

# Lexical influence of Avar on minority languages of Daghestan

Samira Verhees

Linguistic Convergence Laboratory / NRU HSE Moscow

Summer School - Areal Linguistics and Languages of Russia - LCL - Voronovo, 2018

## Abstract

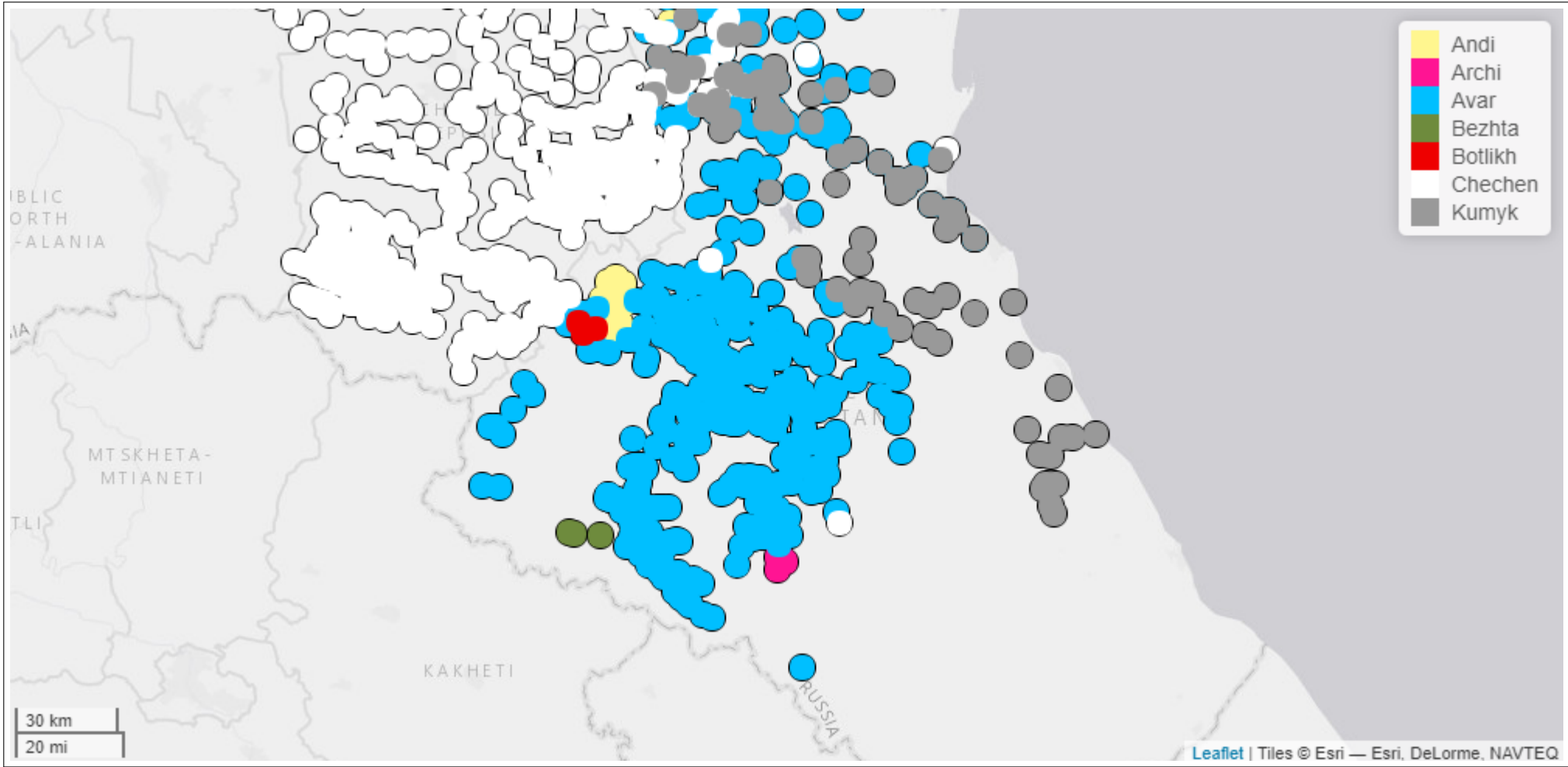
The aim of the Daghestanian Loans project is to investigate borrowing patterns among the languages of Daghestan (one of the North Caucasian republics of the Russian Federation). The focus is on borrowing between local languages, as opposed to the influence of major external languages like Arabic. This poster is about the influence of Avar (a major literary language of the region) on several minority languages for which it has traditionally served as language of literacy and (local) inter-ethnic communication. I compare the translations of a set of 118 meanings. The data show a mixture of genetic and areal clustering. Though the overall quantity of loans from Avar across different languages is similar, the actual words that are borrowed are different.

