-e* | *ih-**u*** |
| be able | *bažar-i* | *bažar-e* | *bažar-**iw*** |

# Basic verbs

Figure 12: Aorist suffixes in the dictionaries



# Derived verbs

Analytic formation of both the habitualis and the aorist with auxiliaries

- be: *b-uk'-e*, *b-uk'-a*
- become: *b-ah-e*, *b-ah-u*

Figure 13: Derived verbs: inflection

| | Infinitive | Habitualis | Aorist |
|---|---|---|---|
| roar | *buda-ɬi* | *buda* ***b-uk'-e*** | *buda* ***b-uk'-a*** |
| bleat | *baʕada-ɬi* | *baʕada* ***b-ah-e*** | *baʕada* ***b-ah-u*** |

# Causative verbs

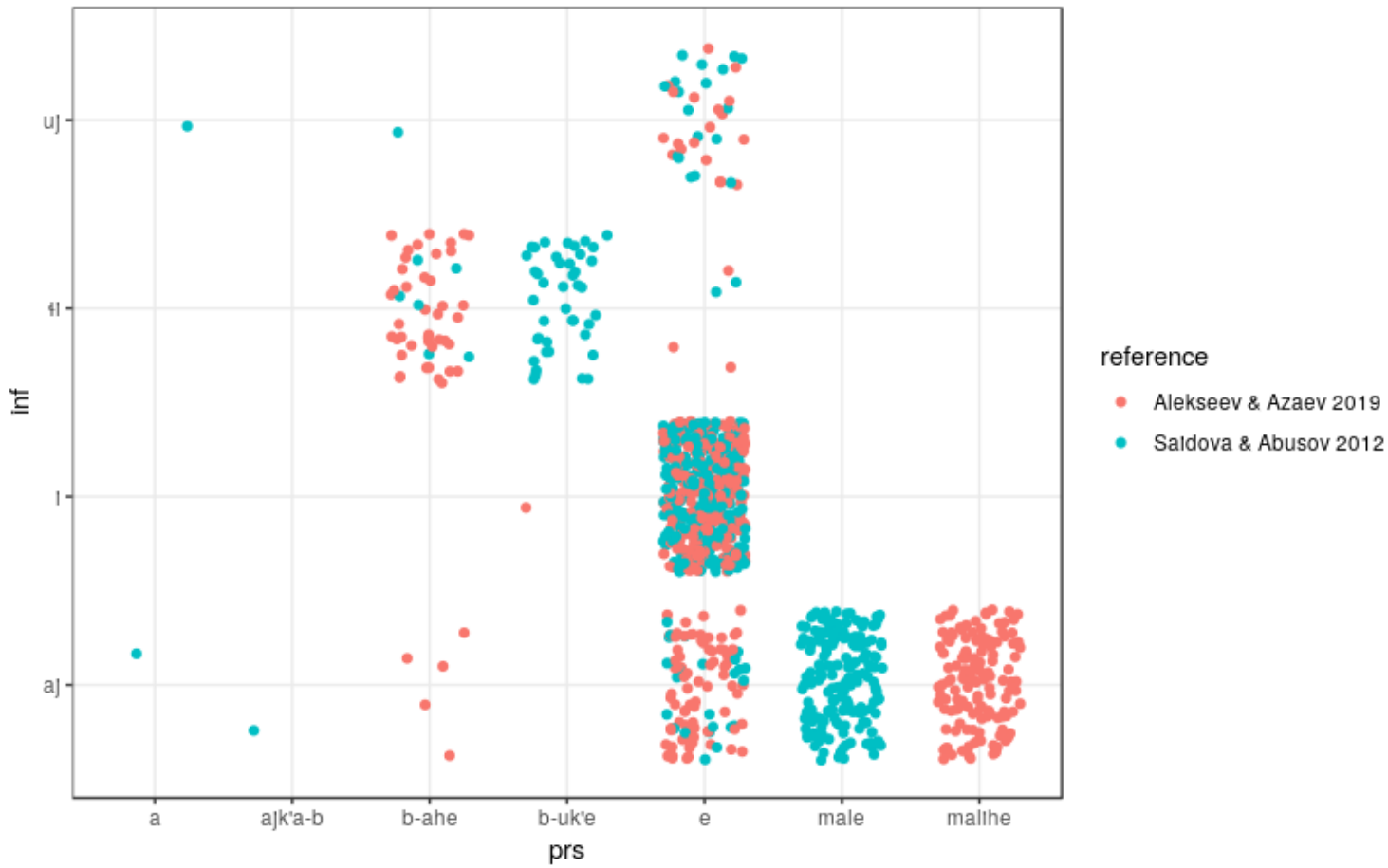
- *-o* < **-a-u* [-CAUS-AOR]
- *-mal-e* a reduced form of *-malih-e*?
- *-mal-o* rarely found in the data

