CORPUS-BASED PROBABILITIES CAN SUBSTITUTE FOR CLOZE IN READING EXPERIMENTS

Anastasiya Lopukhina (National Research University Higher School of Economics, Russia),
Anna Laurinavichyute (National Research University Higher School of Economics, Russia; University of Potsdam, Germany), Konstantin Lopukhin (Scrapinghub, Russia)
alopukhina@hse.ru

During reading or listening, people can generate predictions about the lexical and morphosyntactic properties of upcoming input based on available context (Kuperberg & Jaeger, 2016; Luke & Christianson, 2016). Psycholinguistic experiments that study predictability or control for it conventionally rely on a human-based approach and estimate predictability via the cloze task (Taylor, 1953). Despite its ubiquitous use, the cloze task might not be the most precise measure of predictability (Smith & Levy, 2011; Staub et al., 2015), as probabilities obtained in the cloze task have two major limitations. The first limitation concerns unpredictable words: no probability is available for words that did not come up in the cloze task. So we know very little about how low-predictability words affect comprehension. The second limitation is that the answers provided in the cloze task could be systematically lexically biased in a way that makes cloze probabilities an imprecise estimate of probabilities that comprehenders have in mind.

To find an alternative to cloze probabilities, several studies have compared cloze and corpus-based probabilities in the amount of variance each of them predict in reading times and eye movements while reading (Hofmann et al., 2017; Ong & Kliegl, 2008; Smith & Levy, 2011). These studies provided arguments both in favor of and against substituting cloze with corpus-based probabilities. Our study is the first that compares cloze and corpus-based lexical as well as morphosyntactic probabilities and advocates for the corpus-based approach for measuring probabilities in reading experiments.

First, we trained a long-short-term-memory (LSTM) recurrent neural network language model (Jozefowicz et al., 2016) on the Russian National Corpus (577 million tokens). Then we obtained cloze and corpus-based lexical probabilities for all 1,362 word forms in 144 Russian sentences from (Laurinavichyute et al., 2019). Probabilities obtained from participants in the cloze task and from the LSTM were very close (means were 0.184 and 0.195 respectively; r=0.68, see Fig. 1). Finally, we estimated how much variance in eye movements registered while reading the same sentences was explained by each of the two probabilities. For that, we compared the goodness of fit of the models with cloze and corpus-based probabilities using the k-fold cross-validation (k=10). For all eye fixation measures (FFD, SFD, GD, and TT), models with logit-transformed cloze and corpus-based probabilities did not differ in the goodness of fit. These results indicate that the two types of probability explain comparable amounts of variance in eye movements while reading.

Along with the traditionally studied lexical predictability (the activation of a particular word form), we analyzed word class predictability and its effect on reading over and above lexical predictability. We found that word class can be highly predictable from context and that cloze and corpus-based word class probabilities are strongly correlated (Table 1). We also found that higher word class probabilities facilitated reading over and above lexical probabilities (Fig. 2). The comparison of the models with cloze and corpus-based word class probabilities showed that for FFD, SFD, and TT the models did not differ in the goodness of fit. For GD, the model with cloze probabilities explained more variance than the model with corpus-based probabilities. Taken together, our results indicate that corpus can substitute for cloze in estimating lexical and word class probabilities in reading experiments, sparing the resources required to collect cloze data. Our results also indicate that in languages with rich inflectional morphology, such as Russian, pre-activation of word class features is much more common than prediction of words' full identity.

