

Молодежный конкурс научных исследований НИВЦ  
2 декабря 2021, МГУ им. М.В. Ломоносова

# **Последовательны ли носители языка в своих суждениях? Измерение последовательности и ее параметризация**

Анастасия Герасимова

Аспирант 3го года кафедры теоретической  
и прикладной лингвистики филологического факультета  
Младший научный сотрудник Лаборатории  
автоматизированных лексикографических систем НИВЦ  
МГУ им. М.В. Ломоносова

Приоритетное направление:

*Современные компьютерные технологии в обучении*

Длительность исследования: 6 месяцев

# 1. Проблема исследования, ее актуальность

## Проблема разграничения «правильных» и «неправильных» языковых выражений

Идеальная языковая модель должна уметь порождать грамматически правильные предложения, но не допускать порождения неграмматичных структур.

— Отсутствие непосредственного доступа к языку: сведения о правильности языковых структур могут быть получены косвенным образом.

Корпуса текстов, размеченные по морфологическим и синтаксическим параметрам (НКРЯ, ГИКРЯ, CGEL 1985, Longman Grammar of Spoken and Written English 1999, и многие другие)

Корпуса, размеченные по приемлемости  
(Corpus of Linguistic Acceptability (CoLA) [Warstadt et al. 2019]).

Анализ текстов и оценок как единого целого.

Индивидуальные различия – шум.

## **1. Проблема исследования, ее актуальность**

### **Проблема разграничения «правильных» и «неправильных» языковых выражений**

Идеальная языковая модель должна уметь порождать грамматически правильные предложения, но не допускать порождения неграмматичных структур.

- Отсутствие непосредственного доступа к языку: сведения о правильности языковых структур могут быть получены косвенным образом.
- Лингвистические факторы, которые определяют приемлемость предложения: грамматическая правильность + факторы употребления.

## **1. Проблема исследования, ее актуальность**

### **Проблема разграничения «правильных» и «неправильных» языковых выражений**

Идеальная языковая модель должна уметь порождать грамматически правильные предложения, но не допускать порождения неграмматичных структур.

- Отсутствие непосредственного доступа к языку: сведения о правильности языковых структур могут быть получены косвенным образом.
- Лингвистические факторы, которые определяют приемлемость предложения: грамматическая правильность + факторы употребления.
- Проблема неоднородности суждений: варьирование как в оценках различных носителей, так и в оценках одного индивида.

# 1. Проблема исследования, ее актуальность

## Проблема разграничения «правильных» и «неправильных» языковых выражений

Идеальная языковая модель должна уметь порождать грамматически правильные предложения, но не допускать порождения неграмматичных структур.

- Отсутствие непосредственного доступа к языку: сведения о правильности языковых структур могут быть получены косвенным образом.
- Лингвистические факторы, которые определяют приемлемость предложения: грамматическая правильность + факторы употребления.
- Проблема неоднородности суждений: варьирование как в оценках различных носителей, так и в оценках одного индивида.



Можно ли доверять оценкам приемлемости и использовать их в качестве эмпирической базы для построения языковых моделей?

## 2. Задача

Оценить непротиворечивость языковых данных, получаемых от носителей языка.

Предположение: целостное представление о грамматике языка.

Три аспекта непротиворечивости:

1) *согласованность между носителями*: разные индивиды сходным образом реагируют на определенные условия

2) *индивидуальная последовательность*: индивид последователен в собственных реакциях на однотипные языковые выражения, поскольку обладает целостным представлением о грамматике родного языка

3) *индивидуальная устойчивость*: поведение индивида не должно отличаться при смене речевого модуса – при порождении и понимании

### 3. Материалы исследования

Серия экспериментальных исследований на материале трех конструкций, демонстрирующих вариативность в русском языке.

Анализ *согласованности* разных индивидов и *индивидуальной последовательности*:

(i) отрицательные конструкции с инфинитивными оборотами, содержащими отрицательное местоимение (*никто, ничто*) [Падучева 2013, Корнакова и др. 2016].

