

INTRODUCTION

Eye-movement corpus (read by healthy adults):

- a repository of basic benchmarks of eye-movement characteristics for a language
- a testing ground for language processing theories and reading models
- a control data source of reading behavior of other readers groups (children, elderly people, bilinguals, adults with cognitive impairments, etc.)

Russian sentence corpus: Benchmark Measures of Eye movements in reading in Cyrillic

Anna Laurinavichyute (HSE, Moscow), Irina Sekerina (CSI CUNY, New York, Svetlana Alexeeva (SPbSU, St. Petersburg), Kristine Bagdasaryan (HSE, Moscow)

THE STUDY

Main goal: to create the first systematic corpus of eye movements in reading in Russian

Research Questions:

Q1. Do native speakers of Russian read texts as readers of other alphabetic languages (in particular, German) do concerning basic psycholinguistic variables (length, predictability, frequency)?

Q2. What about morphological characteristics (not tested previously)?

C1. part of speech: are verbs read longer than nouns?

C2. morphosyntactic ambiguity: are ambiguous words read longer than unambiguous words?

C3. base vs. inflected word form: are inflected words read longer than base word forms?

CORPUS DESIGN AND MATERIALS

Based on German PSC (Kliegl et al. 2004; 2006) protocol

Step 1. 144 target words (3*2*2 design): part of speech (adjective/noun/verb), length (3-4, 5-7, and 8-10 characters), and frequency (> 50 ipm or <10 ipm)

Step 2. Sentence selection using RNC and acceptability norming (215 participants) : 5-11 words in length, target word in the middle

Step 3. Collection of predictability norms (750 participants):

Predictability cloze task:

Ваня раскрыл было рот, но понял, ??? (что)

Step 4. Eye-movement data collection (96 participants): Eyelink 1000 plus (with chinrest)

RESULTS

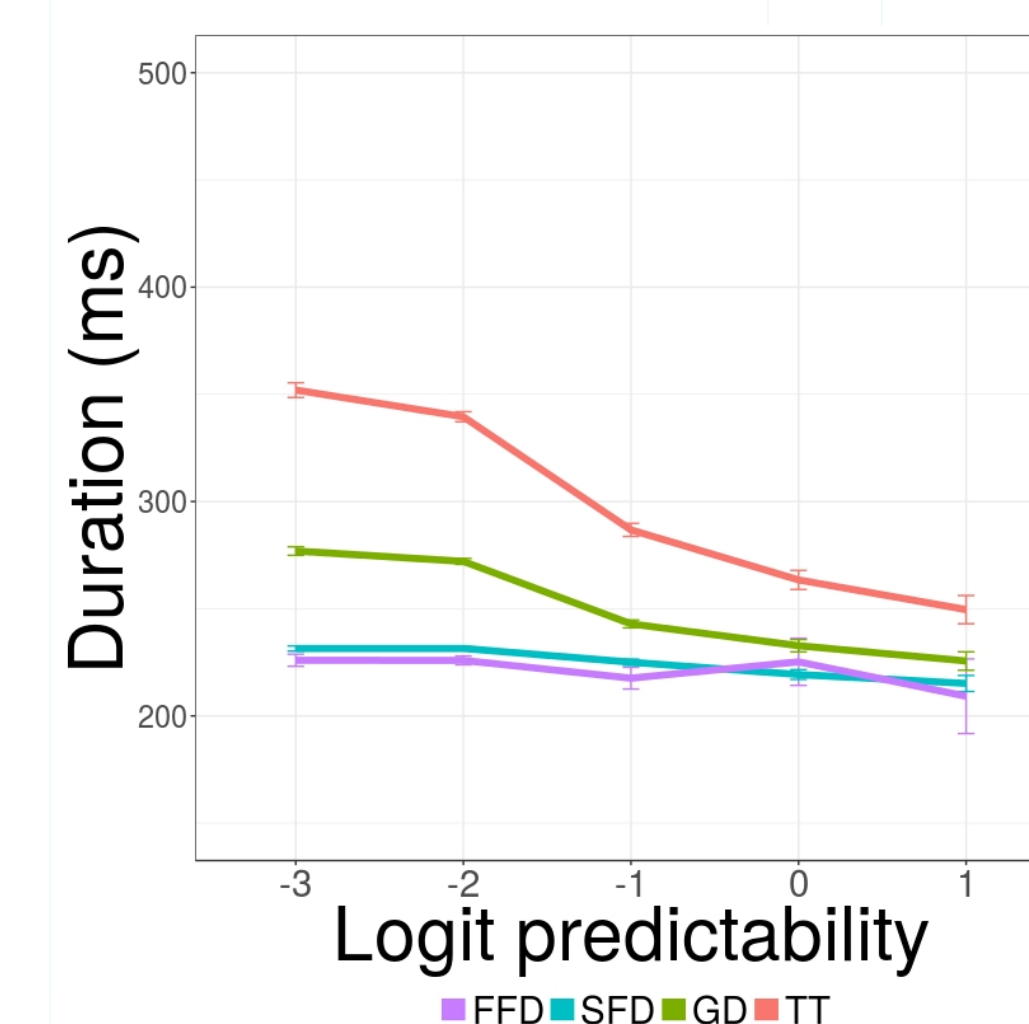
Q1. Corpus Reading. Descriptive statistics

