



Предсказуемость в психолингвистике и NLP

**Предсказуемость слова** — в психолингвистике это вероятность правильного угадывания следующего слова в тексте или предложении на основе предыдущего контекста. В рамках эксперимента участники видят **только левый контекст от целевого слова**. Доказано, что вне зависимости от исследуемого языка, **предсказуемость, длина и частотность оказывают наибольшее влияние на то, насколько легко слово будет обрабатываться при чтении** (Ashby et al., 2005; Ehrlich & Rayner, 1981 и др.).

- 1) С ростом предсказуемости уменьшается вероятность возврата взгляда к предыдущему слову.
- 2) Вместе с длиной слова предсказуемость влияет на вероятность фиксации взгляда на целевом слове: чем выше предсказуемость слова и чем оно короче, тем чаще оно будет пропускаться при чтении (Laurinavichyute, A.K., Sekerina, I.A., Alexeeva, S., 2019).

Считается, что **чем лучше нейросеть предсказывает следующее слово, тем лучше это отражает человеческое поведение при чтении** (Wilcox et al., 2020; Merx and Frank, 2020).  
**Результаты Transformers** на данный момент в **наибольшей степени отражают поведение людей**. Результаты моделей, которые были обучены на **2 миллиардах токенах и более, лучше соответствуют поведению человека**, чем аналогичные нейросетевые модели, но с меньшим количеством параметром (Oh & Schuler, Findings 2023).

TEST YOURSELF

- 1) Клиенты воровали из ресторана \_\_\_\_ атрибутуку
- 2) Приблизительно в центре тайги \_\_\_\_ снегоход
- 3) Елена сидела в кресле, молодая Мурка \_\_\_\_ урчала
- 4) Врач прописал заживляющую \_\_\_\_ мазь
- 5) Выбирая вязаную шапочку, знайте, что лучше шапка цвета \_\_\_\_ хаки

Данные для исследования

**144 предложения из Russian Sentence Corpus** <https://www.hse.ru/en/neuroling/research/RSC>

**144 стимула:** 79 существительных, 33 глагола и 32 прилагательных. Все стимулы сбалансированы по длине и и частотности.

Используемые модели

Эксперимент 1: **ruGPT-3.5, llama-3.1, Qwen/QwQ**.

Архитектура трансформеров; механизма внимания, от 8 миллиардов параметров

Методы сэмплирования: **top\_k, top\_p, temperature, num\_beams**.

Подавался только левый контекст, брались первые несколько токенов из предсказания

Экспермент 2: **RuModernBERT-small** (35 миллионов параметров) и **RuModernBERT-base** (105 миллионов параметров)

Двунаправленная архитектура, специально обучены на “маскированном языке”.

«Видели» как к левый, так и правый контекст, что значительно улучшает точность и качество предсказаний.

Гипотезы:

- Гипотеза 1. У людей будет наблюдаться больше ошибок, связанных с несоответствием предсказанной части речи с ожидаемой. (Модели будут давать более прямолинейные и менее разнообразные варианты. Ожидается, что различные части речи обладают разной степенью предсказуемости).
- Гипотеза 2. Длина слова и его частотность, будут влиять на точность предсказаний и людей, и моделей, но это влияние может быть выражено в разной степени.
- Гипотеза 3. Ошибки, связанные с неверной частеречной принадлежностью предсказаний, могут рассматриваться как более значимые с точки зрения общей сложности предложения.
- Гипотеза 4. Чем больше контекста будет видеть и модели, и люди, тем выше качество ответов будет даваться.
- Гипотеза 5: Модели хуже улавливают синтаксические признаки (например, совершая больше ошибок в управлении)
- Гипотеза 6. BERT покажет более высокие результаты по сравнению и с моделями архитектуры Transformers, и с людьми, засчет использования двунаправленной обработки контекста