- (1) а. Но Людка **ничего** не хотела слушать. [НКРЯ]  
б. Мать Виолетты не хочет **ничего** делать... [НКРЯ]  
с. Я не хочу печалить Вас **ничем**. [НКРЯ]

### 3. Материалы исследования

Серия экспериментальных исследований на материале трех конструкций, демонстрирующих вариативность в русском языке.

Анализ *устойчивости* грамматики индивидов в двух модусах: при порождении и при понимании:

(ii) согласование с существительными, обозначающими человека по профессии и не имеющими женской родовой параллели (*директор, менеджер*) [Pesetsky 2013, Лютикова 2015].

- (2) а. *Это Цветкова — эта толстая врач — сказала!* [НКРЯ]  
б. ... *властный директор ответила с обезоруживающей женской непосредственностью...* [НКРЯ]

### 3. Материалы исследования

Серия экспериментальных исследований на материале трех конструкций, демонстрирующих вариативность в русском языке.

Анализ *устойчивости* грамматики индивидов в двух модусах: при порождении и при понимании:

(iii) падежное оформление прилагательного в конструкциях с малыми числительными (*два, три, четыре*) [Шкапа 2011, Pesetsky 2013]

- (3) а. *Как и положено, победителей ожидают три наградных номинации — бронзовая, серебряная и золотая.* [НКРЯ]  
б. *Все три опытные спортсменки потерпели фиаско на отборе.* [НКРЯ]

## 4. Серия лингвистических экспериментов

Эксперименты на понимание: оценить предложение по шкале от 1 до 7.

Эксперименты на порождение: дополнить предложения путем раскрытия скобок.

- (i) отрицательные конструкции с инфинитивными оборотами. → 128 человек
- (ii) рассогласование → 106 человек
- (iii) малые числительные → 106 человек

## 5. Анализ непротиворечивости

### 5.1. Согласованность ответов респондентов

(i) отрицательные конструкции с инфинитивными оборотами

Индивидуальные наборы рангов условий.

Сопоставление с эталонным набором (по всей выборке) при помощи коэффициента ранговой корреляции  $\tau$  Кендалла.

$\tau$	1	0.8165		0.5		0		-0.5		-0.8165		-1
Последовательность рангов	1,1,2	1,2,3	2,1,3	1,2,2	2,1,2	3,1,2	1,3,2	1,2,1	2,1,1	2,3,1	3,2,1	2,2,1
Число респондентов	14	29	15	7	4	16	13	6	7	6	7	4
%	10,9	22,7	11,7	5,5	3,1	12,5	10,2	4,7	5,5	4,7	5,5	3,1
$\Sigma$ %		45,3		8,6		22,7		23,4				

- 45% респондентов имеют коэффициент ранговой корреляции  $\tau = 1$  и  $\tau = 0.8$ , что задает основную тенденцию выборки
- 23% респондентов имеют отрицательное значение коэффициента

## 5. Анализ непротиворечивости

### 5.2. Индивидуальная последовательность

(i) отрицательные конструкции с инфинитивными оборотами

Индивидуальная последовательность ответов респондентов оценивалась через среднее стандартное отклонение в ответах для каждого условия.

Коэффициент $\tau$	Среднее стандартное отклонение		
	< 1	$\geq 1$ & < 2	$\geq 2$
-0.8165 & -0.5	2%	17%	4%
0 & 0.5	2%	24%	5%
0.8165 & 1	5%	35%	6%
<b><math>\Sigma</math> %</b>	<b>9%</b>	<b>76%</b>	<b>15%</b>

- Большая часть респондентов (76%) демонстрируют средний уровень индивидуальной последовательности (варьирование от 1 до 2 баллов по шкале 1-7).
- Строго последовательны в своих суждениях меньше 10% респондентов (варьирование в 1 балл).

## 5. Анализ непротиворечивости

### 5.3. Устойчивость при порождении и понимании

(ii) рассогласование & (iii) малые числительные

а) какой вариант респондент использовал в первом эксперименте;

б) какой из вариантов респондент в среднем оценил как более приемлемый.

