Effects of linguistic context and reading abilities on comprehension of unknown words

Margarita Ryzhova, Emilia Ellsiepen, Katharina Trinley, Iza Škrjanec, Vera Demberg (Saarland University, Germany)
mryzhova@lst.uni-saarland.de

Words that are unfamiliar to us can elicit processing difficulties. Word familiarity can be modulated by the intrinsic properties of the word, like frequency and length [2, 6]. However, the literature shows that the context also affects comprehension [3, 4, 7]. For example, scientific or technical texts may contain more specialized vocabulary that is unfamiliar to the general reader. In contrast, everyday texts such as newspapers or novels may contain more familiar language. In such common contexts, the reader can be surprised to encounter an unknown word or attribute it to a typo, while in a more scientific context, the reader might expect to encounter special domain terms they don't know. On the other hand, substantial evidence indicates that reading comprehension is influenced by the reader's literacy. More skilled readers are better at monitoring their comprehension, recognizing when additional processing is necessary, and are more motivated to fully understand the texts they read, thereby investing greater cognitive effort [1, 5].

In our studies on processing unknown words in German, we manipulate the type of context to explore its effect on readers' sensitivity to unfamiliar words. Additionally, we assess each participant's reading proficiency to investigate potential interactions between context and literacy (ART and vocabulary size). We conducted two self-paced reading experiments (with two sets of materials that were only partially overlapping) and asked participants to read texts for comprehension. Each text includes a target word: either a *real word* or a *pseudoword*. The target words were embedded into two types of contexts: *everyday* and *scientific*, making both studies follow a 2x2 design. Everyday stories concern familiar events from daily life (e.g., children playing in a park), while scientific stories occur in less common settings with characters with a specialized profession (e.g., researchers conducting experiments in a laboratory). The scientific stories themselves are not expository texts but rather narratives describing a less familiar scenario.

Our results confirm that readers are sensitive to pseudowords in *everyday* and *scientific* contexts, leading to increased reading times. However, evidence across the two studies is mixed regarding whether the context influences the processing of unknown words – see Table 1 for model specifications and results. Overall, the trend indicates that pseudowords are read more slowly in *everyday* than in *scientific* context, which may suggest that unknown words, despite their lack of a defined meaning, are more expected in domain-specific texts than in general narratives, resulting in faster reading. This effect, however, was only significant in one of the studies. We also find that high-literacy readers take more time to process pseudowords than low-literacy readers, regardless of context. This may reflect a greater effort by high-literacy readers to understand and integrate unfamiliar words into the context.

At the time of abstract submission, we are collecting data for an eye-tracking counterpart of this study. Our motivation is that eye-tracking can provide deeper insights into the processing of pseudowords and the nature of reading times through regressions and second-pass fixation durations. For literacy, early eye-tracking measures may reveal that high-literacy readers recognize pseudowords more quickly, while later measures may indicate that these readers invest more effort in integrating unknown words into the context.