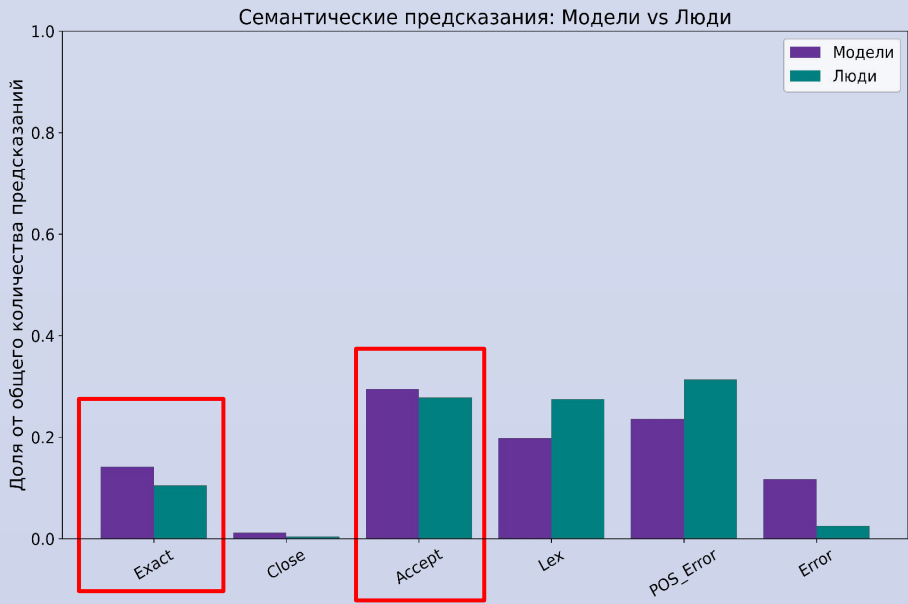
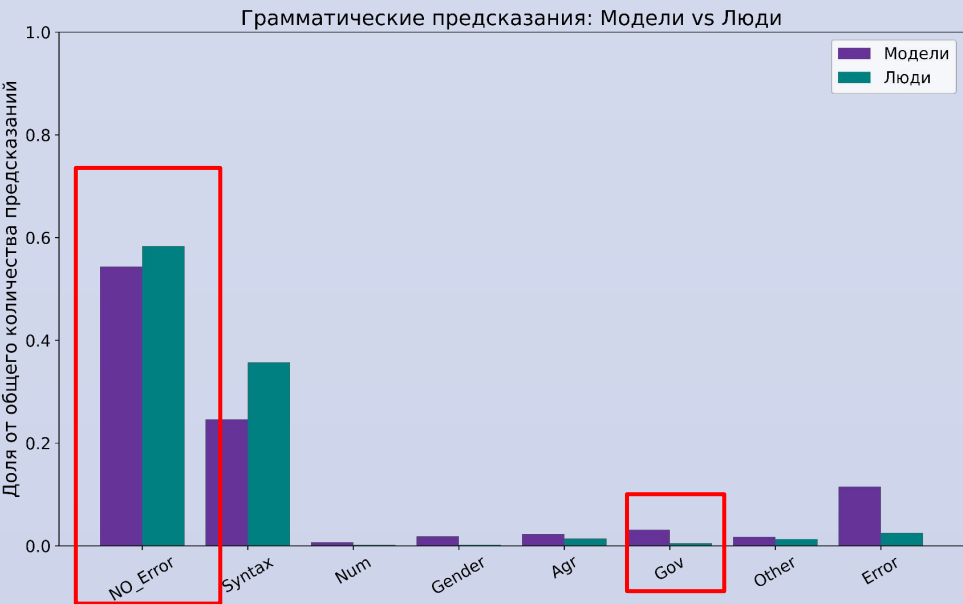
Избранная литература:

Laurinavichyute, A.K., Sekerina, I.A., Alexeeva, S. *et al.* **Russian Sentence Corpus: Benchmark measures of eye movements in reading in Russian.** *Behav Res* 51, 2019. 1161–1178. URL: <https://doi.org/10.3758/s13428-018-1051-6>

