

The speakers of minority languages are more multilingual

Nina Dobrushina and George Moroz

Linguistic Convergence Laboratory, NRU HSE, Moscow, Russia

16 April 2019

Typology of small-scale multilingualism

Laboratoire Dynamique du Langage, Lyon, France

Presentation is available here: tinyurl.com/y6jjp38y



Multilingualism in Daghestan

The Republic of Daghestan is an area of high language density and diversity.

- three language families (the estimated number of languages ranges from 30 to 45):
 - East Caucasian (Nakh-Daghestanian),
 - Turkic: Kumyk, Nogai, Azerbaijani
 - Indo-European: Tat and Russian

Multilingualism in Daghestan

The Republic of Daghestan is an area of high language density and diversity.

- three language families (the estimated number of languages ranges from 30 to 45):
 - East Caucasian (Nakh-Daghestanian),
 - Turkic: Kumyk, Nogai, Azerbaijani
 - Indo-European: Tat and Russian
- The vast majority of Daghestanians are Muslims

Multilingualism in Daghestan

The Republic of Daghestan is an area of high language density and diversity.

- three language families (the estimated number of languages ranges from 30 to 45):
 - East Caucasian (Nakh-Daghestanian),
 - Turkic: Kumyk, Nogai, Azerbaijani
 - Indo-European: Tat and Russian
- The vast majority of Daghestanians are Muslims
- The main occupation in highland villages was shepherding and, to a certain extent, crop farming
- In poor highland settlements people often practiced seasonal jobs outside the village

Multilingualism in Daghestan

The Republic of Daghestan is an area of high language density and diversity.

- three language families (the estimated number of languages ranges from 30 to 45):
 - East Caucasian (Nakh-Daghestanian),
 - Turkic: Kumyk, Nogai, Azerbaijani
 - Indo-European: Tat and Russian
- The vast majority of Daghestanians are Muslims
- The main occupation in highland villages was shepherding and, to a certain extent, crop farming
- In poor highland settlements people often practiced seasonal jobs outside the village
- **Endogamy:** in most Daghestanian villages, wives were taken from the same village

Multilingualism in Daghestan

The Republic of Daghestan is an area of high language density and diversity.

- three language families (the estimated number of languages ranges from 30 to 45):
 - East Caucasian (Nakh-Daghestanian),
 - Turkic: Kumyk, Nogai, Azerbaijani
 - Indo-European: Tat and Russian
- The vast majority of Daghestanians are Muslims
- The main occupation in highland villages was shepherding and, to a certain extent, crop farming
- In poor highland settlements people often practiced seasonal jobs outside the village
- **Endogamy**: in most Daghestanian villages, wives were taken from the same village
- **linguae francae**: **Avar, Azeri, Kumyk, Russian** (in XX century)

Multilingualism in Daghestan

- More than forty languages spoken in Daghestan
- Widespread multilingualism
- Multilingual repertoires were village specific:
 - each village had its own set of second languages
 - occasional knowledge of additional languages was rare
- Multilingualism was distributed unevenly across villages – some were very highly multilingual, some were almost monolingual

Problem setting

Question:

What influences the richness of language repertoire?

Hypothesis:

The number of speakers plays a role

(“Numbers count: a larger culture is likely to be a dominant culture”

– [Thomason 2001: 6]

The aim

To test quantitatively whether the size of language group influences the number of languages they speak

Our data

Data obtained during interviews on language usage from about 15 fieldtrips (see [Dobrushina 2013] for methodology details) and collected into **Atlas of Multilingualism in Daghestan** [Dobrushina et al. 2017]:

- field trips to 17 clusters of villages (2 to 4 villages per cluster); totality of 54 villages
- 24 languages (Russian excluded)

Our data

Data obtained during interviews on language usage from about 15 fieldtrips (see [Dobrushina 2013] for methodology details) and collected into **Atlas of Multilingualism in Daghestan** [Dobrushina et al. 2017]:

- field trips to 17 clusters of villages (2 to 4 villages per cluster); totality of 54 villages
- 24 languages (**Russian excluded**)
- 3210 people born between 1900 and 1959
 - 1564 females (48.7%)
 - 1646 males (51.3%)
- variable containing the number of second languages spoken by each speaker

Our data

Data obtained during interviews on language usage from about 15 fieldtrips (see [Dobrushina 2013] for methodology details) and collected into **Atlas of Multilingualism in Daghestan** [Dobrushina et al. 2017]:

