

Catching variation during fieldwork on Nakh-Daghestanian languages

George Moroz

Linguistic Convergence Laboratory, NRU HSE

21 June 2022

The Second International Conference ANATOLIA-THE CAUCASUS-IRAN:
dedicated to the 25th anniversary of “Iran and the Caucasus”



Introduction

Investigating variation:

- In variationism (e.g. (Labov 1963) on Martha's Vineyard /ai/ ~ /au/, (Trudgill 1974) on Norwich speech, (Wolfram 1969) on Afro-American speech from Detroit) researchers get mad about social stratification, mostly urban.

Investigating variation:

- In variationism (e.g. (Labov 1963) on Martha's Vineyard /ai/ ~ /au/, (Trudgill 1974) on Norwich speech, (Wolfram 1969) on Afro-American speech from Detroit) researchers get mad about social stratification, mostly urban.
- “Two equally interesting questions are at the heart of this book: how an extraordinary degree of idiosyncratic linguistic variation can coexist with an extraordinarily homogeneous speaker population, and how linguists might overlook the possibility of their coexistence.” (Dorian 2010: 3)

Investigating variation:

- In variationism (e.g. (Labov 1963) on Martha's Vineyard /ai/ ~ /au/, (Trudgill 1974) on Norwich speech, (Wolfram 1969) on Afro-American speech from Detroit) researchers get mad about social stratification, mostly urban.
- “Two equally interesting questions are at the heart of this book: how an extraordinary degree of idiosyncratic linguistic variation can coexist with an extraordinarily homogeneous speaker population, and how linguists might overlook the possibility of their coexistence.” (Dorian 2010: 3)
- In this talk we explore variation in a small, homogeneous speaker population and the probability that an average researcher of Nakh-Daghestanian languages catches this variation.

Data

Data were collected from

- 44 speakers of Andi (Nakh-Daghestanian) during fieldwork in Zilo (Botlikh district, Dagestan) in 2019



Created with [lingtypology](#) (Moroz 2017)

Data were collected from

- 44 speakers of Andi (Nakh-Daghestanian) during fieldwork in Zilo (Botlikh district, Dagestan) in 2019



Created with [lingtypology](#) (Moroz 2017)

- and 23 researchers of Nakh-Daghestanian languages via an online questionnaire

Zilo Data

44 Zilo speakers were asked to translate 16 stimuli:

- 1 'big butterfly'
- 2 'big butterflies'
- 3 'big grasshopper'
- 4 'big grasshoppers'
- 5 'the parents left'
- 6 'pour your father some water'
- 7 'pour your cow some water'
- 8 'third girl'
- 9 'the bride was beautiful at the wedding'
- 10 'heels'
- 11 'sons-in-law'
- 12 'they are eating'
- 13 'when he comes, we will eat'
- 14 'rainbow'
- 15 'north'
- 16 'thousand'

Stimuli: class attribution

lect	number	m	f	an	inan 1	inan 2	inan 3	non-h
Andi	sg	w	j	b	b	r	-	-
Andi	pl	w	j	j	b	r	-	-
Gagatli	sg	w	j	b	b	r	-	-
Gagatli	pl	w	j	j	b	r	-	-
Rikvani	sg	w	j	b	b	r	b	-
Rikvani	pl	w	j	j	b	r	r	-
Zilo	sg	w	j	b	b	r	-	-
Zilo	pl	w	j	j	b	r	-	-
Muni	sg	w	j	-	-	-	-	b
Muni	pl	w	j	-	-	-	-	b

- *odoruk'a* 'butterfly' (1, 2) and *kats'a* 'grasshopper' (3, 4) belong to the rare *inan 3* class in Rikvani (Suleymanov 1957)
- *odoruk'a* 'butterfly' (1) is one of the most variable words according to (Moroz and Verhees 2019)

- *ila-ima* ‘parents’ can agree with **m** or **inan 2** (5).