Oh, B.-D., Schuler, W. **Transformer-Based Language Model Surprisal Predicts Human Reading Times Best with About Two Billion Training Tokens** // *Findings of the Association for Computational Linguistics: EMNLP 2023*. Singapore: Association for Computational Linguistics, 2023. C. 1915–1921. URL: <https://aclanthology.org/2023.findings-emnlp.128/>

Результаты модели vs люди

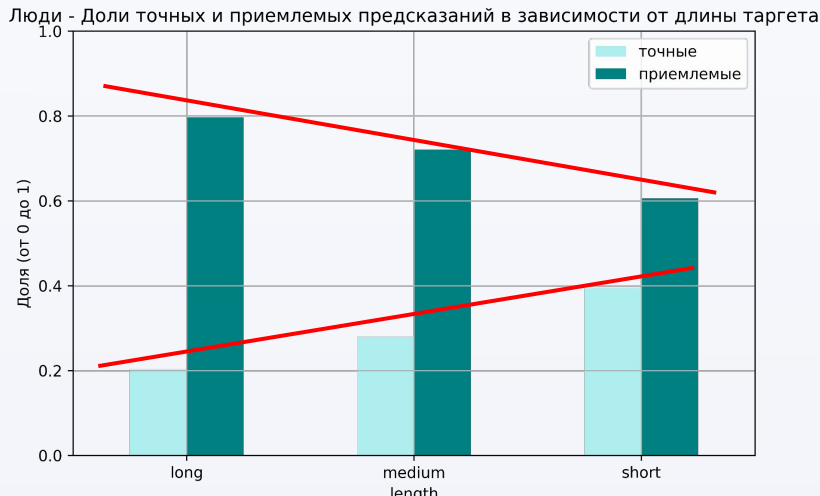
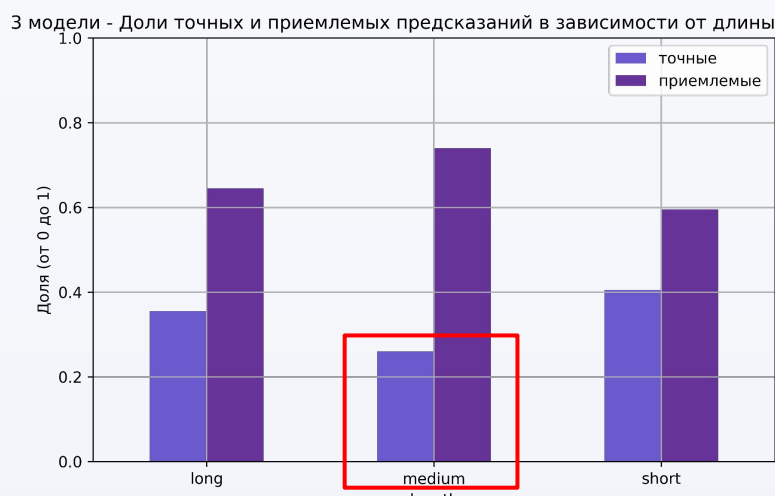
Все предсказания были размечены по **семантическим / грамматическим признакам**



	Модели			Люди		
	correct POS	correct sem	correct gram	correct POS	correct sem	correct gram
noun	<b>79,84</b>	55,93	<b>68,38</b>	<b>71,19</b>	43,05	<b>69,92</b>
verb	<b>63,32</b>	38,69	55,28	79,24	36,73	54,07
adj	42,02	27,66	30,85	<b>26,27</b>	18,54	18,54

Её сын Гриша умер **<бездетным>**, и младший сын остался единственным наследником.  
Предсказания моделей: **рано, недавно, сегодня**  
Предсказания людей: **от** (болезни), **месяц** (назад), **прошлой** (весной/летом/зимой)