	Эксперимент 1 порождение		Эксперимент 2 оценка	
Условие А	NOM	Условие А	NOM	5
			GEN	2
Условие Б	GEN	Условие Б	NOM	2
			GEN	5

## 5. Анализ непротиворечивости

### 5.3. Устойчивость при порождении и понимании

(ii) рассогласование & (iii) малые числительные

а) какой вариант респондент использовал в первом эксперименте;

б) какой из вариантов респондент в среднем оценил как более приемлемый.

	Эксперимент 1 порождение		Эксперимент 2 оценка
Условие А	NOM	Условие А	NOM 5 GEN 2
Условие Б	GEN	Условие Б	NOM 2 GEN 5

## 5. Анализ непротиворечивости

### 5.3. Устойчивость при порождении и понимании

(ii) рассогласование & (iii) малые числительные

	Согласование по роду (ii)	Малые числительные (iii)
<b>Стратегия А.</b> Вариант, используемый при порождении, оценивается как наиболее приемлемый	57%	39%
<b>Стратегия Б.</b> В одном эксперименте используются/оцениваются как приемлемые обе альтернативы, в другом эксперименте предпочтение отдается какому-то одному варианту	30%	37%
<i>Оба варианта использованы при порождении</i>	14%	23%
<i>Оба варианта равно приемлемы по оценкам</i>	16%	14%
<b>Стратегия В.</b> В двух экспериментах предпочитают различные варианты.	13%	24%

Респонденты последовательны при переходе между модальностями в среднем в половине случаев; при этом последовательность различается для разных типов феноменов.

## 6. Результаты

Установлены три аспекта непротиворечивости языковых данных:

- согласованность между носителями
- индивидуальная последовательность
- (индивидуальная) устойчивость

Предложены способы измерения и интерпретации неоднородности.

Многовариантность языка:

Несмотря на то, что большая часть ответов индивидов формирует общую тенденцию выборки, мы наблюдаем отклонения от общей тенденции.

Варьирование неслучайно и происходит в ограниченном диапазоне.

(Не)последовательность респондентов в разных модальностях

→ разные стадии развития одного из вариантов.

## **7. Научная значимость и новизна**

При построении компьютерных моделей языка имплицитно предполагается единство и стабильность грамматики.

В действительности → не одна идеальная грамматика, а множество идиолектов, которые складываются в отдельные языковые профили носителей языка.

Компьютерные модели предсказывают грамматическую правильность на основании частотности [Lau et al. 2017, Sprouse et al. 2018].

Непротиворечивость собственно языковых данных – вне поля зрения.

## **8. Практическая значимость**

Важная задача современной лингвистики – учитывать это внутриграмматическое варьирование и задавать такие параметры языковой модели, которые могли бы его допускать.

## 9. Перспективы продолжения исследования

1. Классификация и кластеризация вариативных феноменов с точки зрения того, на каком основании происходит варьирование.
2. Еще один аспект непротиворечивости: динамика изменения поведения респондента в процессе восприятия языкового материала.



Учет вариативности необходим для развития научных представлений о языке.