## 1 Introduction

Daghestanian Loans: a shortlist of borrowable concepts based on the World Loanword Database (WOLD) [Haspelmath and Tadmor 2009]. Minimizing borrowings from prestigious languages like Arabic or Russian. A total of 164 meanings / nouns, for efficient data collection in the field. This poster is based on mixed data - the intersection of various lists of translations resulted in 118 concepts (sometimes with multiple translations).

## 2 The Avar zone

Avar is a major literary language (about 1.000.000 speakers) of the East Caucasian / Nakh-Daghestanian language family. One of the 14 official languages of the Republic of Daghestan. Also spoken in the Balaken and Zaqatala regions of Azerbaijan and by a small community in Georgia (not shown on the map below). Has served as the language of literacy for speakers of Andic, Tsezic, Archi and Mehweb Dargwa and is a local lingua franca.



### 2.1 The sample

Data comes from four minority languages (seven lects in total) that have been influenced by Avar. A distant, divergent Avar dialect (Kusur, the lonely blue dot on the map above) was included for comparison.

Avar	
Avar	Standard (dictionary at <a href="http://avar.me/">http://avar.me/</a> ) Kusur (fieldwork)
Andic	
Andi	Andi (dictionary: Alisultanova [2009]) Gagatli (dictionary: Salimov [1968]) Rikvani (fieldwork) Zilo (fieldwork)
Botlikh	Botlikh (fieldwork)
Tsezic	
Bezhta	Bezhta (WOLD: Comrie and Khalilov [2009])
Lezgic	
Archi	Archi (WOLD: Chumakina [2009])

### 2.2 Possible borrowing sources

There is a functional distinction among contact languages in the area.

#### Vertical (prestigious languages, cultural borrowings)

- Arabic (religion, science, politics)
- Persian (various)
- Russian (technological innovation)

#### Horizontal (neighboring language, everyday life terms)

- Lak
- Georgian
- Chechen

#### Variable (everyday life terms and cultural borrowings)

- Avar
- Turkic (in this case mostly Kumyk of the Kipchak branch)

Mediated borrowing can be problematic.

## 3 Data annotation

Translations of one word are grouped by similarity. The full dataframe also shows the alleged source of a lexeme. In case a lexeme is likely to have entered the language through another language, this is indicated in a separate column ("intermediary"). (Data are available at Github: [https://github.com/sverhees/2018\\_Areal\\_AvarInfluence](https://github.com/sverhees/2018_Areal_AvarInfluence).)

English	Archi_word	Archi	Bezhta_word	Bezhta	Avar_word	Avar	Andi_word	Andi
the mirror	daxoni	1	c'ika	2	mat'u	3	mat'u	3

## 4 Loans and lexical similarity

### 4.1 Quantity of loans

The table below shows the number of loans from a specific donor language.