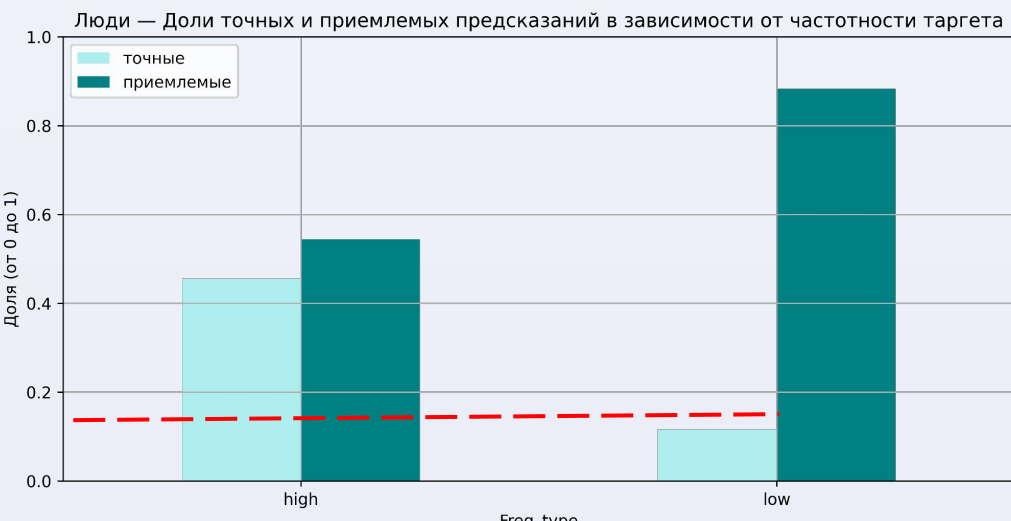
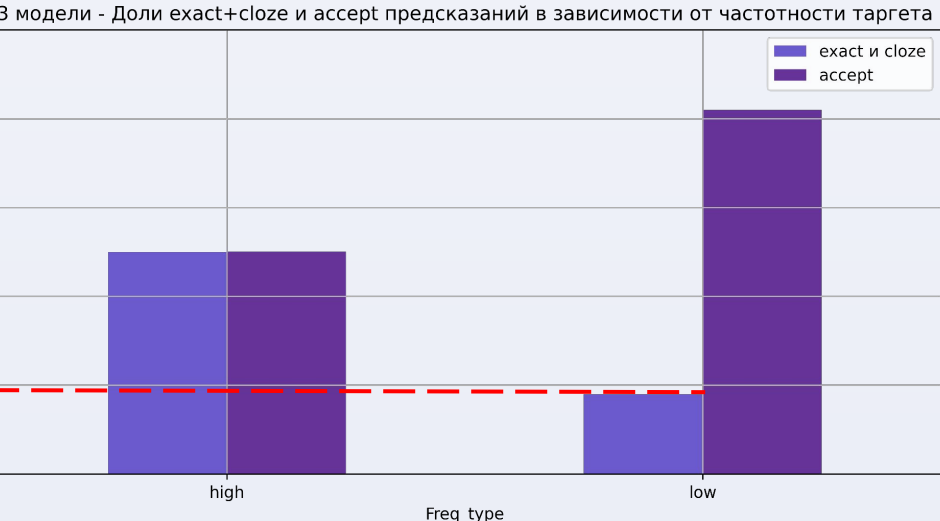
Влияние длины таргета



**Гипотеза 2 подтвердилась частично.**

Явной **корреляции между длиной таргетного слова и успешностью предсказания нет**. Длина таргетного слова влияет только на **точность предсказания**. Люди более чувствительны к этому параметру, чем модели.