- field trips to 17 clusters of villages (2 to 4 villages per cluster); totality of 54 villages
- 24 languages (**Russian excluded**)
- 3210 people born between 1900 and 1959
 - 1564 females (48.7%)
 - 1646 males (51.3%)
- variable containing the number of second languages spoken by each speaker
- we grouped all languages into three categories according to the number of speakers at the present time
 - **big** — 100 000 speakers and more
 - **medium** — 10 000–30 000 speakers
 - **small** — one village languages, 1 000–2 000 speakers

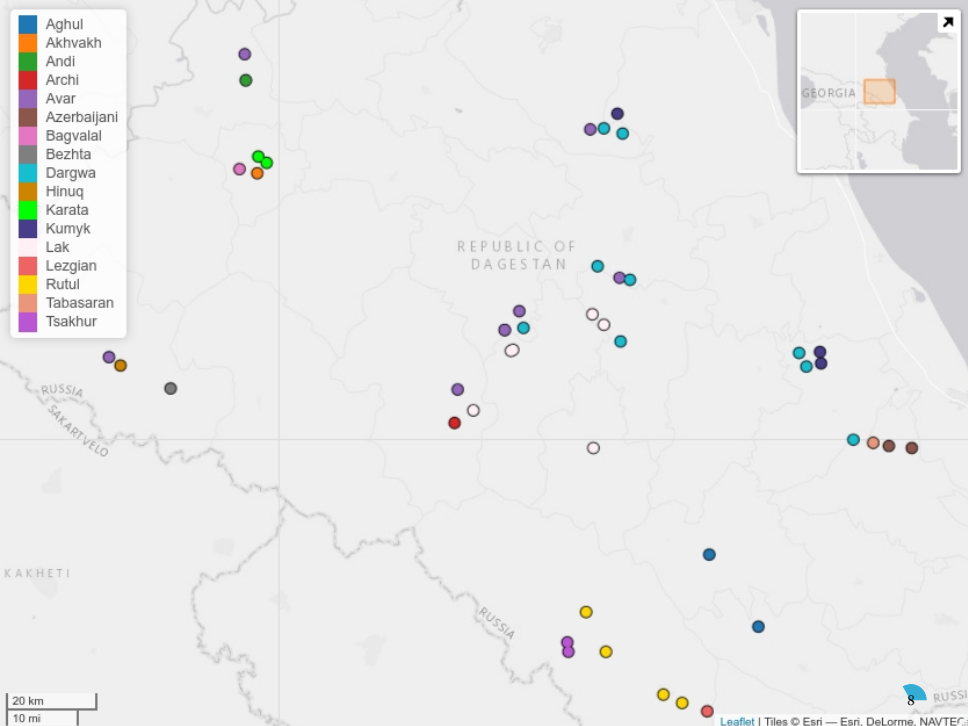
Retrospective family interviews, [Dobrushina 2013]

- Rate of bilingualism at the community level is taken to be a proxy for the intensity of language contact
- Short interviews about language repertoire of locals are taken
- The respondent reports the data not only about himself but also about all his elder relatives whom ((s)he thinks) (s)he remembers

Name	Akaj
Born in	Chabanmakhi
The interviewer was talking to	Umaidat
Family relation to the respondent	Father of Umaidat
Years of birth and death	1900 - 1973
Native language	Kadar Dargwa
Education and living outside the village	worked as a mason, also in other villages
Did he read the Koran?	Yes, could not translate
Did he speak Avar?	yes
Did he speak Kumyk?	yes
Did he speak Russian?	yes
Did he speak any other languages?	no
Literate in	Arabic, Cyrillic

Why retrospective?

- From the establishment of Soviet schools in the 1930s, Russian quickly spread over Daghestan as L2
- Traditional patterns of language contact have been almost completely substituted by Russian as a lingua franca



- Aghul
- Akhvakh
- Andi
- Archi
- Avar
- Azerbaijani
- Bagvalal
- Bezhta
- Dargwa
- Hinuq
- Karata
- Kumyk
- Lak
- Lezgian
- Rutul
- Tabasaran
- Tsakhur