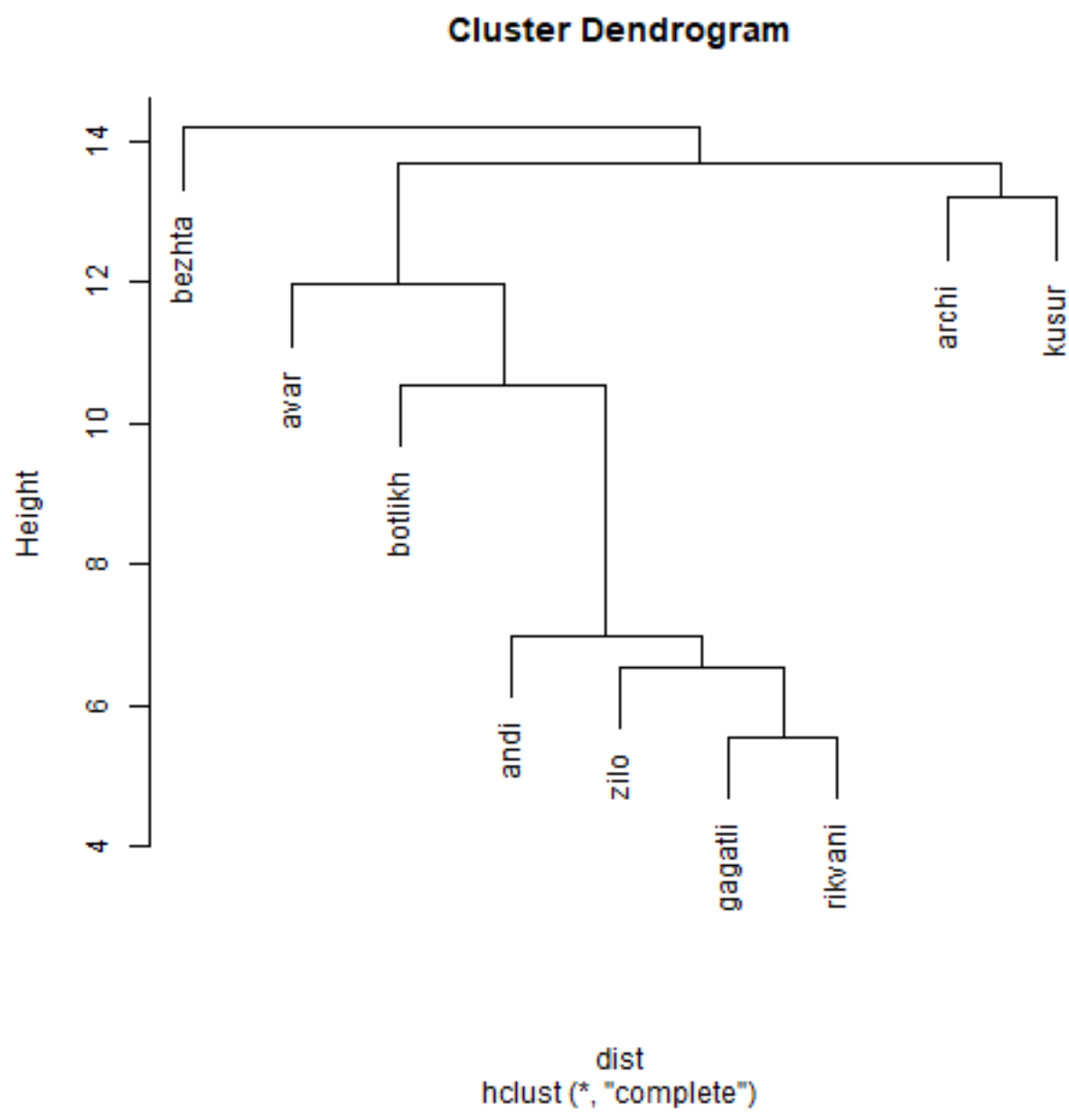
(Keep in mind that the list is not supposed to give an impression of which language has had more influence on the target languages overall, since we attempted to minimize the portion of borrowings from prestigious languages (i.e. vertical borrowing).)

	andi	gagatli	rikvani	zilo	botlikh	bezhta	archi
avar	19	18	18	15	27	17	12
chechen	1	1	0	2	0	0	0
georgian	0	0	0	0	0	16	0
lak	0	0	0	0	0	0	6
turkic	10	17	16	12	12	9	11
arabic	12	9	8	10	11	11	5
persian	12	13	11	10	9	12	14
russian	7	5	6	3	7	16	6
non loan	65	69	67	68	63	54	62
unidentified	2	2	1	2	4	2	2
total stems	128	134	127	122	133	137	118

There is quite a large portion of non loans in each language, but this concerns different meanings (there is only one shared non loan across all languages: 'tongs'). The same goes for Avar loans: the quantity is similar, but only the word *xalbac* 'lion' is shared by all.

### 4.2 Similarity: hierarchical clustering

The dendrogram below shows clustering based on the number of shared stems across languages.



In terms of shared lexicon, the Kusur dialect of Avar is about as similar to Standard Avar as a dialect of Andi. It patterns more like a language of the area where Lezgic languages are spoken, which is why it shares a branch with Archi.

## 5 Further research

- Even out the sample with representative one-to-one meaning-word correspondences
- Compare the intersection between lists and the role of different types of donors, e.g., do non loans have a significant effect on the similarity of Andic lects + Standard Avar, or is similarity caused by a shared set of vertical loans (which differs from the vertical loans set of languages from other areas).

The map was drawn with [Lingtypology](#) for R [Moroz 2017]. I would like to thank George Moroz for converting my table into a distance matrix and helping me make (some) sense of my data.

## References

- Alisultanova, M. (2009). *Lexika andijskogo jazyka [The lexicon of the Andi language]*. Makhachkala: Phd thesis.
- Chumakina, M. (2009). Archi. In M. Haspelmath and U. Tadmor (Eds.), *World Loanword Database*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Comrie, B. and M. Khalilov (2009). Bezhta. In M. Haspelmath and U. Tadmor (Eds.), *World Loanword Database*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Haspelmath, M. and U. Tadmor (Eds.) (2009). *WOLD*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Moroz, G. (2017). *lingtypology: Linguistic Typology and Mapping*. <https://CRAN.R-project.org/package=lingtypology>. <https://github.com/agricolanz/lingtypology/>.
- Salimov, K. S. (2010 (1968)). *Gagatlijskij govor andijskogo jazyka [The Gagatli dialect of Andi]*. Makhachkala: Institut yazyka, literatury i iskusstva im. G. Tsadasi.