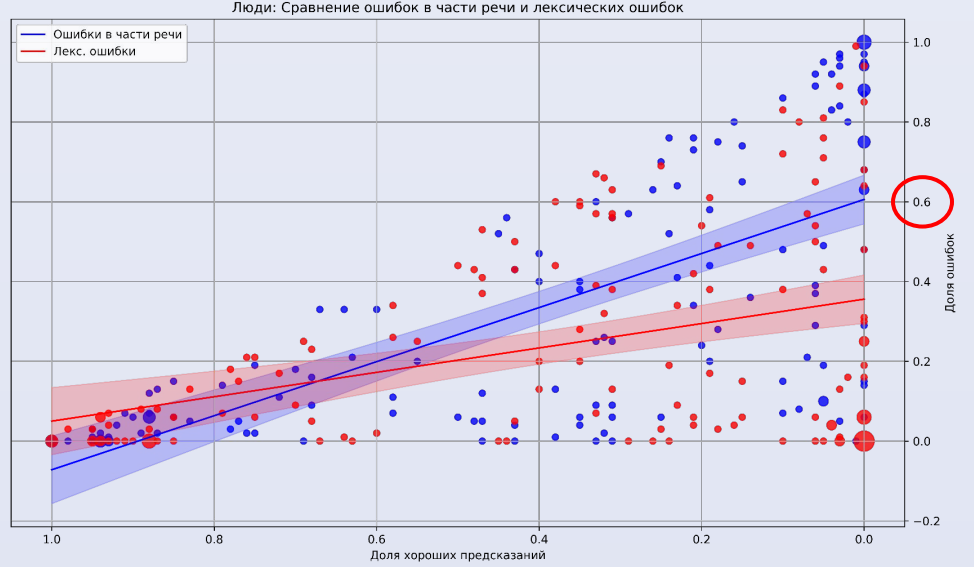
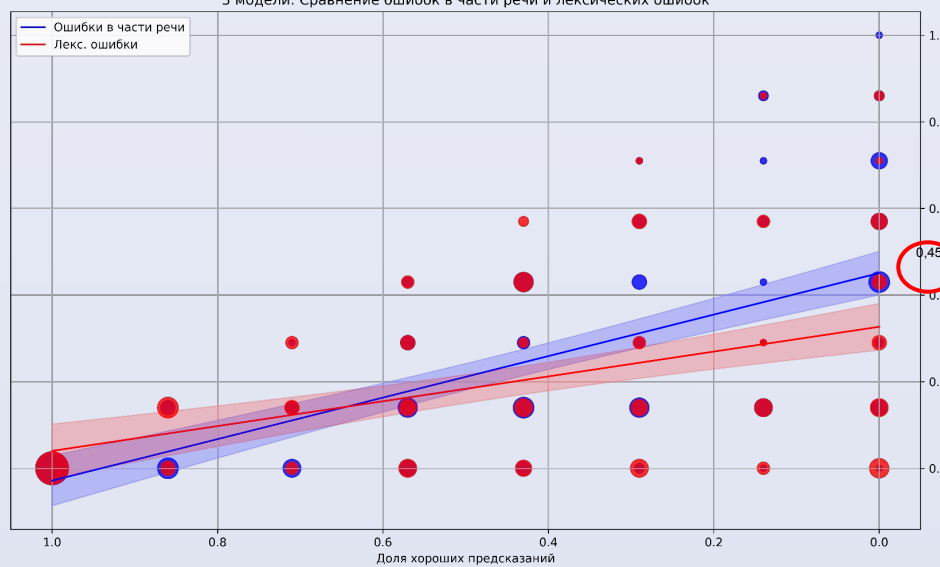
Влияние частотности таргета



**Гипотеза 2 подтвердилась частично.**

**Частотность не влияет на семантическую приемлемость предсказаний**, также **нет линейной зависимости семантической корректности предсказаний от частотность биграмм, образуемых таргетным словом и предыдущим**. **Есть небольшая положительная зависимость между точностью предсказаний и частотностью целевого слова**. Люди более чувствительны к увеличению частотности слова, чем модели.

