Quantifying lexical contact in Daghestan

M. Daniel, I. Chechuro, S. Verhees

20.02.2018 Linguistic Convergence Lab Seminar

# Similarity matrices

**Similarity matrix SV style**

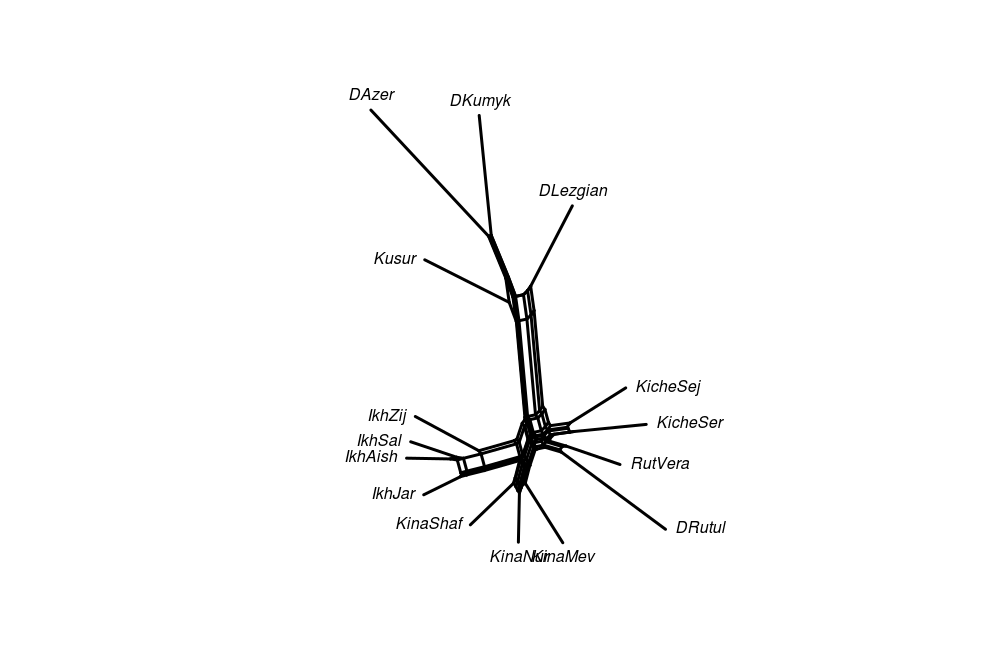
Each meaning can correspond to one or more words in a given language. Each distinct lexeme gets its own row; languages are marked 1 if the lexeme/root is represented, 0 if this is not the case. Similarity is measured by the number of times languages share a 1 on the same row.

**Similarity matrix MD style**

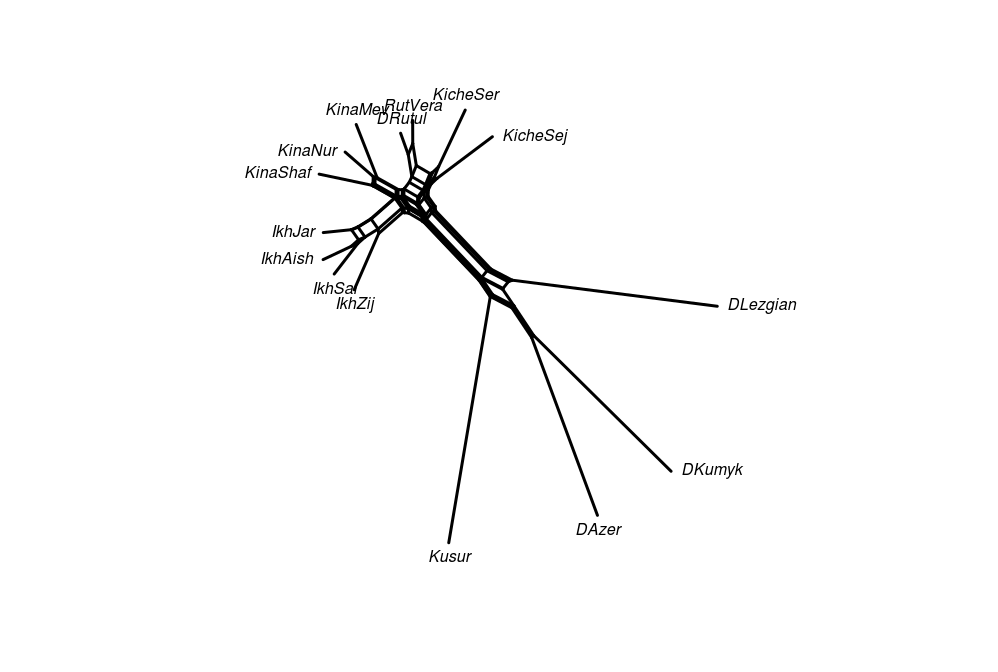
Each meaning can correspond to one or more words in a given language. Each distinct lexeme is assigned a number; a distance matrix is created from the number of times languages share the same lexeme (i.e. share 1 or more numbers on a row) / number of meanings.

