

Word Familiarity Classification From a Single Trial Based on Eye-Movements. A Study in German and English

Margarita Ryzhova*
Iza Škrjanec*
mryzhova@coli.uni-saarland.de
skrjanec@lst.uni-saarland.de
Saarland University
Saarbrücken, Germany

Nina Quach
Saarland University
Saarbrücken, Germany
s8niquac@stud.uni-saarland.de

Alice Chase
Saarland University
Saarbrücken, Germany
avchase@coli.uni-saarland.de

Emilia Ellsiepen
Saarland University
Saarbrücken, Germany
emilia@lst.uni-saarland.de

Vera Demberg
Saarland University
Saarbrücken, Germany
vera@coli.uni-saarland.de

ABSTRACT

Identifying processing difficulty during reading due to unfamiliar words has promising applications in automatic text adaptation. We present a classification model that predicts whether a word is (un)known to the reader based on eye-movement measures. We examine German and English data and validate our model on unseen subjects and items achieving a high accuracy in both languages.

CCS CONCEPTS

• **Applied computing** → *Psychology*; • **Computing methodologies** → **Cross-validation**; • **General and reference** → **Experimentation**.

KEYWORDS

Eye-tracking, Word familiarity, Classification

ACM Reference Format:

Margarita Ryzhova, Iza Škrjanec, Nina Quach, Alice Chase, Emilia Ellsiepen, and Vera Demberg. 2023. Word Familiarity Classification From a Single Trial Based on Eye-Movements. A Study in German and English. In *2023 Symposium on Eye Tracking Research and Applications (ETRA '23)*, May 30–June 02, 2023, Tübingen, Germany. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3588015.3590118>

1 INTRODUCTION

Comprehension difficulty during reading can arise from unfamiliar words, i.e. reading a word we have not encountered before. Research shows that this difficulty is evident from eye movements [Lowell and Morris 2014]. In this work we present a classification model that uses reading measures to predict whether or not a reader is familiar with the target word. While most related studies [Ahn et al. 2020;

*Both authors contributed equally to this research.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
ETRA '23, May 30–June 02, 2023, Tübingen, Germany
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0150-4/23/05.
<https://doi.org/10.1145/3588015.3590118>

Bingel et al. 2018; Makowski et al. 2018; Reich et al. 2022] focus on predicting comprehension and difficulty from eye movements on longer text passages, our study attempts classification based on a single trial. We evaluate the robustness of our model by testing on unseen subjects and items. The study was conducted for English and German.

2 EXPERIMENT

The materials comprised 12 sentences each in two conditions: control and pseudo. In the German example “Viel Geld wurde investiert, bevor der **Satellit**|**Latendit** endlich in die Umlaufbahn gelangte”, “Satellit” is the control, while “Latendit” is the pseudoword. An English example is “Sally was impressed by the **hospitality**|**turbitolity** she experienced during her travel”. Control words were commonly-known¹ real words (7-14 characters long), placed in the middle of the sentence. Pseudowords were generated wrt. the control words using Wuggy [Keuleers and Brysbaert 2010]. Each subject read 3 sentences in the pseudo condition, 9 in the control condition, and 16 fillers. Subjects’ eyes were tracked with an EyeLink 1000+ (SR Research) eye tracker. Subjects were native speakers of the corresponding language with normal or corrected-to-normal vision. For German, we had 18 subjects (8 female; 19-53 y.o.) and for English, 20 (7 female, 1 non-binary; 18-33 y.o.).

3 MODELLING

The analysis procedure was the same for both languages. First, we use filler words to fit a regression model for each reading measure separately (first-fixation duration, first-pass duration, number of regressions out, number of fixations, and dwell time) with word length, log word frequency and log trial order as main effects and by-subject random intercepts [Ferreira and Clifton 1986]. Then, we use these models to residualize the predictors of a logistic regression model to classify critical words as known vs. unknown to the reader. The classifier is trained with k -fold cross validation. We test the classifier’s robustness in three scenarios: 1) known subjects and items, 2) held-out subjects, 3) held-out items. The classification model included only significant predictors. For both languages,

¹Word frequency estimates were obtained from the CELEX database [Baayen et al. 1995].

dwel time was the most important predictor in the classifier. For English, the number of fixations furthermore significantly improved the model fit. In both languages, the performance degrades slightly for unseen subjects, and also for unseen items in English (Table 3). The confusion matrices (Tables 1 and 2, for German and English respectively) show known words are rarely misclassified, while unknown words might sometimes be read faster and therefore be misinterpreted by the model as known.

Table 1: Confusion matrix for single-trial word familiarity classification in German

		Predicted	
		Control	Pseudo
True	Control	141	2
	Pseudo	7	42

Table 2: Confusion matrix for single-trial word familiarity classification in English

		Predicted	
		Control	Pseudo
True	Control	170	4
	Pseudo	10	50

Table 3: Pseudoword F-scores in 3 scenarios.

	German	English
1	0.90	0.88
2	0.89	0.84
3	0.90	0.87

ACKNOWLEDGMENTS

The work reported here was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 232722074 – SFB 1102 "Information Density and Linguistic Encoding".

REFERENCES

- Seoyoung Ahn, Conor Kelton, Aruna Balasubramanian, and Greg Zelinsky. 2020. Towards Predicting Reading Comprehension From Gaze Behavior. In *ETRA 2020* (Stuttgart, Germany). Association for Computing Machinery, New York, NY, USA, Article 32, 5 pages.
- R. Harald Baayen, Richard Piepenbrock, and Leon Gulikers. 1995. *The CELEX Lexical Database. Release 2 (CD-ROM)*. Linguistic Data Consortium.
- Joachim Bingel, Maria Barrett, and Sigrid Klerke. 2018. Predicting misreadings from gaze in children with reading difficulties. In *Proceedings of BEA 13*. ACL, New Orleans, Louisiana, 24–34.
- Fernanda Ferreira and Charles Clifton. 1986. The independence of syntactic processing. *Journal of Memory and Language* 25, 3 (1986), 348–368.
- Emmanuel Keuleers and Marc Brysbaert. 2010. Wuggy: A multilingual pseudoword generator. *Behavior Research Methods* 42 (2010), 627–633.
- Randy Lowell and Robin K. Morris. 2014. Word length effects on novel words: Evidence from eye movements. *Attention, Perception, & Psychophysics* 76, 1 (01 Jan 2014), 179–189.
- Silvia Makowski, Lena A. Jäger, Ahmed AbdelWahab, Niels Landwehr, and Tobias Scheffer. 2018. A Discriminative Model for Identifying Readers and Assessing Text Comprehension from Eye Movements. In *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2018, Dublin, Ireland, September 10-14, 2018, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 11051)*. Springer, 209–225.
- David Robert Reich, Paul Prasse, Chiara Tschirner, Patrick Haller, Frank Goldhammer, and Lena A. Jäger. 2022. Inferring Native and Non-Native Human Reading Comprehension and Subjective Text Difficulty from Scanpaths in Reading. In *ETRA 2022*. ACM, NY, USA, Article 23, 8 pages.