## Библиография

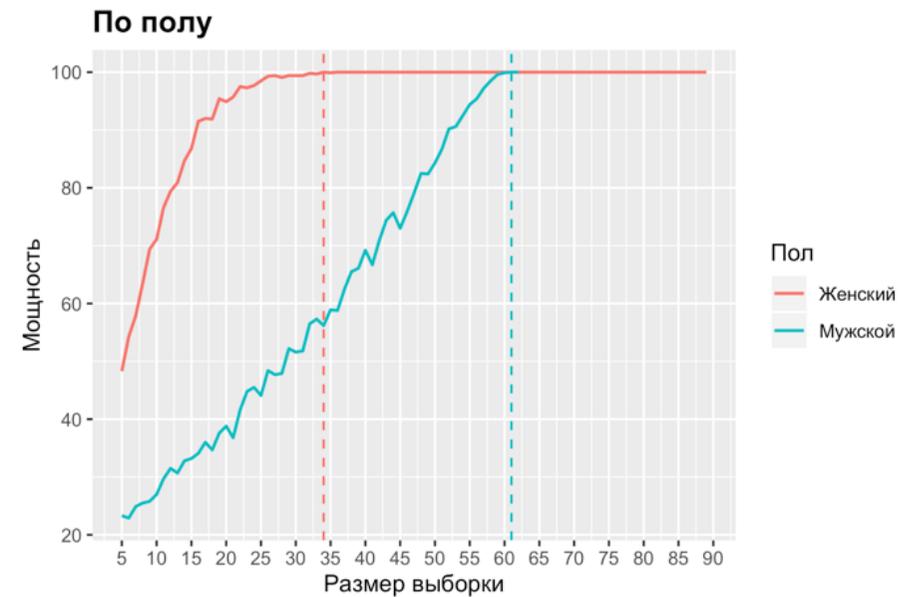
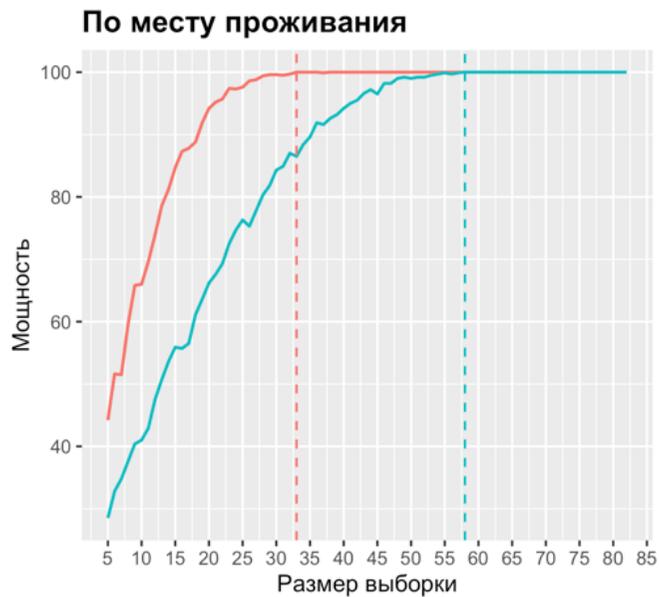
- Корнакова Е. В., Лютикова Е. А., Гращенков П. В. Лицензирующие свойства отрицания в русском языке // Rhema. Рема. 2016. № 4. С. 58-82.
- Лютикова Е.А. Согласование, признаки и структура именной группы в русском языке // Русский язык в научном освещении. 2015. № 30. С. 44-74.
- Падучева Е.В. Русское отрицательное предложение. М.: ЯСК. 2013.
- Шкапа М.В. Согласование определения с существительным при числительных два, три, четыре // Проблемы русской стилистики по данным НКРЯ. 2011. [URL: [https://studiorum.ruscorpora.ru/stylistics/syntax\\_numeral/](https://studiorum.ruscorpora.ru/stylistics/syntax_numeral/)]
- Lau, J. H., Clark, A., Lappin, S. Grammaticality, Acceptability, and Probability: A probabilistic view of linguistic knowledge. // Cognitive Science. 2017. 41:5, 1201-1241.
- Pesetsky D. Russian case morphology and the syntactic categories. Cambridge, MA: MIT Press, 2013.
- Sprouse J., Yankama B., Indurkha S., Fong S., Berwick R.C. Colorless green ideas do sleep furiously: gradient acceptability and the nature of the grammar // GLOW Issue. The Linguistic Review. 2018. Vol. 35. № 3. P. 575–599.
- Warstadt A., Singh A., Bowman S. R. Neural network acceptability judgments // Transactions of the Association for Computational Linguistics. 2019. Vol. 7. P. 625-641.