- *ila-ima* 'parents' can agree with *m* or *inan 2* (5).
- Different dative morphemes with masculine experiencers
-j/-tu (6, 7)

- *ila-ima* 'parents' can agree with *m* or *inan 2* (5).
- Different dative morphemes with masculine experiencers
-j/-tu (6, 7)
- Ordinal numeral marker sometimes agrees with noun class (8)

- *ila-ima* 'parents' can agree with *m* or *inan 2* (5).
- Different dative morphemes with masculine experiencers
-j/-tu (6, 7)
- Ordinal numeral marker sometimes agrees with noun class (8)
- Different plural forms for certain nouns (10, 11)

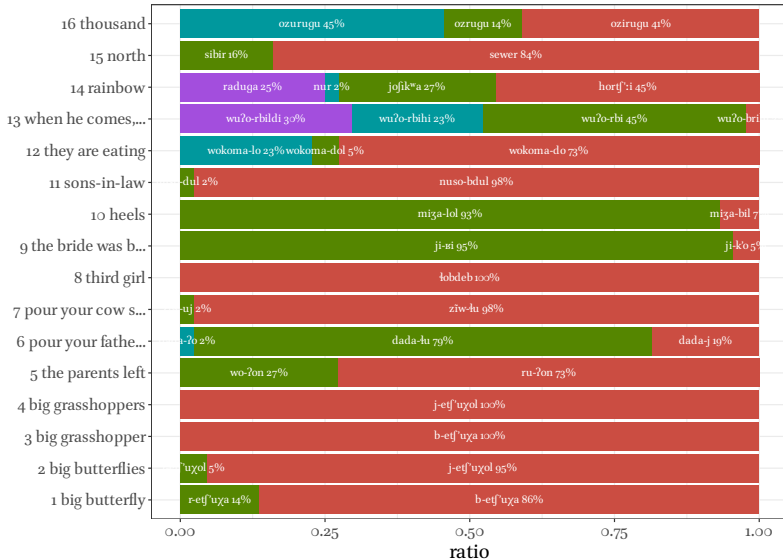
- *ila-ima* 'parents' can agree with *m* or *inan* 2 (5).
- Different dative morphemes with masculine experiencers
-j/-tu (6, 7)
- Ordinal numeral marker sometimes agrees with noun class (8)
- Different plural forms for certain nouns (10, 11)
- Two past copulas *b-ik'o* and *b-iki* with no functional distinction

- *ila-ima* ‘parents’ can agree with **m** or **inan 2** (5).
- Different dative morphemes with masculine experiencers
-j/-tu (6, 7)
- Ordinal numeral marker sometimes agrees with noun class (8)
- Different plural forms for certain nouns (10, 11)
- Two past copulas *b-ik’o* and *b-iki* with no functional distinction
- Temporal forms meaning “When X happened/happens/will happen” have a long (*-rbihi*) and a short form (*-rbi*), and can appear in ergative case with the same meaning (*-rbil-di*)

- *ila-ima* ‘parents’ can agree with *m* or *inan* 2 (5).
- Different dative morphemes with masculine experiencers
-j/-tu (6, 7)
- Ordinal numeral marker sometimes agrees with noun class (8)
- Different plural forms for certain nouns (10, 11)
- Two past copulas *b-ik’o* and *b-iki* with no functional distinction
- Temporal forms meaning “When X happened/happens/will happen” have a long (*-rbihi*) and a short form (*-rbi*), and can appear in ergative case with the same meaning (*-rbil-di*)
- Optional plural suffix *-l* can be added to progressive verb forms in *-rado/-mado* (12)

- *ila-ima* ‘parents’ can agree with *m* or *inan* 2 (5).
- Different dative morphemes with masculine experiencers
-j/-tu (6, 7)
- Ordinal numeral marker sometimes agrees with noun class (8)
- Different plural forms for certain nouns (10, 11)
- Two past copulas *b-ik’o* and *b-iki* with no functional distinction
- Temporal forms meaning “When X happened/happens/will happen” have a long (*-rbihi*) and a short form (*-rbi*), and can appear in ergative case with the same meaning (*-rbil-di*)
- Optional plural suffix *-l* can be added to progressive verb forms in *-rado/-mado* (12)
- ‘rainbow’, ‘north’, and ‘thousand’ (14, 15, and 16) show variation in word choice (14, 15) and adaptation (16): *ozrugu*, *ozirugu*, *ozurugu*

