

Visual and auditory lexical decision in German: An online megastudy

Sun, C.C.¹, Wientzek, T.² & Hendrix, P.³

¹University of Tübingen, ching-chu.sun@uni-tuebingen.de

²University of Tübingen, tim.wientzek@uni-tuebingen.de

³Tilburg University, p.h.g.hendrix@tilburguniversity.edu

Keywords: Visual lexical decision, Auditory lexical decision, German, Megastudy, Online experiment

In recent years, there has been a growing inclination towards the collection and examination of extensive quantities of behavioral data through an online platform. The implementation of online methods in psycholinguistic research offers several advantages (Aguasvivas et al, 2020), including the ability to reach a large and diverse sample of participants. By recruiting participants from a wide range of demographic backgrounds, online experiments can enhance the generalizability of the results and provide a more representative sample of the population.

The lexical decision task is a widely adopted methodology in the field of psycholinguistics, primarily used for investigating the cognitive processes involved in word recognition. While massive data sets of reaction times collected from lexical decision tasks are available in several languages (Keuleers et al, 2010, 2015; Brysbaert et al, 2016; Tucker et al, 2019), to the best of our knowledge, no large-scale lexical decision experiments in German have been conducted through crowdsourcing. In light of this, the use of online platforms for conducting large-scale lexical decision experiments in German could prove to be a valuable tool for advancing our understanding of the cognitive processes involved in word recognition in German.

In the current study, we present two online lexical decision experiments, one in the visual domain (Lexical Decision: LD) and the other in the auditory domain Auditory Lexical Decision: ALD). Both experiments include a total of 27,060 real words and 24,850 pseudo-words. The real words were obtained from the deCow corpus, a corpus of German texts compiled by Schäfer et al. (2014). Pseudo-words were generated using the Wuggy pseudo-word generator (Keuleers & Brysbaert, 2010) and further selection was done through custom code to closely mimic the phonotactic patterns of the German language. For the ALD, all real words and pseudo-words were recorded by a single speaker, who pronounced the pseudo-words as naturally as possible.

Although additional participants are needed for both experiments, sufficient data from the LD have been collected to perform an initial statistical analysis. Thus we present preliminary results of analyzing the LD reaction times using a linear regression model. At the time of writing this abstract, a total of 451,832 observations have been collected in the lexical decision (LD) experiment. For the purpose of the preliminary analysis, incorrect responses, reaction times that fall outside of a specified range (shorter than 300ms or longer than 2,000ms), and responses by non-native speakers were removed.

A linear regression model fit to the data, revealed that both frequency and word length play a significant role in reaction times, consistent with the findings of previous studies. Furthermore, we observed that reaction times were shorter on computers compared to smartphones. Additionally, the analysis indicated that female participants had faster reaction times than male participants, while the accuracy of males and females was similar. We intend to continue the collection of data in both the visual and auditory domain and publicly release the collected data.

References

- Aguasvivas, J., Carreiras, M., Brysbaert, M., Mander, P., Keuleers, E., & Duñabeitia, J. A. (2020). How do Spanish speakers read words? Insights from a crowdsourced lexical decision megastudy. *Behavior research methods*, 52(5), 1867-1882.
- Brysbaert, M., Stevens, M., Mander, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, 42(3), 441.
- Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior research methods*, 42(3), 627-633.
- Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in psychology*, 1, 174.
- Keuleers, E., Stevens, M., Mander, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *The Quarterly Journal of Experimental Psychology*, 68(8), 1665-1692.
- Schäfer, R., Barabesi, A., & Bildhauer, F. (2014, April). Focused web corpus crawling. In *Proceedings of the 9th Web as Corpus workshop (WAC-9)* (pp. 9-15).
- Tucker, B. V., Brenner, D., Danielson, D. K., Kelley, M. C., Nenadić, F., & Sims, M. (2019). The massive auditory lexical decision (MALD) database. *Behavior research methods*, 51(3), 1187-1204.