Figure 14: Causative verbs: inflection

| | Infinitive | Habitualis | Aorist |
|---|---|---|---|
| resettle | *guč-a-j* | *guč-**e*** | *guč-**o*** |
| roast | *žad-a-j* | *žad-a-j-**mal-e*** | *žad-a-j-**mal-o*** |
| sew up | *mik'-a-j* | *mik'-a-j-**mal-e*** | *mik'-**o*** |
| sew up | *mik'-a-j* | *mik'-a-j-**malih-e*** | *mik'-a-j-**malih-u*** |

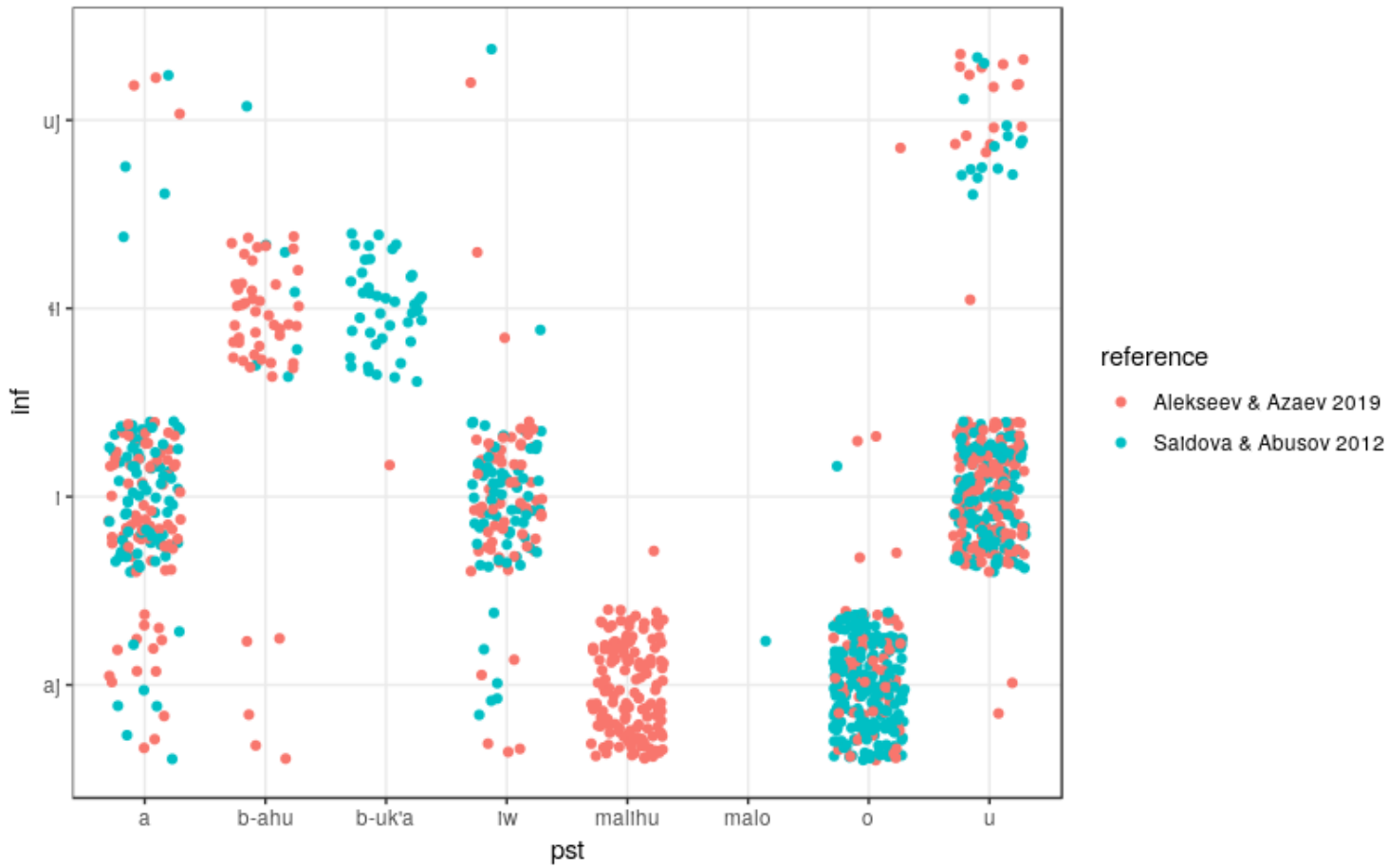
# Habitualis in the dictionaries

Figure 15: Habitualis



# Aorist in the dictionaries

Figure 16: Aorist



# Variation

- Basic verbs display little variation in the aorist (and no variation at all in the habitualis)
- Most variation is observed for derived and causative verbs:
  - Derived verbs: preference for auxiliary 'be' in **saidovaabusov2012** vs. 'become' in **alekseev2019**
    BUT it seems that this is just a matter of personal taste for citation forms: examples in the dictionary entries show that both variants are possible
  - Causative verbs: full (older?) forms *-malih-e* and *-malih-u* in **alekseev2019** vs. reduced form *-mal-e* and synthetic *-o* in **saidovaabusov2012**
    This might be interpreted as diachronic variation, since the data in **alekseev2019** are older

# Variation

The comparison of two dictionaries allowed to identify different types of variation in Botlikh

- Phonological variation
  - stress patterns
  - syllable structure
- Morphological variation
  - nouns: plural suffixes and oblique stem formation
  - verbs: present (habitualis) and past (aorist) forms

Variation seems to affect (to a greater extent) specific groups of words

- nouns: borrowings and masdars (both at the phonological and at the morphological level)