Zilo questionnaire (44 speakers): results



Information entropy

In order to measure how the count configuration c affects our sampling method, we use the information entropy, introduced in (Shannon 1948):

$$H(X) = - \sum_{i=1}^n P(x_i) \times \log_2 P(x_i)$$

Information entropy

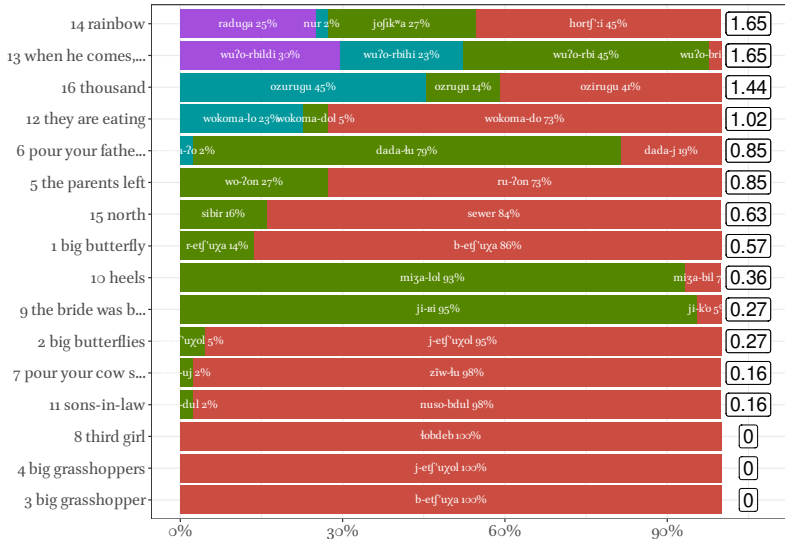
In order to measure how the count configuration c affects our sampling method, we use the information entropy, introduced in (Shannon 1948):

$$H(X) = - \sum_{i=1}^n P(x_i) \times \log_2 P(x_i)$$

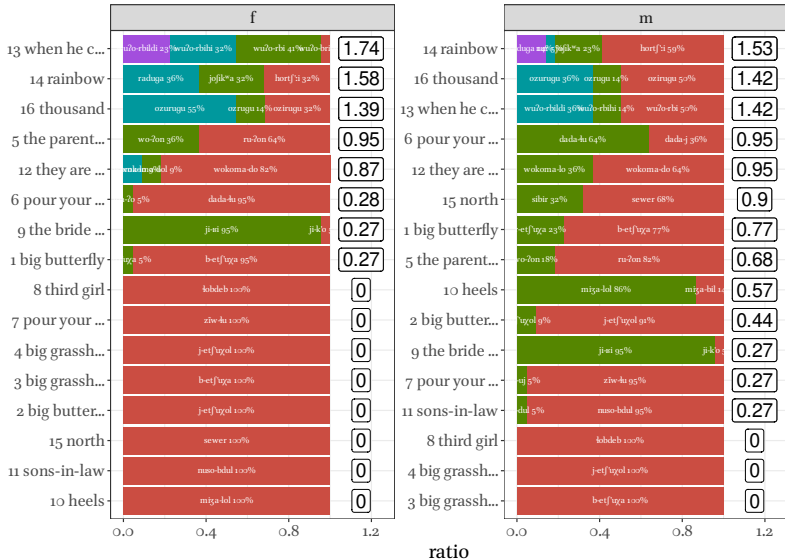
The range of the information entropy is $H(X) \in [0, +\infty]$:

data	entropy
A-A-A-A-A	0.00
A-A-A-A-B	0.72
A-A-A-B-B	0.97
A-A-B-B-B	0.97
A-A-B-B-C	1.52
A-B-C-A-B	1.52
A-B-C-D-E	2.32

