Исследование вариативности в Зиловском андийском

Г. А. Мороз (с участием С. Ферхеес)

Международная лаборатория языковой конвергенции, НИУ ВШЭ

05–06 апреля 2022, НИУ ВШЭ, Нижний Новгород Корпусные технологии и компьютерные науки в гуманитарном знании (КонКорт-2022)



Исследование вариативности в зиловском диалекте андийского языка

Зиловские данные

Исследование нахско-дагестанских исследователей

Что если 10^5 "среднестатистических" исследователя … приедет в 3ило?

Вариативность

• "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]

Вариативность

- "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]
- Я сейчас представлю результаты анализа вариативность в моноэтничном селении Зило (андийский язык), а также покажу, как мы пробовали оценить, как "среднестатистический" исследователь получил бы похожие результаты.

Данные

Данные были собраны у:

 44 носителей андийского языка (нахско-дагестанская семья) во время полевого исследования (Ботлихский район, Дагестан) в 2019 году



4

Данные

Данные были собраны у:

 44 носителей андийского языка (нахско-дагестанская семья) во время полевого исследования (Ботлихский район, Дагестан) в 2019 году



 и 23 исследователей нахско-дагестанских языков при помощи онлайн опроса.



Исследование вариативности в зиловском диалекто андийского языка

Зиловские данные

Исследование нахско-дагестанских исследователей

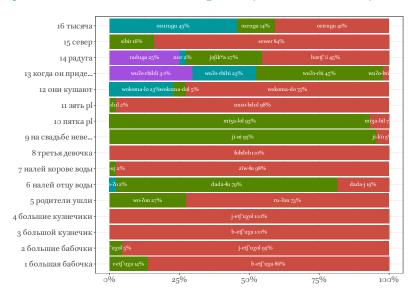
Что если 10^5 "среднестатистических" исследователя … приедет в 3ило?



44 носителей зиловского перевели следующие предложения:

- 1. 'большая бабочка'
- 2. 'большие бабочки'
- 3. 'большой кузнечик'
- 4. 'большие кузнечики'
- 5. 'родители ушли'
- 6. 'налей отцу воды'
- 7. 'налей своей корове воды'
- 8. 'третья девочка'
- 9. 'на свадьбе невеста была красивая'
- 10. 'пятки'
- 11. 'зятья'
- 12. 'они едят'
- 13. 'когда он придет, мы будем есть'
- 14. 'радуга'
- 15. 'север'
- 16. 'тысяча'

Результаты зиловского опроса (44 носителей)



Информационная энтропия

Чтобы измерить вариативность каждого вопросы, мы решили использовать информационную энтропию, введенную в [Shannon 1948]:

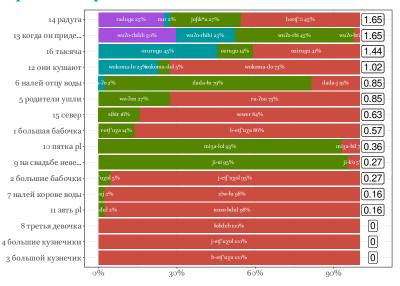
$$H(X) = -\sum_{i=1}^{n} P(x_i) \times \log_2 P(x_i)$$

Область значения энтропии $H(X) \in [0, +\infty]$:

данные	энтропия
A-A-A-A	0.00
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

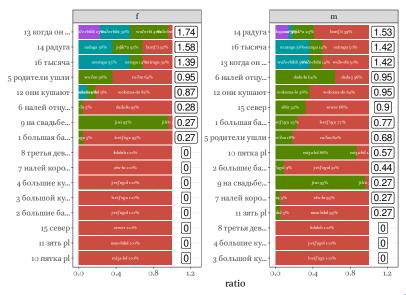


Зиловский опрос (44 носителей): значение энтропии справа

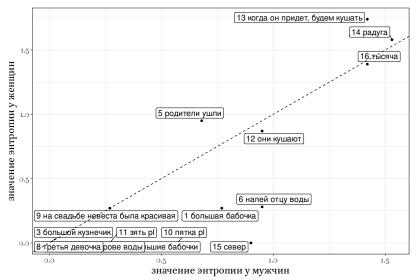




Зиловский опрос (44 носителей): гендерные различия



Зиловский опрос (44 носителей): значения энтропии в зависиимости от гендера



Исследование вариативности в зиловском диалекто андийского языка

Зиловские данные

Исследование нахско-дагестанских исследователей

Что если 10^5 "среднестатистических" исследователя … приедет в 3ило?



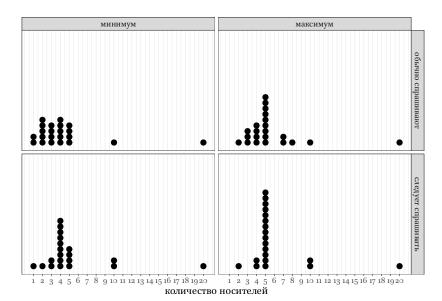
23 нахско-дагестанских исследователей заполнили следующую анкету:

- образование
- лингвистические интересы
- изучалась ли лингвистика в университете
- участие в полевой работе в качестве студента
- год получения степени
- место учебы/работы
- предпочтительное количество людей в полевой работе
- цели полевой работы
- количество носителей, которые, согласно мнению исследователя, *следует* опрашивать
- количество носителей, которые исследователь *обычно* опрашивает

• ...