All words	Russian	German	All words	Russian Mean *	German Mean *
N of sentences	144	144	FF	222 ms.	207 ms.
Sentence length	Range: 5-13 words, M = 9	Range: 5-11 words, M = 7.9	SF	232 ms.	210 ms.
N of words	1362	1138	GD	231 ms.	241 ms.
Word length	M = 5.6 let.		TT	283 ms.	245 ms.
Predictability	M = 18%		P0 (PSK)	17.8%	9.1%
Guesses per word	20-151	83	P1	68.8%	74.1%
% of short words	35%	41%	P2+	4.9%	17.0%
% of LF words	61%	45%	RO	17.8%	12.5%
% of LP words	65%	66%	RG	7%	0.4%

LF – low frequency; LP – low predictability; SF – single fixation duration, FF – first fixation duration (SF included), GD – gaze duration, TT – total reading time; P0, P1, P2+ – probability of zero, one, two+ first-pass fixations, PSK – probability of skipping (not fixated at all); ILP – relative initial landing position; RO, RG – probability of origin, goal of regressive eye movement. Intercept of the linear mixed-effects model with length, frequency and predictability as covariates and subject, word and sentence as random intercepts

The corpora are available for English (Kennedy & Pynthe 2005), German (Kliegl et al. 2004; 2006), French (Kennedy & Pynthe 2005), Dutch (Kuperman et al. 2010a), Chinese (Yan et al. 2006; 2010), Japanese (Sainio et al. 2007), Thai (Winskel, Radach, & Luksaneeyanawin 2009), Uighur (Yan et al. 2014) etc.