## 10. Приложение

### Согласованность социолингвистических групп

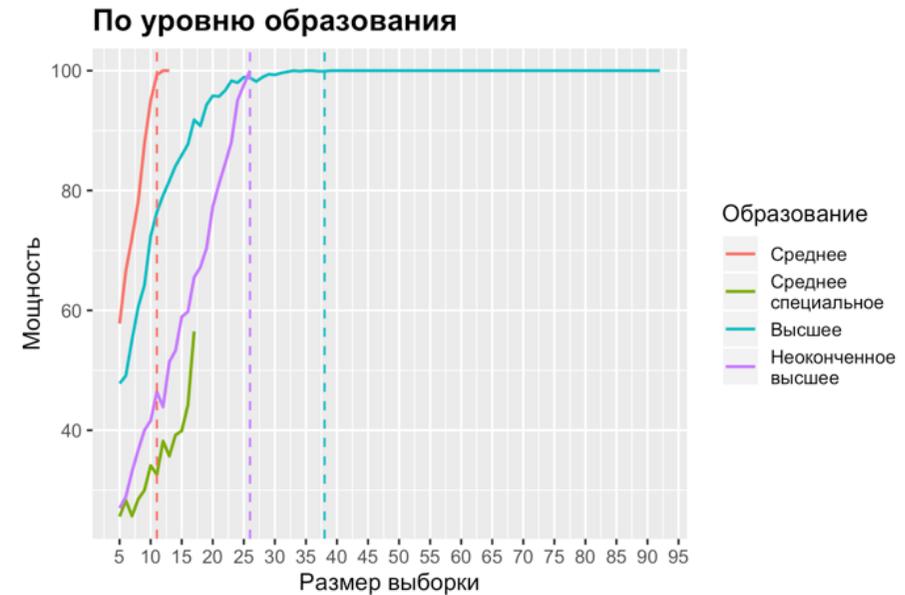
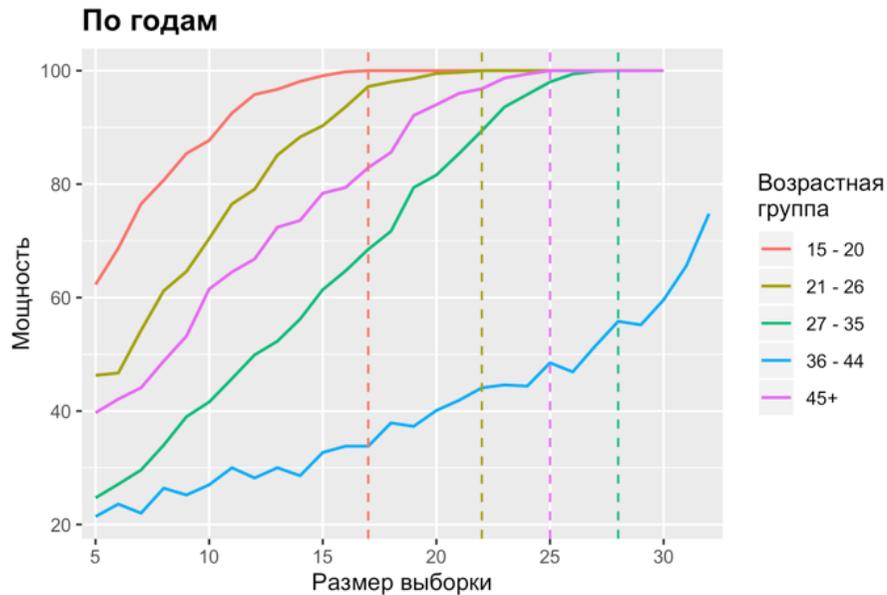
Скорость достижения максимальной мощности для подгрупп респондентов, выбранных по социолингвистическим параметрам.

Вероятность обнаружить различие между условиями в рамках конкретного эксперимента.



# 10. Приложение

## Согласованность социолингвистических групп



- Существенные отличия наблюдаются для различных социолингвистических групп.
- При наименьшем объеме выборки максимальная мощность достигается для респондентов в возрасте от 15 до 20 лет.