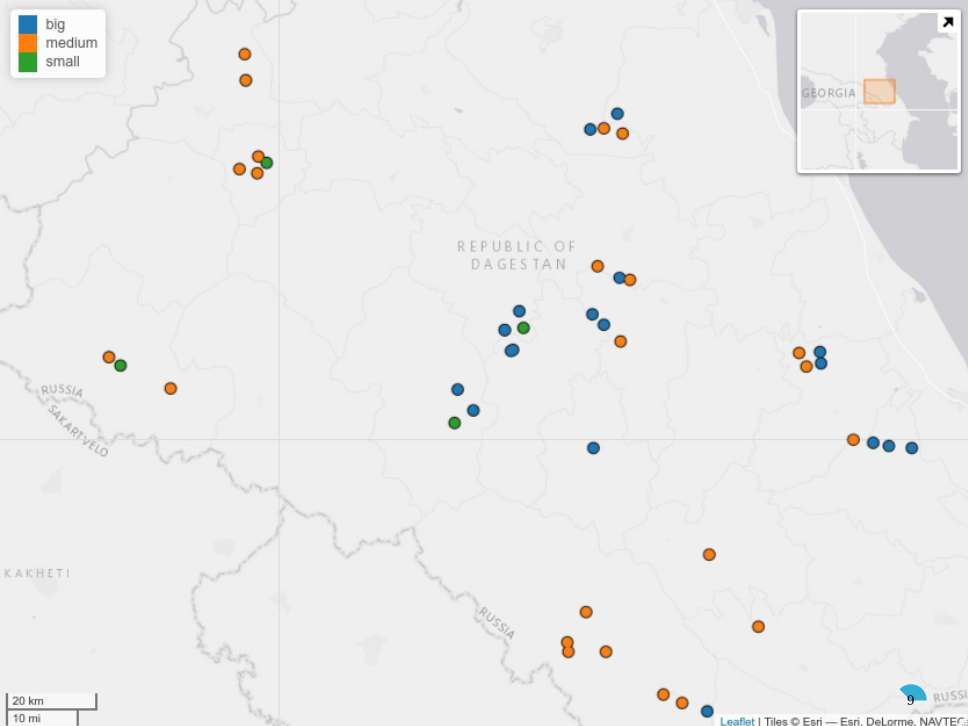
REPUBLIC OF DAGESTAN

RUSSIA
SAKARTVELO

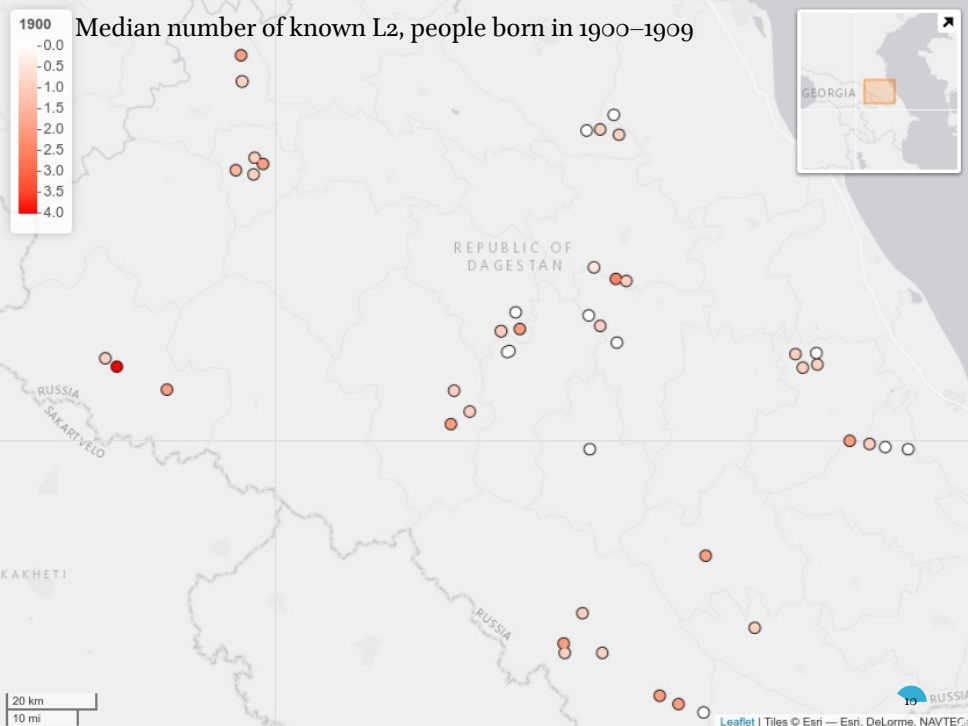
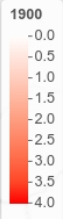
KAKHETI