- verbs: derived and causative

# Variation

How do we explain the variation observed?

- Idiolectal variation?
- Diachronic variation?
- Personal preferences of the author?
- ...?

# Methodological remarks

- Two similar datasets collected independently in a small language community approximately in the same time period can nevertheless display considerable variation
- This demonstrates the importance of transparency in data collection i.e. metadata on the speakers consulted
- And methodological decisions
  e.g. what is included or not as a separate dictionary entry

# Methodological remarks

- The availability of comparable material on which quantitative investigations can be conducted is a rare luck for such small languages like Botlikh
- This precious information can be used to provide numerical approximations for impressionistic observations reported in the available literature on the language

# Methodological remarks

- 

# Further research

- Cleaning the data by removing superfluous entries (multi-word expressions, inflected forms)
- Completing the annotation of masdars and borrowings to analyze them separately
- How much variation is left if we do not consider such “hot spots”?

# Further research

- Other Avar-Andic dictionaries are being added (Avar, Godoberi, Karata, Tindi, Chamalal, Bagvalal, Akhvakh)
- The same procedure of data processing and extraction of grammatical information is being carried out
- This will give us the opportunity to conduct comparative analysis at different linguistic levels (phonology, morphology, lexicon) of the different languages of the Avar-Andic branch of EC languages

# The end

# References I