# Visualizations

**Rutul area (SV style)**

1. 

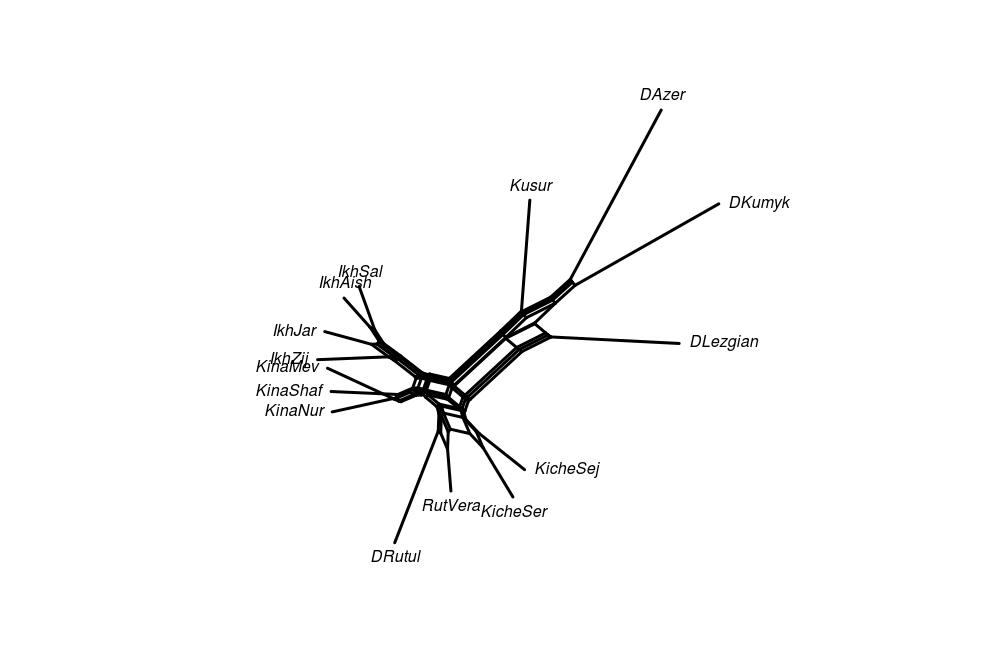
**Rutul area (MD style)**

1. 

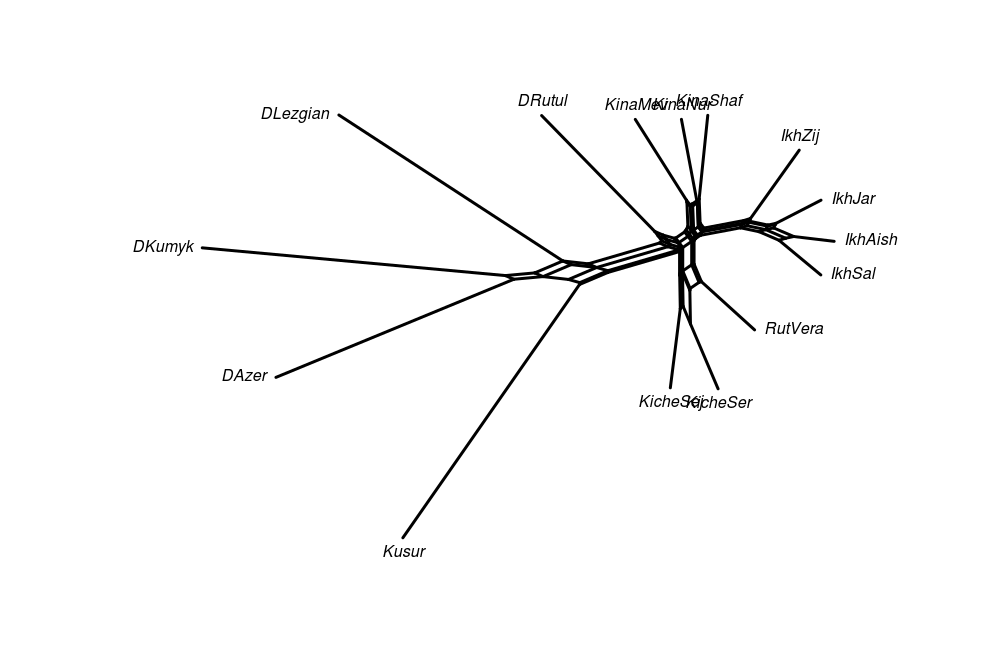
### Without “stop words”

If a meaning has a different lexeme in each language, remove it. The idea is to measure distance based on loans only.