RUSSIA

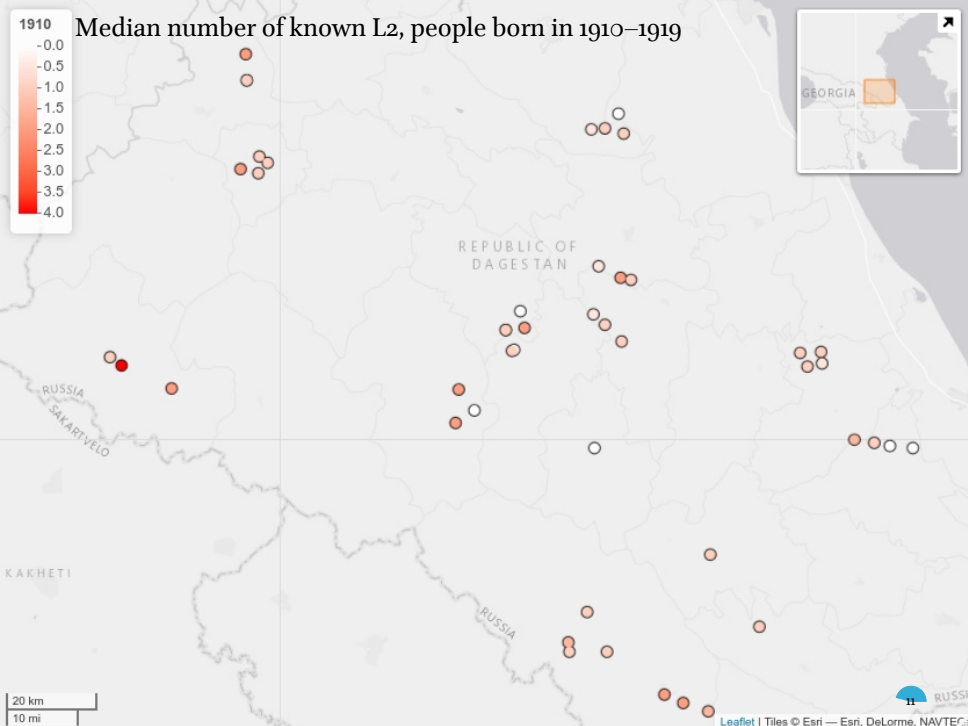
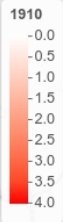
20 km
10 mi



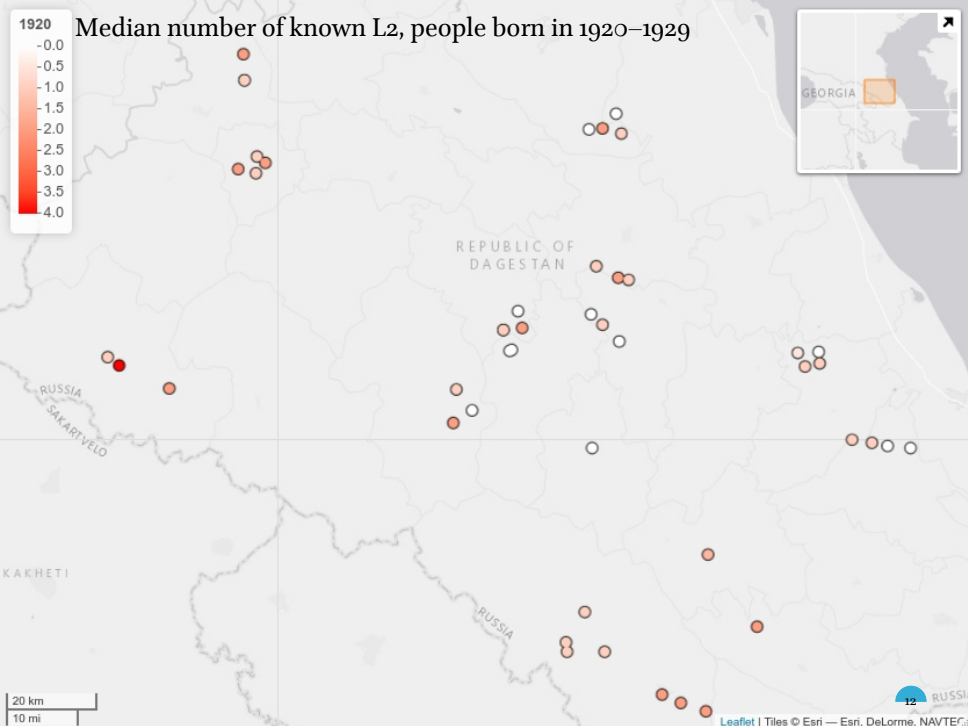
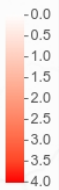
1900 Median number of known L2, people born 1900–1909