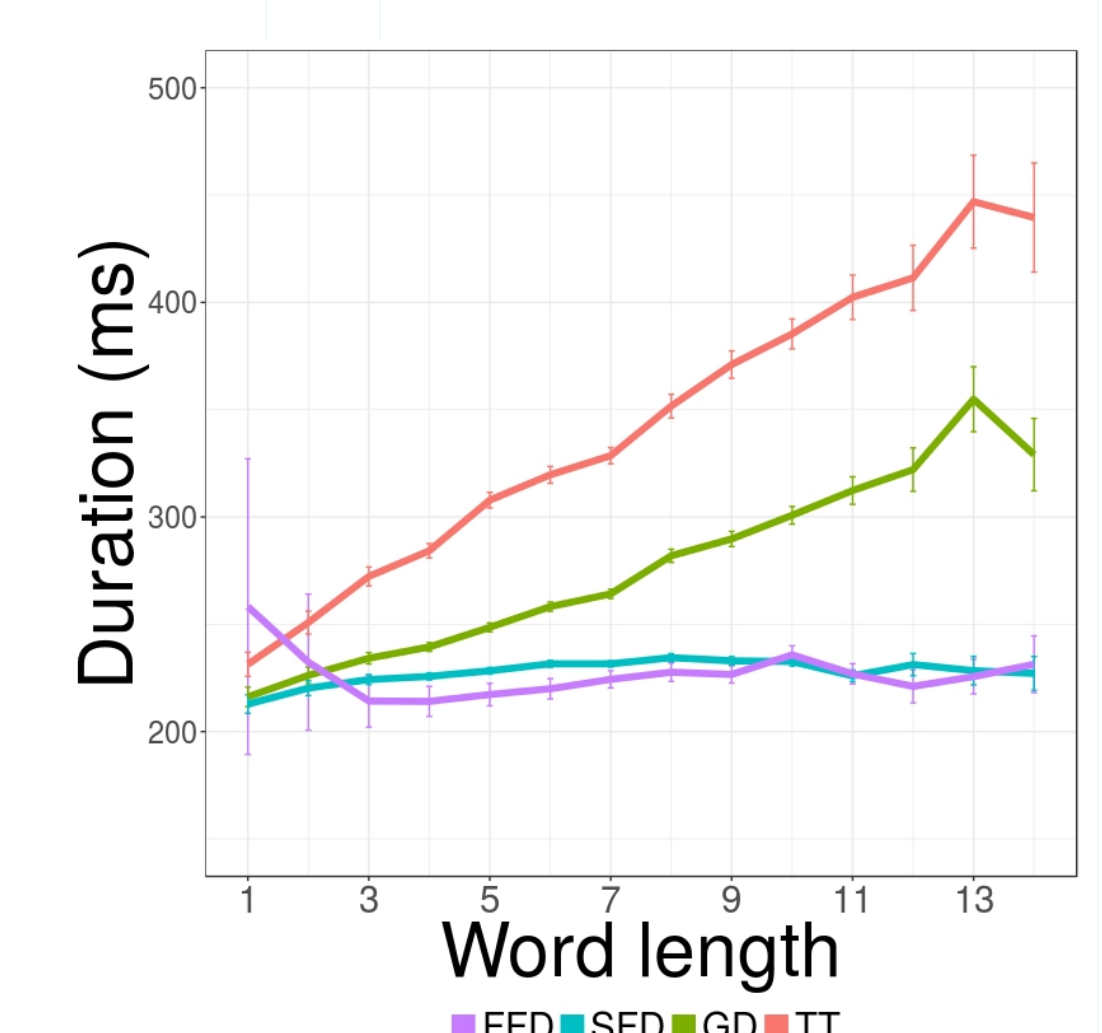
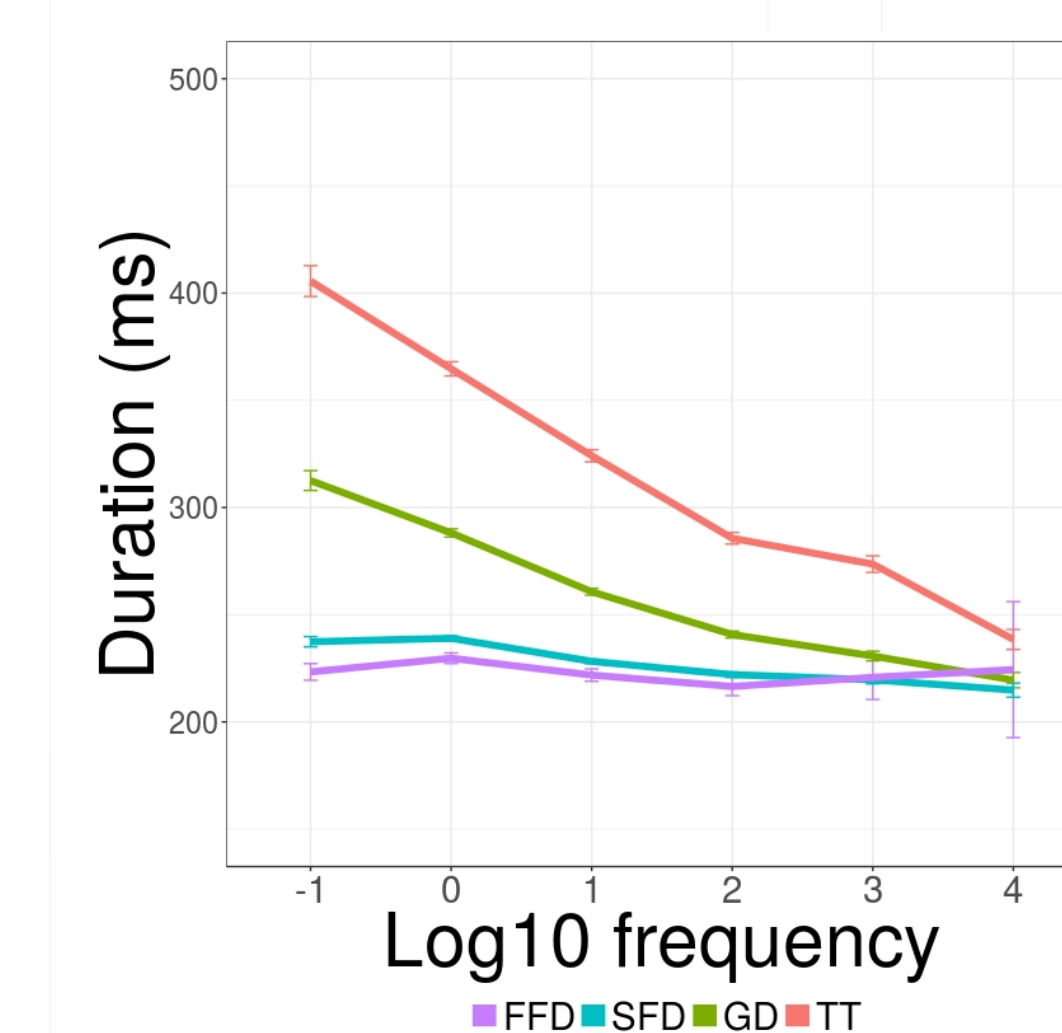
У нас в Волгограде многие придерживаются иной точки зрения.