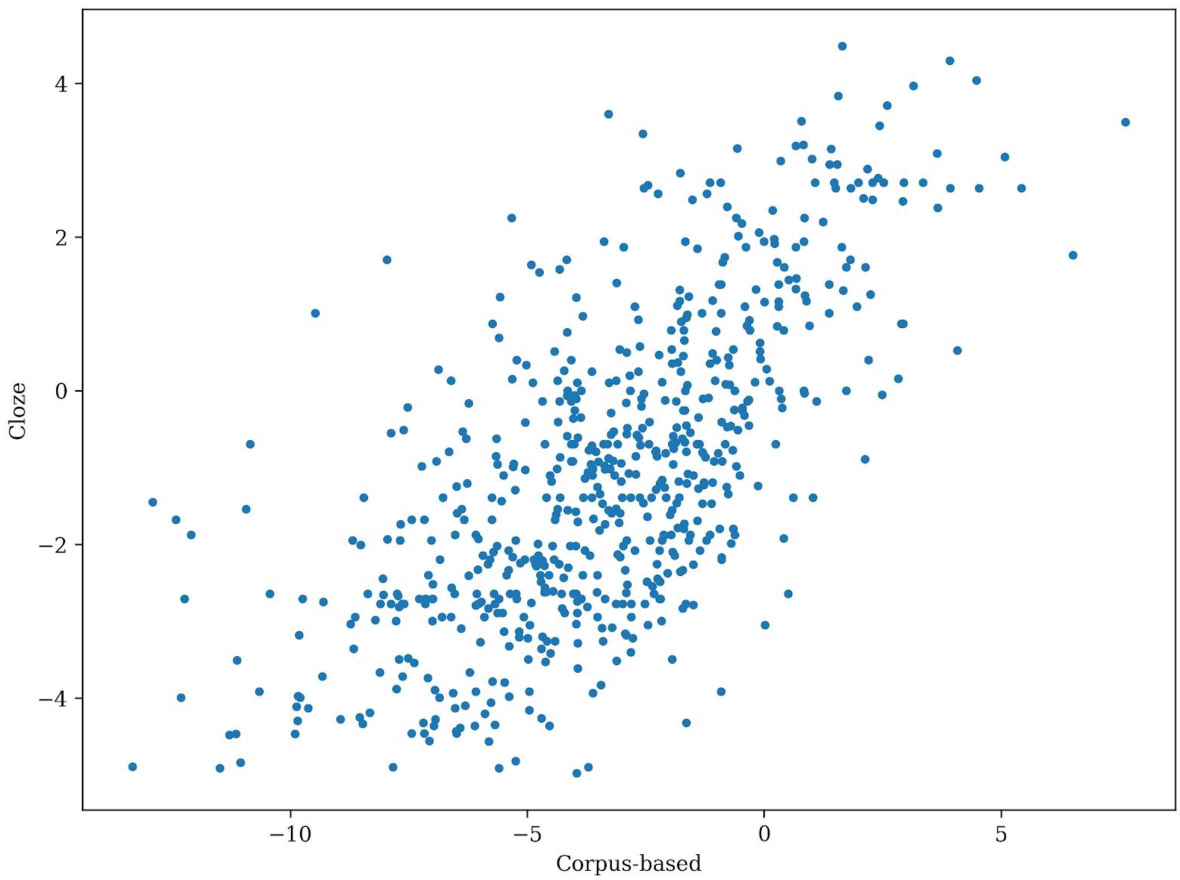

Fig. 1. Cloze versus corpus-based probabilities in logit space. Each point represents a single word for which probabilities were provided by both the LSTM model and at least one participant in the cloze experiment (there were 651 such words out of 1,218 total words).