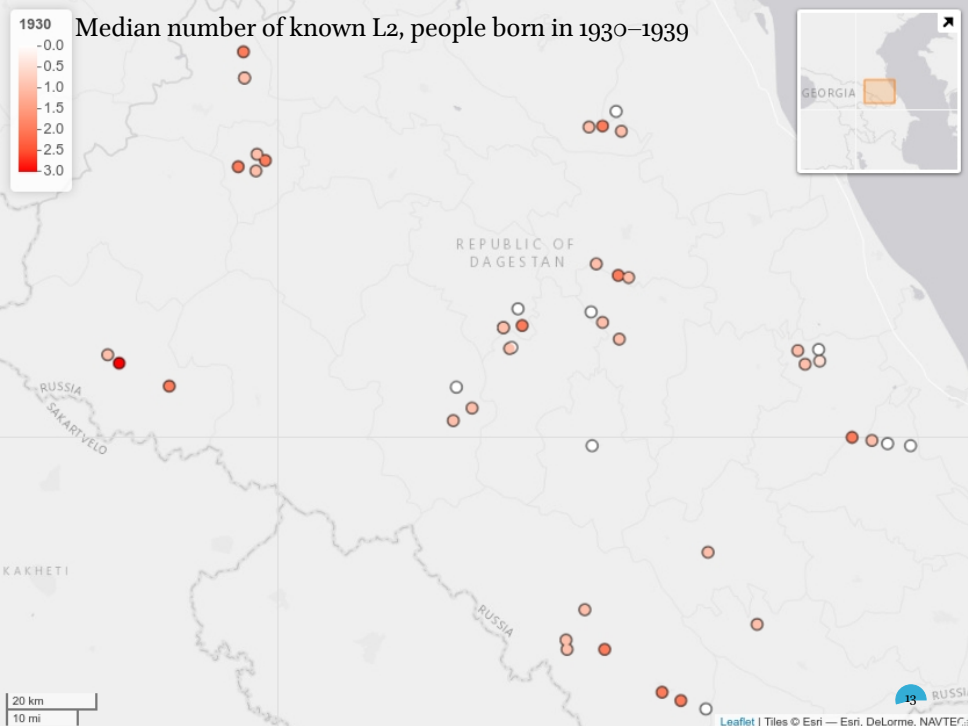
1910 Median number of known L2, people born 1910–1919



1920 Median number of known L2, people born 1920–1929

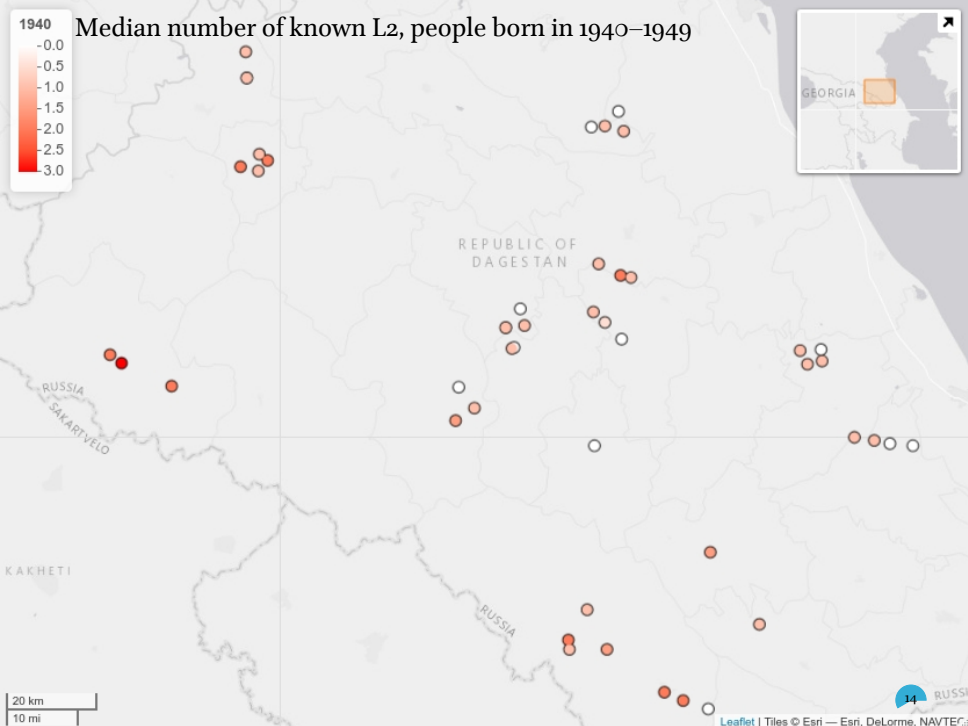
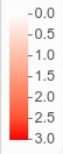