**Rutul area (SV style)**

1. 

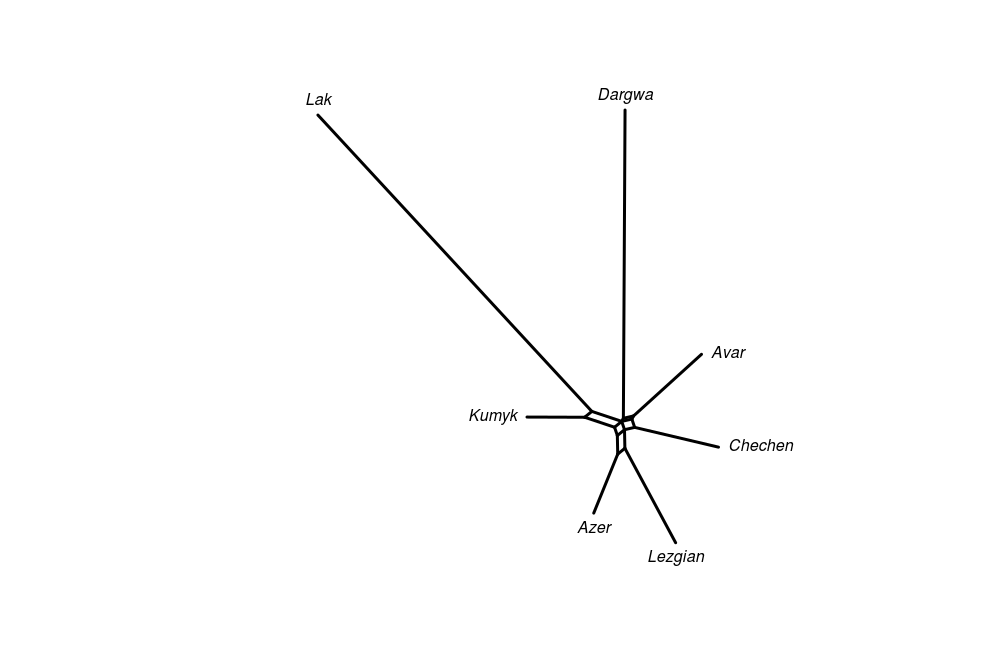
**Rutul area distance matrix (MD style)**

1. 

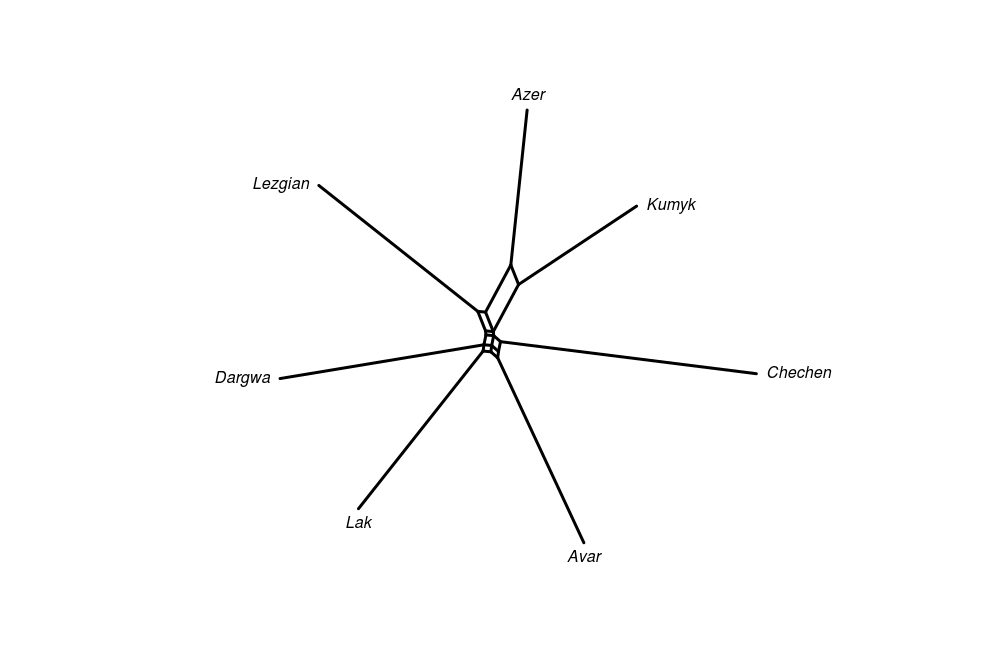
### Distance between major languages

Based on a larger subset of the WOLD questionnaire. Data gathered from dictionaries. Visualizations based on the same variables as the Rutul data.