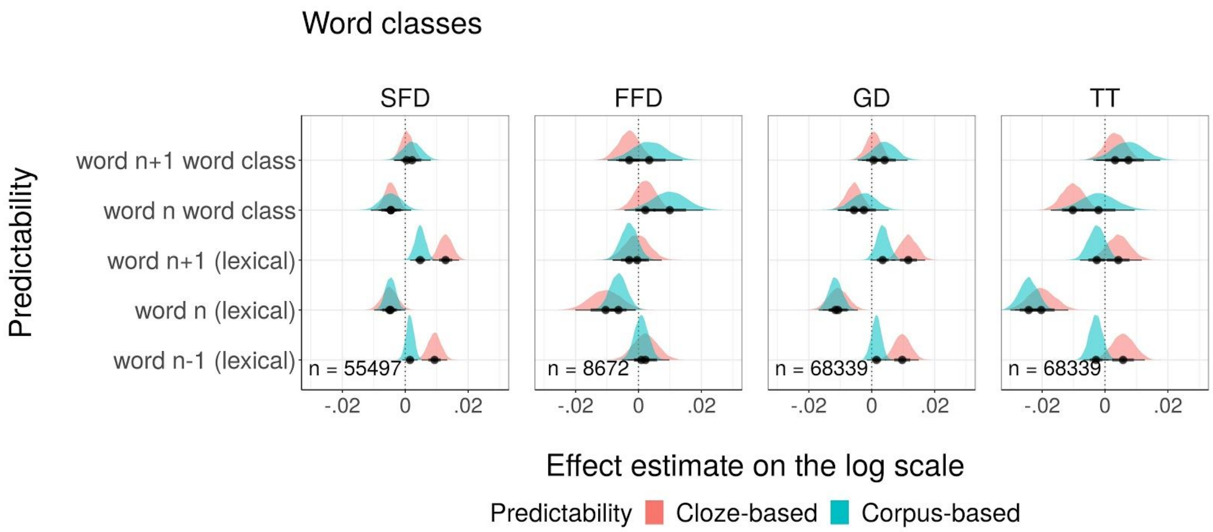

Fig. 2. Visualization of modeling estimates for four fixation duration measures.

Table 1
Mean word class probabilities and standard deviations

Word classes	# words	Mean word class cloze probabilities	Mean word class corpus-based probabilities	Pearson correlations
Content words				
nouns	439	0.76 (0.01)	0.81 (0.02)	0.71
verbs (finite forms)	190	0.66 (0.02)	0.70 (0.03)	0.63
verbs (infinitives)	52	0.65 (0.05)	0.71 (0.06)	0.72
adjectives	165	0.35 (0.02)	0.32 (0.04)	0.57
adverbs	44	0.30 (0.05)	0.16 (0.06)	0.72
numerals	6	0.45 (0.15)	0.50 (0.20)	0.00
All content words	896	0.63 (0.01)	0.66 (0.02)	0.70
Function words				
personal pronouns	69	0.47 (0.03)	0.36 (0.06)	0.67
prepositions	117	0.71 (0.03)	0.60 (0.05)	0.59
conjunctions	64	0.74 (0.03)	0.55 (0.06)	0.64
particles	32	0.52 (0.05)	0.50 (0.09)	0.71
All function words	282	0.63 (0.02)	0.52 (0.03)	0.65
All words	1178	0.63 (0.01)	0.62 (0.01)	0.68

References

Hofmann, M. J., Biemann, C., & Remus, S. (2017). Benchmarking n-grams, topic models and recurrent neural networks by cloze completions, EEGs and eye movements. In B. Sharp, F. Sedes, & W. Lubaszewski (Eds.), *Cognitive approach to natural language processing* (pp. 197–201). London: Elsevier, ISTE Press.

Jozefowicz, R., Vinyals, O., Schuster, M., Shazeer, N., & Wu, Y. (2016). Exploring the limits of language modeling. *arXiv preprint arXiv:1602.02410* / Accessed 31 March 2020.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?. *Language, cognition and neuroscience, 31(1)*, 32–59.

Laurinavichyute, A. K., Sekerina, I. A., Alexeeva, S., Bagdasaryan, K., & Kliegl, R. (2019). Russian Sentence Corpus: Benchmark measures of eye movements in reading in Russian. *Behavior research methods, 51(3)*, 1161–1178.

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology, 88*, 22–60.

Ong, J. K., & Kliegl, R. (2008). Conditional co-occurrence probability acts like frequency in predicting fixation durations. *Journal of Eye Movement Research, 2(1)*, 1–7.

Smith, N. J., & Levy, R. (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1637–1642). Austin, TX: Cognitive Science Society.

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language, 82*, 1–17.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly, 30*, 415– 433.