1930 Median number of known L2, people born 1930–1939

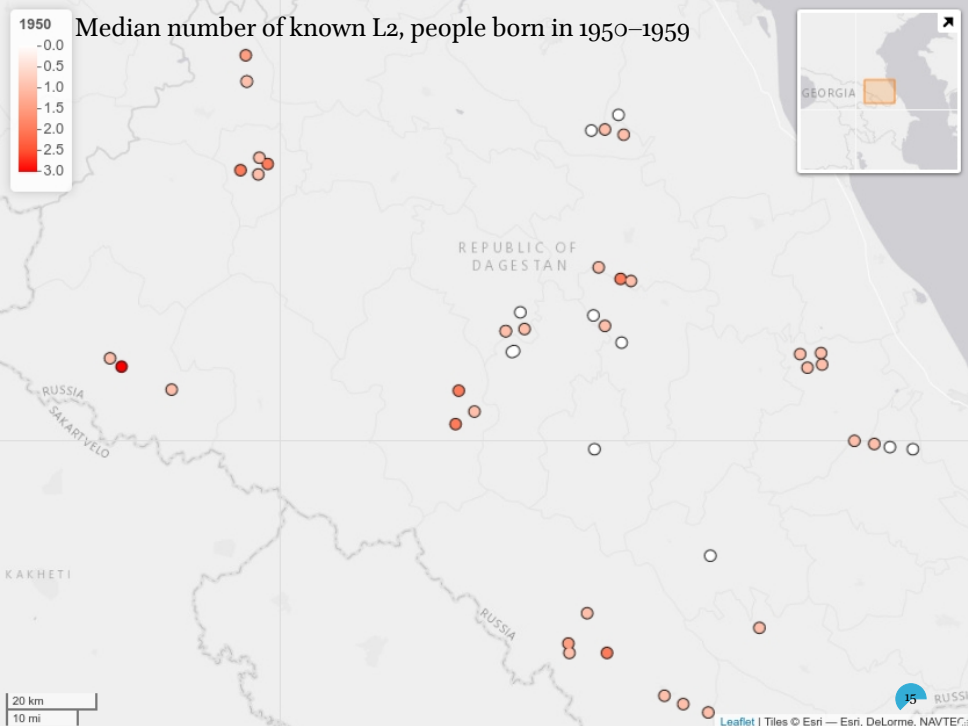
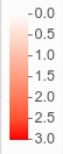