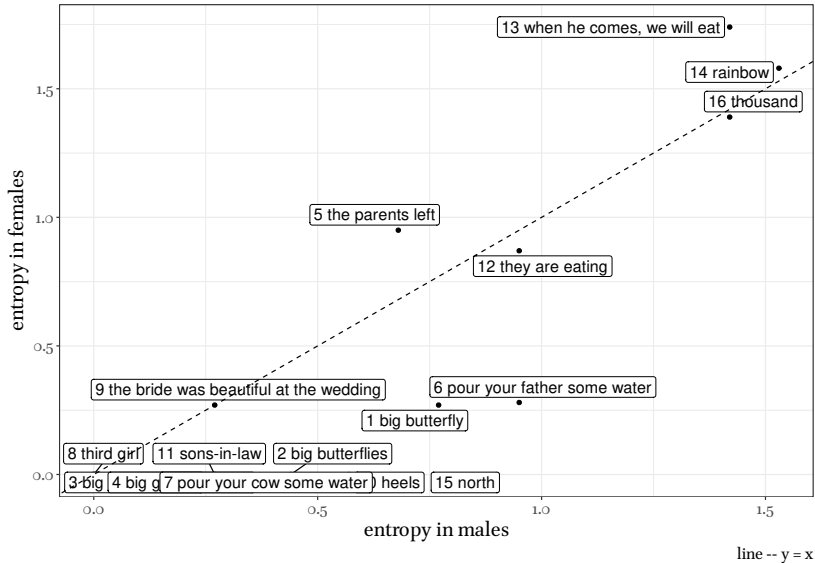
Zilo questionnaire (44 speakers): sorted by entropy



Zilo questionnaire (44 speakers): gender differences



Zilo questionnaire (44 speakers): gender differences



More about ‘butterfly’:

- previously we claimed:

lect	number	m	f	an	inan 1	inan 2	inan 3
Rikvani	sg	w	j	b	b	r	b
Rikvani	pl	w	j	j	b	r	r
Zilo	sg	w	j	b	b	r	-
Zilo	pl	w	j	j	b	r	-

- in (Moroz and Verhees 2019) we asked only the singular form of ‘butterfly’
- in our new experiment:

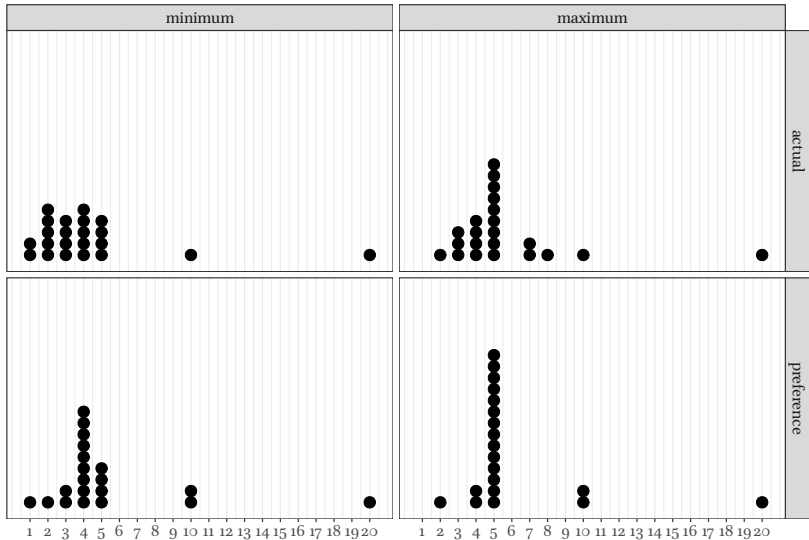
sg	pl	n	class
b-etʃʷuxa	j-etʃʷuxol	38	an
r-etʃʷuxa	j-etʃʷuxol	4	???
r-etʃʷuxa	r-etʃʷuxol	2	inan 2

Nakh-Daghestanian Fieldwork
Survey (thanks to Samira Verhees
for the help)

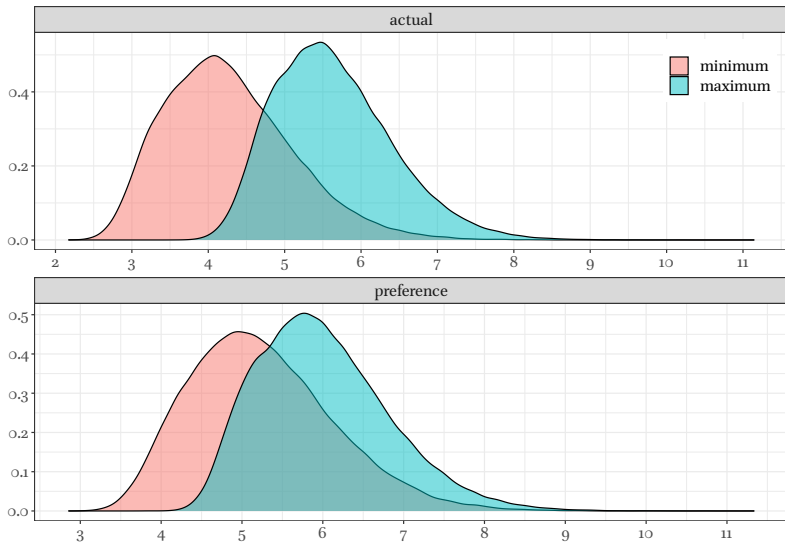
23 ND researchers were asked about:

- level of education
- linguistic interest
- studying linguistics at university
- fieldwork participation as a student
- year in which they finished their degree
- place of study and work
- number of people who participated in their fieldtrips
- preferred number of participants in fieldtrips
- goals of fieldwork
- use of elicitation and corpora
- **number of speakers a researcher *should* consult with**
- **number of speakers the researcher *usually* consults with**
- how researchers need to deal with interspeaker variability
- how researchers need to deal with intraspeaker variability
- whether speakers under the age of 13 are reliable consultants
- whether speakers older than 70 are reliable consultants
- personal (dis)preferences about the choice of consultants

Number of speakers

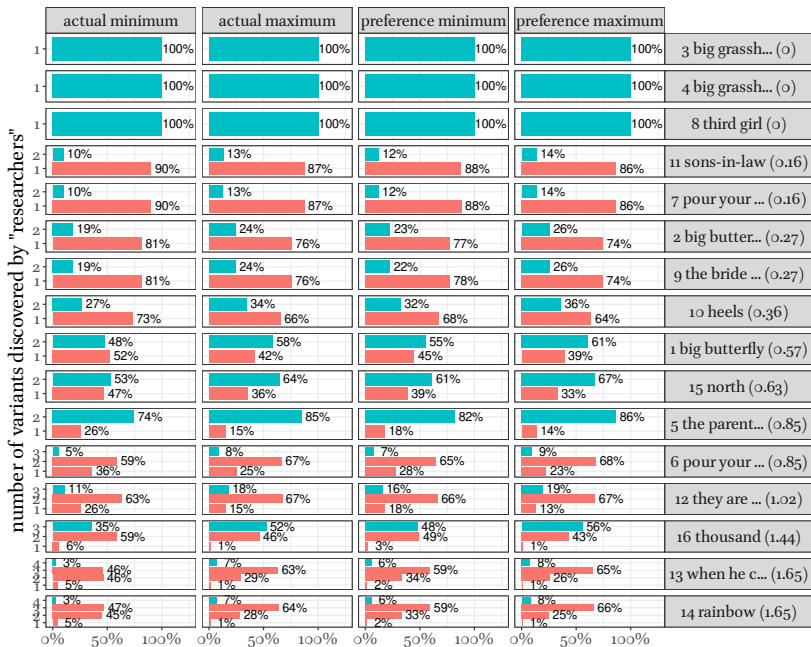


Bootstrapped mean number of speakers (10^5 iterations)



What if 10^5 “average researchers” ...
come to Zilo?

10⁵ samples from experiment results (entropy value)



Conclusions

Conclusions:

- Variation can be explained with the entropy measure
- Some questions have higher variability within female Zilo speakers and some — within male Zilo speakers
- An “**average researcher**” is a meaningful unit of meta analysis
- An “**average researcher**” might overlook a significant amount of the variation we observed due to the low number of speakers they usually consult with
- However, our experiment with 44 speakers also failed to show some of the variation we found in prior research on this dialect

References

- Dorian, N. C. (2010). *Investigating variation: The effects of social organization and social setting*. Oxford University Press.
- Labov, W. (1963). The social motivation of a sound change. *Word*, 19(3):273–309.
- Moroz, G. (2017). *lingtypology: easy mapping for Linguistic Typology*.
- Moroz, G. and Verhees, S. (2019). Variability in noun classes assignment in Zilo Andi: experimental data. *Iran and the Caucasus*, 23(3):268–282.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423.

References

- 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)*]. PhD thesis, Institut Jazykoznanja AN SSSR.
- Trudgill, P. (1974). *The social differentiation of English in Norwich*. Cambridge University Press.
- Wolfram, W. A. (1969). *A Sociolinguistic Description of Detroit Negro Speech., No. 5.*, volume 5 of *Urban language*. Center for Applied Linguistics.