Length, frequency and predictability effects on eye-movement measures

The frequency and predictability of the word increase, the reading times decrease (all measures);

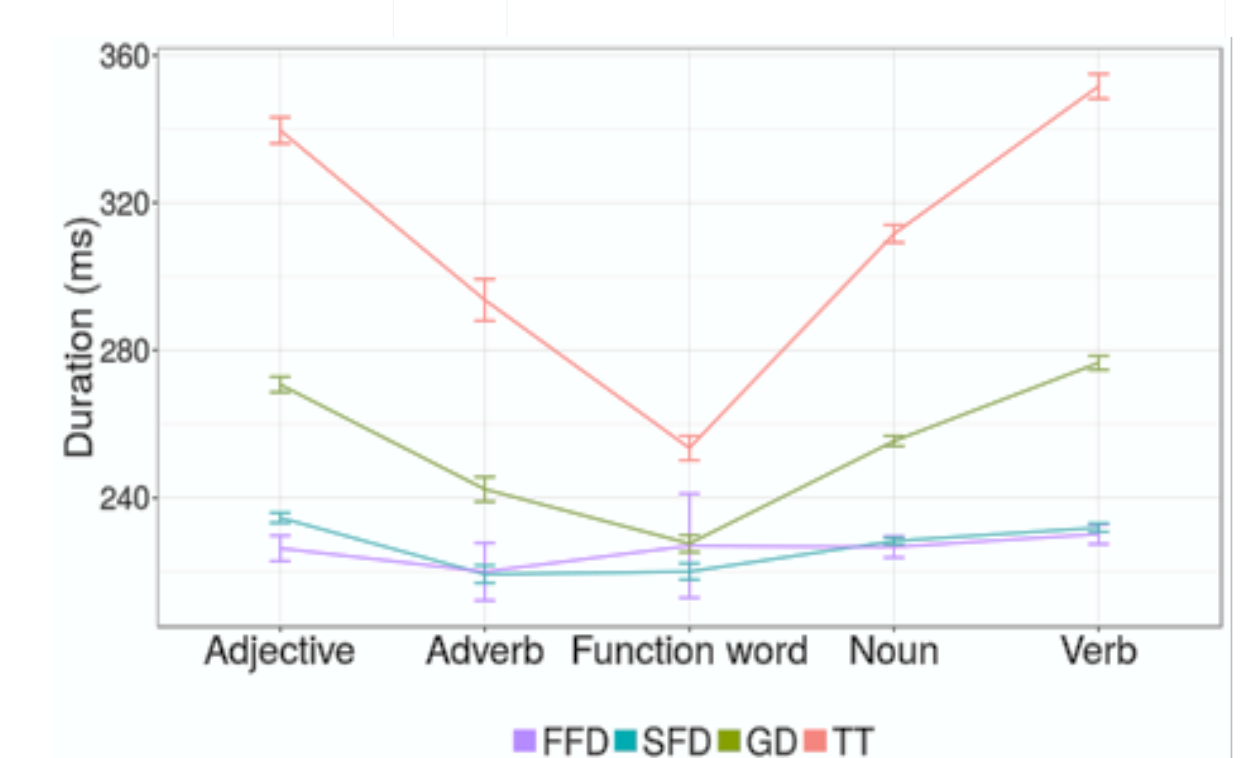
The word length increases, the reading times also increase