	Critical region (Study 1)				Spillover region (Study 1)				Critical region (Study 2)				Spillover region (Study 2)			
Predictors	*Est*	*Std. Error*	*t-value*	*p*	*Est*	*Std. Error*	*t-value*	*p*	*Est*	*Std. Error*	*t-value*	*p*	*Est*	*Std. Error*	*t-value*	*p*
(Intercept)	795.91	24.76	32.15	**<0.001**	627.76	18.92	33.17	**<0.001**	843.46	19.80	42.61	**<0.001**	680.27	17.39	39.13	**<0.001**
Story (scientific)	-22.64	12.75	-1.78	0.077	-7.43	8.44	-0.88	0.379	12.61	14.81	0.85	0.395	-19.69	13.74	-1.43	0.152
Word (pseudo)	76.78	16.62	4.62	**<0.001**	92.68	11.95	7.76	**<0.001**	197.76	18.73	10.56	**<0.001**	233.57	17.20	13.58	**<0.001**
Story:Word	-12.89	13.18	-0.98	0.328	-18.28	8.08	-2.26	**0.024**	-23.66	22.36	-1.06	0.290	-28.28	18.54	-1.53	0.128
Trial	-32.84	13.55	-2.42	**0.016**	-7.31	8.28	-0.88	0.378	-35.95	9.03	-3.98	**<0.001**	-13.68	6.02	-2.27	**0.023**
Word:Trial	-27.09	13.60	-1.99	**0.047**	-2.44	8.73	-0.28	0.780	-51.99	16.24	-3.20	**0.001**	-0.92	11.06	-0.08	0.934
Chunk	-15.49	19.53	-0.79	0.428	-4.88	10.76	-0.45	0.650	-21.75	11.57	-1.88	0.061	-5.91	8.85	-0.67	0.504
log(Frequency)	-7.56	16.89	-0.45	0.655	-16.77	11.74	-1.43	0.154	-14.62	11.39	-1.28	0.200	-3.33	4.15	-0.80	0.423
Num. Chars Chunk	50.23	19.46	2.58	**0.010**	12.19	13.40	0.91	0.363	11.50	11.27	1.02	0.308	3.61	10.61	0.34	0.734
Num. Words Chunk					9.72	13.53	0.72	0.473					4.93	9.83	0.50	0.616
ART									41.05	17.36	2.37	**0.018**	12.19	13.99	0.87	0.384
Vocab									-24.77	20.48	-1.21	0.227	17.61	14.48	1.22	0.224
Story:Vocab									1.61	17.53	0.09	0.927	3.62	11.03	0.33	0.743
Word:Vocab									2.86	16.23	0.18	0.860	54.51	14.70	3.71	**<0.001**
Story:Word:Vocab									23.77	22.24	1.07	0.285	8.82	17.09	0.52	0.606
Story:ART									-6.46	15.96	-0.40	0.686	-15.40	12.10	-1.27	0.203
Word:ART									26.06	15.98	1.63	0.103	-23.66	16.26	-1.46	0.146
Story:Word:ART									-22.14	33.71	-0.66	0.511	-9.68	16.18	-0.60	0.550

Table 1. Regression coefficients and test statistics from the Generalized Gamma mixed-effects models (with identity link) of reading times in critical and spillover regions in Studies 1 and 2.

References:

1. **Broek, P. v. d., White, M. J., Kendeou, P., & Carlson, S. (2009).** Reading between the lines: Developmental and individual differences in cognitive processes in reading comprehension. In R. K. Wagner, C. Schatschneider, & C. Phythian-Sence (Eds.), *Beyond decoding: The behavioral and biological foundations of reading comprehension* (pp. 107–123). The Guilford Press.
2. **Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004).** Length, frequency, and predictability effects of words on eye movements in reading. European journal of cognitive psychology, 16 (1-2), 262–284. https://doi.org/10.1080/09541440340000213
3. **Lowell, R., Morris, R.K. (2014).** Word length effects on novel words: Evidence from eye movements. Atten Percept Psychophys 76, 179–189. https://doi.org/10.3758/s13414-013-0556-4
4. **Nieuwland, M. S., & Van Berkum, J. J. (2006).** When peanuts fall in love: N400 evidence for the power of discourse. Journal of cognitive neuroscience, 18(7), 1098–1111. https://doi.org/10.1162/jocn.2006.18.7.1098
5. **Oakhill, J.V., Hartt, J., & Samols, D. (2005).** Levels of Comprehension Monitoring and Working Memory in Good and Poor Comprehenders. Reading and Writing, 18, 657-686. https://doi.org/10.1007/s11145-005-3355-z
6. **Rayner, K. (1998)**. Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372
7. **Williams, R. S., & Morris, R. K. (2004).** Eye movements, word familiarity, and vocabulary acquisition. European Journal of Cognitive Psychology, 16(1-2), 312–339. https://doi.org/10.1080/09541440340000196