20 km
10 mi

1940 Median number of known L2, people born 1940–1949



1950 Median number of known L2, people born 1950-1959



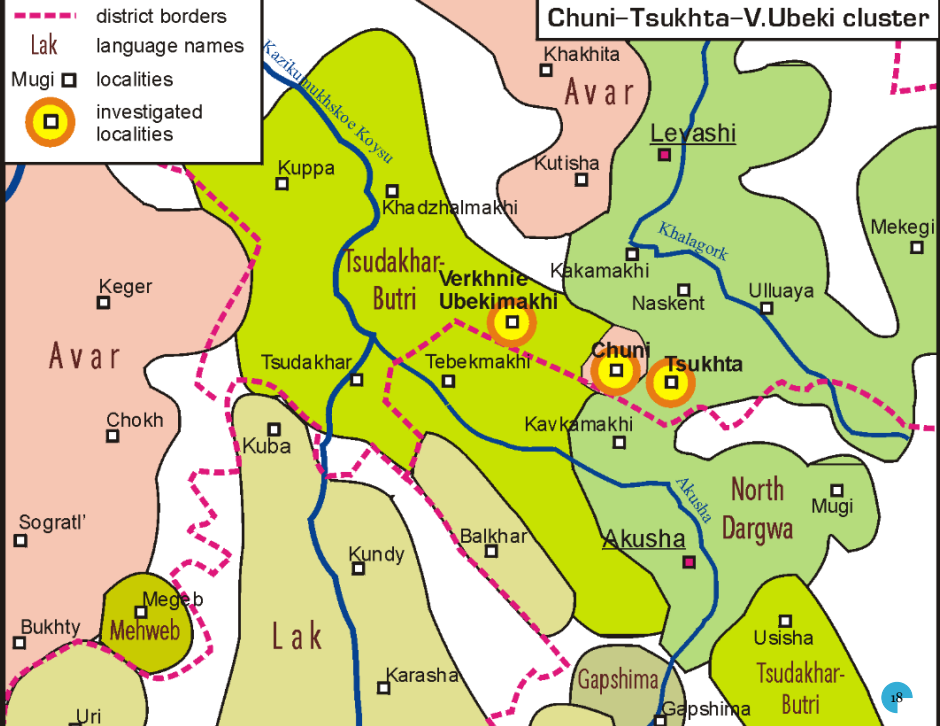
20 km
10 mi



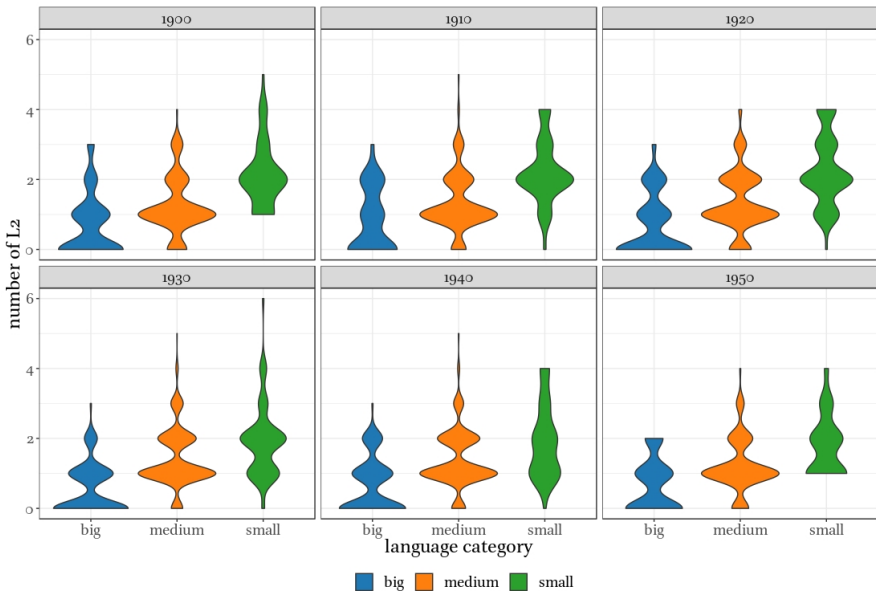
What's going on in Chuni?

- Chuni is an Avar village
- Avar is the biggest Nakh-Daghestanian language (about 700 000)
- Other Avar villages in our sample are close to being monolingual (Chittab, Durangi, Kizhani, Obokh)
- Chuni is an Avar enclave surrounded by Dargwa varieties (Akusha Dargwa and Tsudakhar Dargwa)
- Being a linguistic minority, Chuni people speak both languages

Chuni-Tsukhta-V. Ubeki cluster



Number of L2 in each village by decade and language category



Poisson Mixed Effects Model

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)

Family: poisson (log)

Formula: sum_langs ~ status + (1 | residence.en) + (1 | decade)

Data: df

AIC	BIC	logLik	deviance	df.resid
7502.1	7532.5	-3746.1	7492.1	3195

Scaled residuals:

Min	1Q	Median	3Q	Max
-1.4215	-0.4501	-0.1821	0.3202	3.7685

Random effects:

Groups	Name	Variance	Std.Dev.
residence.en	(Intercept)	0.3679222	0.60657
decade	(Intercept)	0.0004188	0.02046

Number of obs: 3200, groups: residence.en, 46; decade, 6

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	0.7151	0.3052	2.343	0.0191	*
statusmedium	-0.5011	0.3311	-1.514	0.1301	
statusbig	-1.5692	0.3412	-4.599	0.00000424	***

SMALL is stat. s. from o
MEDIUM is NOT stat. s. from SMALL
BIG is stat. s. from SMALL

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

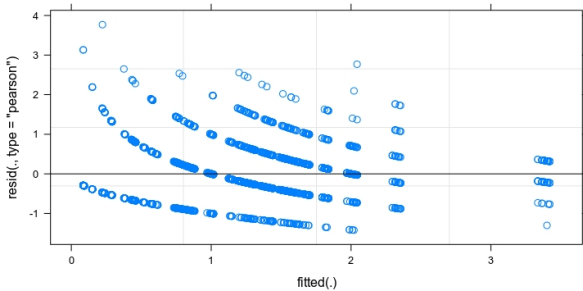
	(Intr)	sttsmd
statusmedim	-0.921	
statusbig	-0.893	0.824

Overdispersion test

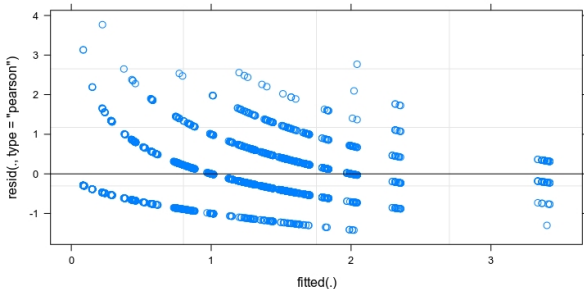
dispersion ratio = 0.4329
Pearson's Chi-Squared = 1383.2471
p-value = 1.0000

No overdispersion detected.