Q2. Focus on morphology

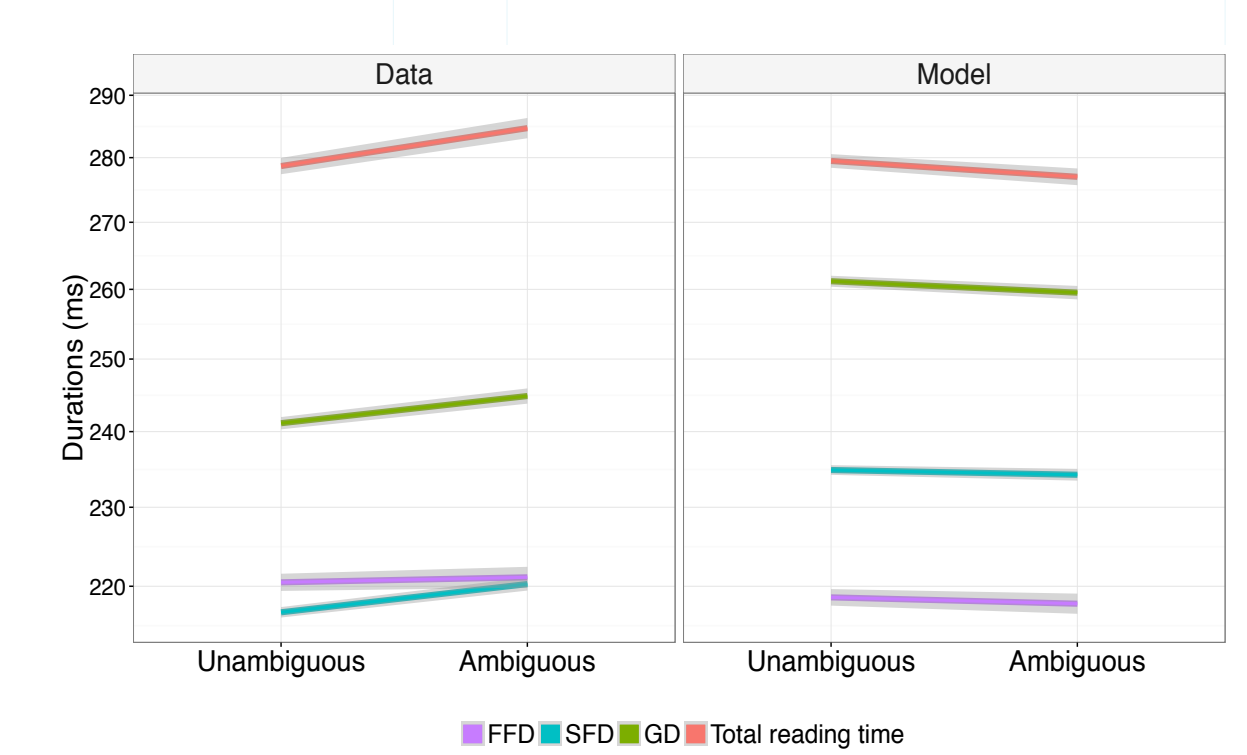
C1. Part of Speech

Nouns are read faster than verbs in the GD and TT measures. Adjectives, adverbs, and function words did not differ from the verbs in any of the measures



C2. Morphosyntactic ambiguity

There was no difference in reading times between morphosyntactically ambiguous and unambiguous word forms in the Russian



C3. Base vs. inflected word form

Inflected word forms took longer to read, but the effect was only significant in the TT measure



CONCLUSIONS

Q1. Gaze patterns are similar to the data reported for German (Kliegl et al. 2004; 2006) concerning basic word variables.

Q2. Part of speech and case influence eye-movement measures in reading, whereas morphosyntactic ambiguity does not.



ACKNOWLEDGEMENTS This project is supported by Grant N 17-34-01052 from RFBF.

Anna Laurinavichyute
annlaurin@gmail.com
Neurolinguistics Laboratory,
National Research University
Higher School of Economics,
Moscow, Russia

Irina Sekerina
irina.sekerina@csi.cuny.edu
Department of Psychology,
College of Staten Island of
The City University of New
York, New York, the USA

Svetlana Alexeeva
mail@s-alexeeva.ru
Laboratory for Cognitive Studies,
St. Petersburg State University,
St. Petersburg, Russia

Kristine Bagdasaryan
kristinabagdasaryan14@gmail.com
National Research University
Higher School of Economics,
Moscow, Russia