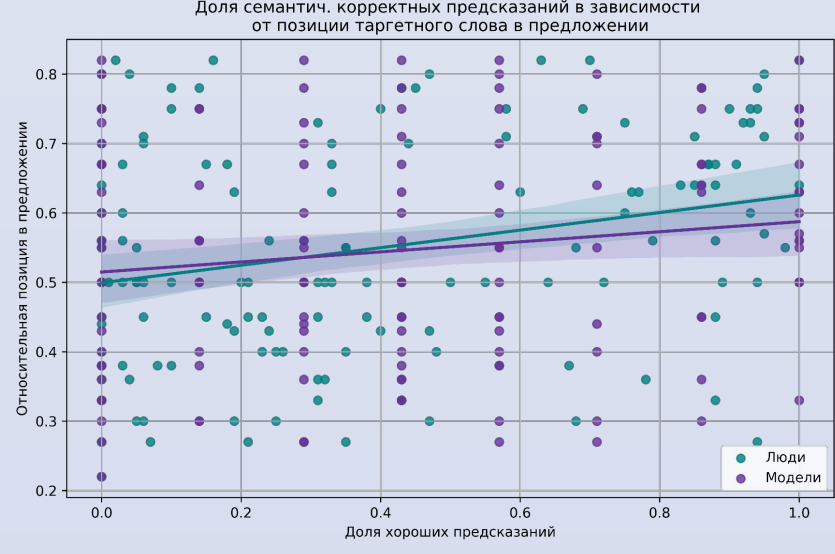
Lex vs POS\_Error



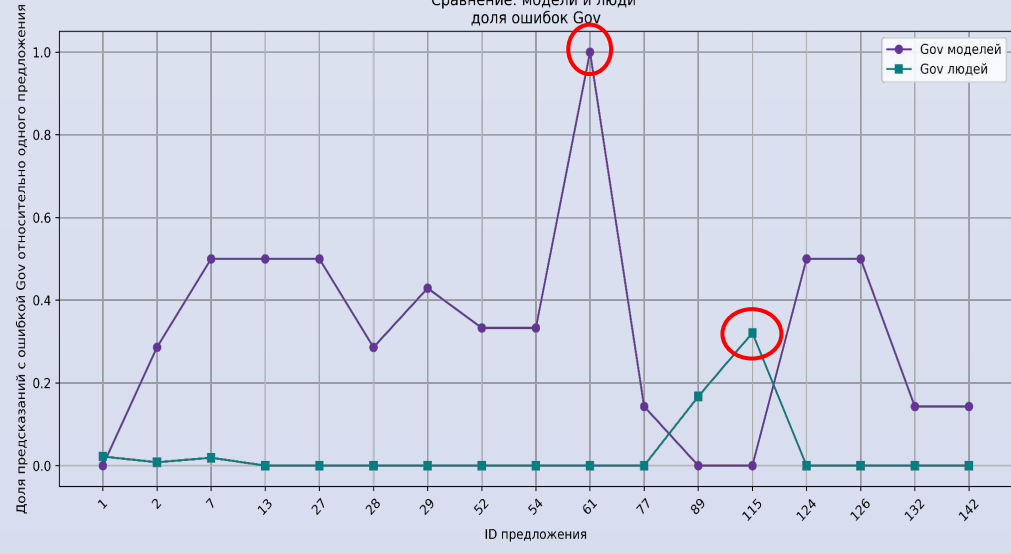
**Гипотеза 3 подтвердилась.**

И у моделей, и у людей **снижение доли приемлемых предсказаний сильнее коррелирует с увеличением частеречных ошибок, чем с лексическими**. При этом среди ответов, данных людьми, **снижение семантической приемлемости на фоне роста доли частеречных ошибок более заметно**.  
Люди и языковые модели демонстрируют **схожую чувствительность к лексическим ошибкам**.

Влияние объема контекста



Ошибки в управлении



**Гипотеза 4 подтвердилась частично.**

Есть **небольшая положительная корреляция между объемом видимого контекста и семантической приемлемостью предсказаний**. На людей этот фактор влияет в большей степени, чем на модели.

**Гипотеза 5 подтвердилась.**

**Модели чаще и больше ошибаются в выборе падежа.**

BERT результаты