**Major languages (SV style)**

1. 

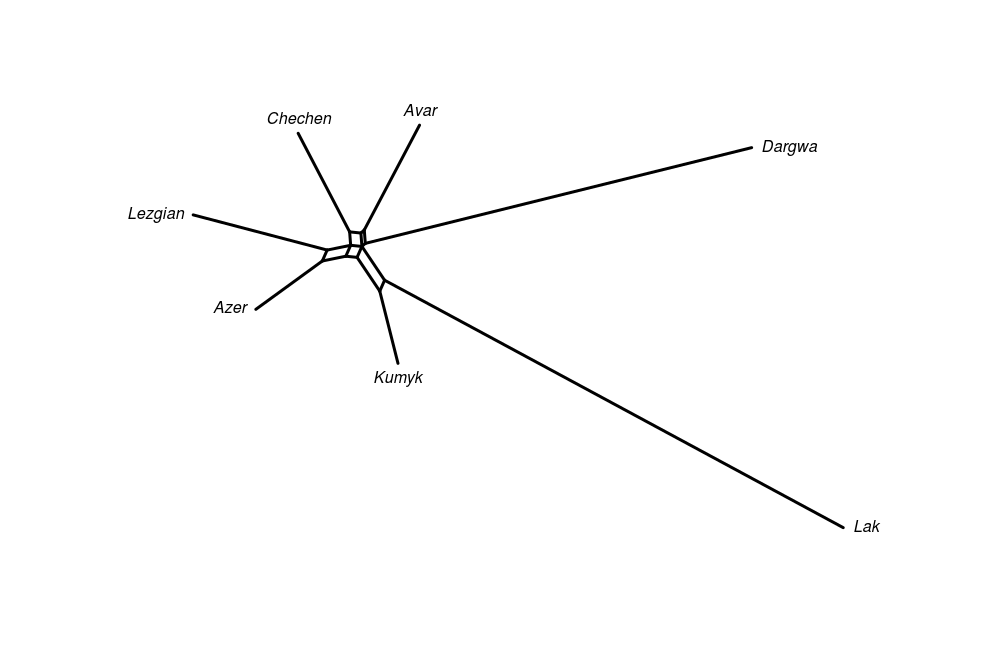
\*\* Major languages (MD style)\*\*

1. 

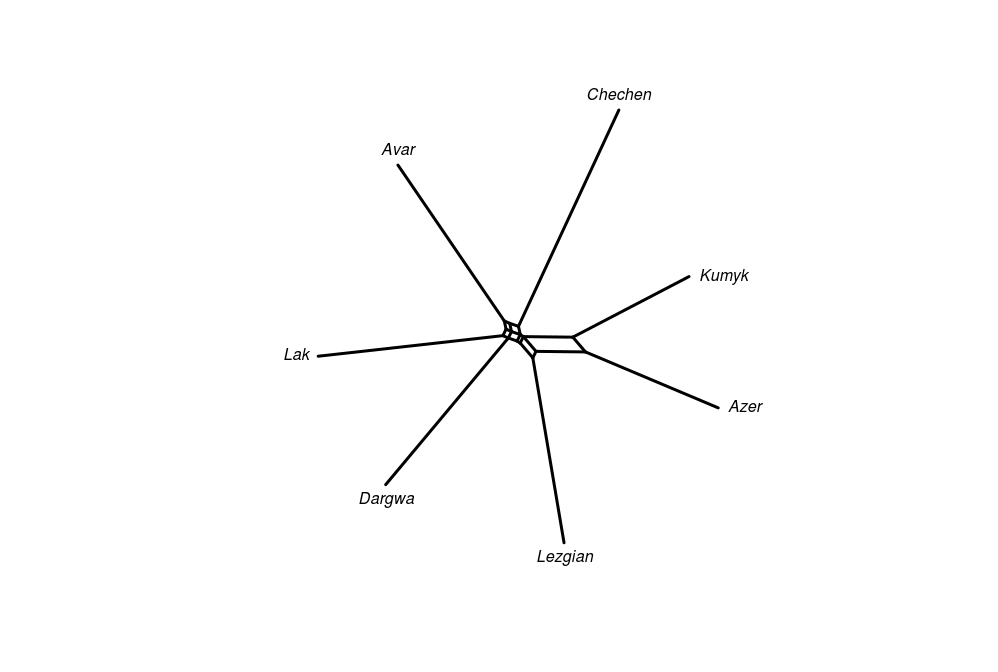
### Without “stop words”

If a meaning has a different lexeme in each language, remove it. The idea is to measure distance based on loans only.

**Major languages (SV style)**

1. 

**Major languages (MD style)**

1. 

## Avar area (MD style)

1. 