Количество носителей





Бутстрэп

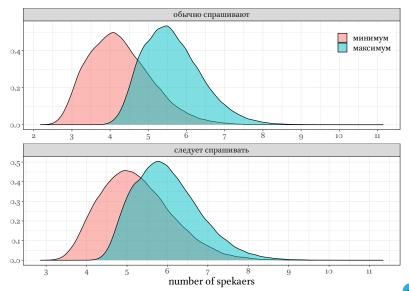


"To pull oneself over a fence by one's bootstraps".

Бутстрэп — это такой статистический подход, в рамках которого некоторый статистический параметр оценивается на основе большого количества выборок из имеющихся данных с повторением (т. е. каждое наблюдение может встретиться в выборке \circ раз, 1 раз, 2 раза и т. д.). В результате, вместо одной оценки параметра получается столько оценок, сколько у нас выборок, а все эти оценки формируют распределение.



Бустрэп среднего количества опрашиваемых носителей (10^5 iterations)



Исследование вариативности в зиловском диалекто андийского языка

Зиловские данные

Исследование нахско-дагестанских исследователей

Что если 10^5 "среднестатистических" исследователя … приедет в 3ило?



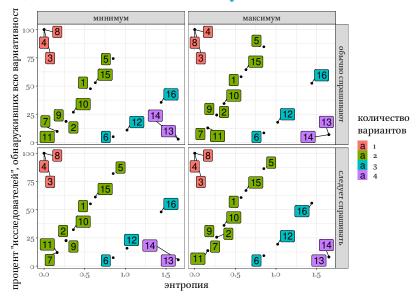
10^5 выборок из данных эксперимента

	ſ	actual minimum	actual maximum	preference minimum	preference maximum	
: 		48% 52%	58% 42%	55% 45%	61% 39%	1 большая бабочка
"исследователями"	2 { 1 {	19%	24%	23%	26%	2 большие бабочки
дова	1 {	100%	100%	100%	100%	3 большой кузнечик
SCTe,	ı {	100%	100%	100%	100%	4 большие кузнечики
Ĭ 2 ×	2 [26%	15%	18%	14%	5 родители ушли
обнаруженых	₫+	5% 59% 36%	8% 67%	28% 65%	9% 68%	6 налей отцу воды
Py z	2 -	10%	13%	12%	14%	7 налей корове воды
бна	1 {	100%	100%	100%	100%	8 третья девочка
OB),	2 -	19%	24%	22%	26% 74%	9 на свадьбе неве
OTBETOB,		27%	34%	32%	36%	10 пятка pl
	2 -	10%	13%	12%	14%	11 зять pl
вариантов	1 t	11% 63%	15% 67%	16%	19% 67%	12 они кушают
Bapi		3% 46%	7% 63%	6% 59% 2%	8% 65% 1% 26%	13 когда он приде
	1 (3% 47% 5%	7% 1% 28% 64%	6% 59%	8% 66%	14 радуга
количество		53% 47%	36%	61% 39%	33%	15 север
KOZI		35% 59%	152% 146%	48% 49%	1% 43%	16 тысяча
	0	% 50% 100%	0% 50% 100%	0% 50% 100%	0% 50% 100%	

10^5 выборок из данных эксперимента (энтропия)

		actual minimum	actual maximum	preference minimum	oreference maximum	
1	ı [100%	100%	100%	100%	3 большой ку (0)
"исследователями"	· {	100%	100%	100%	100%	4 большие ку (o)
това	· {	100%	100%	100%	100%	8 третья дев (0)
79IS	<u> </u>	10%	13%	12%	14%	11 зять pl (0.16)
X 1	1	10%	13%	12%	14%	7 налей коро (0.16)
бнаруженых	2 {	19%	24%	23%	26%	2 большие ба (0.27)
ж.d	1	19%	24%	22%	26%	9 на свадьбе (0.27)
2 2	} }	27%	34%	32%	36%	10 пятка pl (0.36)
0 g) 0	: {	48% 52%	58% 42%	55% 45%	61% 39%	1 большая ба (0.57)
OTBETOB,	} }	53% 47%	36%	61% 39%	33%	15 север (0.63)
O g 1	1	26%	85% 15%	18%	14%	5 родители ушли (0.85)
вариантов	1	5% 59% 36%	25% 67%	7% 28% 65%	23%	6 налей отцу (0.85)
Bapr	1	11% 63%	18% 15%	16% 66%	19% 13%	12 они кушают (1.02)
		35% 59%	152% 146%	48% 49%	156% 143%	16 тысяча (1.44)
количество		3% 46%	7% 63% 1% 29%	6% 59% 2%	8% 65%	13 когда он (1.65)
HON T	1	3% 47% 5% 45%	7% 64% 1% 28% 64%	6% 59% 59%	1% 25% 66%	14 радуга (1.65)

Когда "исследователи" найдут меньше?



Номер на графике соотносится с номером вопроса в анкете.



Исследование вариативности в зиловском диалекте андийского языка

Зиловские данные

Исследование нахско-дагестанских исследователей

Что если 10^5 "среднестатистических" исследователя … приедет в 3ило?



- вариативность можно описывать при помощи энтропии
- "среднестатистического" исследователя осмысленная единица метаанализа, которую следует дальше исследовать
- естественно: количество обнаруженной любыми исследователями зависит от энтропии вопроса



References

Dorian, N. C. (2010). *Investigating variation: The effects of social organization and social setting.* Oxford University Press.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423.