Poisson Mixed Effects Model: Residuals

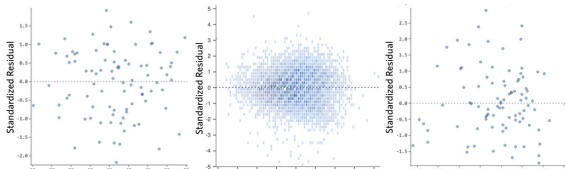


Poisson Mixed Effects Model: Residuals



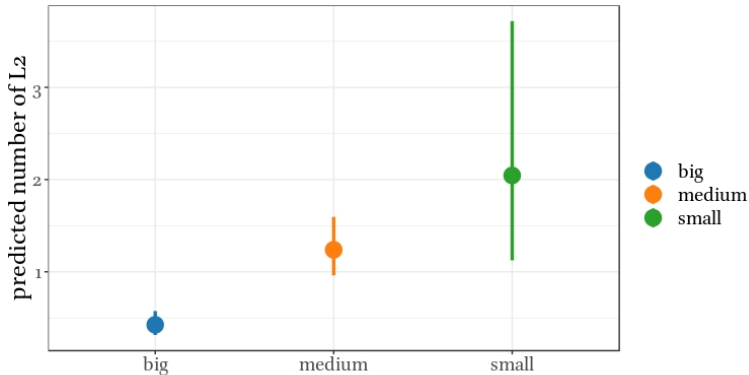
Statistical model is not ideal... Compare with some examples of “good” plots:

from <http://docs.statwing.com/interpreting-residual-plots-to-improve-your-regression/>



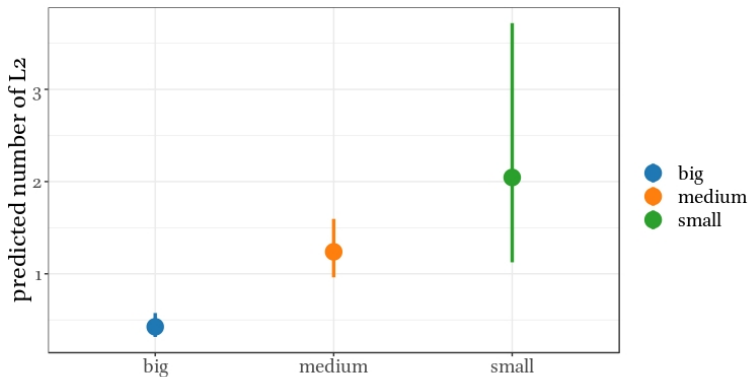
Conclusions:

- The variable language size is statistically significant.
- The obtained coefficients could be interpreted as follows:



Conclusions:

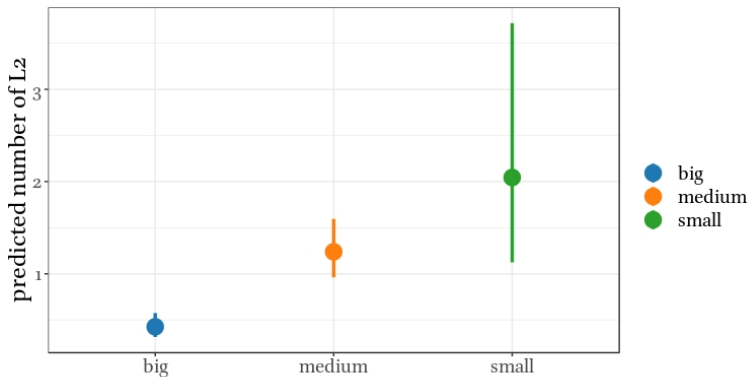
- The variable language size is statistically significant.
- The obtained coefficients could be interpreted as follows:



- Special case: Chuni

Conclusions:

- The variable language size is statistically significant.
- The obtained coefficients could be interpreted as follows:



- Special case: Chuni
- This is not only the case with Daghestanian languages:
 - Circassians in Arabic communities in Israel [Kreindler et al. 1995]
 - Any small community inside the larger one...

Send us a letter!
nina.dobrushina@gmail.com
agricolamz@gmail.com

Presentation is available here: tinyurl.com/y6jjp38y



All visualisation and statistical analysis were made in R version 3.5.3 [R Core Team 2019] with packages ggplot2 [Wickham 2016], lme4 [Bates et al. 2015],
lingtypology [Moroz 2017]

References

- Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48.
- Dobrushina, N. (2013). How to study multilingualism of the past: Investigating traditional contact situations in daghestan. *Journal of sociolinguistics* 17(3), 376–393.
- Dobrushina, N., D. Staferova, and A. Belokon (2017). Atlas of Multilingualism in Daghestan Online. <https://multidagestan.com/about>. Accessed: 2019-04-14.
- Kreindler, I., M. Bensoussan, E. Avinor, and C. Bram (1995). Circassian israelis: Multilingualism as a way of life. *Language, Culture and Curriculum* 8(2), 149–162.
- Moroz, G. (2017). *lingtypology: easy mapping for Linguistic Typology*.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Thomason, S. G. (2001). *Language contact*. Citeseer.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.