

The German Lexicon Project

Aliona Petrenco¹, Jana Hasenäcker², Benjamin Gagl³, Louis Schiekiera⁴, Fritz Günther⁵, Xenia Schmalz⁶, Tanja C. Roembke⁷

¹ Humboldt State University, Germany

² Linguistics Department, Universität Erfurt, Germany

³ Universität Köln, Germany

⁴ Institute of Psychology, Humboldt Universität Berlin, Germany & Department of Education and Psychology, Freie Universität Berlin, Germany

⁵ Department of Psychology, Humboldt Universität Berlin, Germany

⁶ Department of Child and Adolescent Psychiatry, University of Munich, Germany

⁷ Institute of Psychology, RWTH Aachen University, Germany

Background: Large-scale lexical databases provide an empirical foundation for testing (psycho)linguistic hypotheses and for building and refining computational and cognitive models of word processing. The chronometric measurements collected in such megastudies allow continuous, regression-based analyses across tens of thousands of items, enable researchers to examine functional relationships between continuous predictors and outcome variables, and allow the evaluation of the overall predictive accuracy of models [1].

Building on the substantial impact of earlier lexicon projects [2–7], the German Lexicon Project (GLP) addresses a long-standing gap for German by launching a large-scale, multi-lab data collection involving more than a dozen research centers across German-speaking regions of Europe.

Method: The GLP follows the TRUST network [8] guidelines for transparency, transferability, and sustainability for psycholinguistic reading studies in German. It is designed to collect lexical decision data from over a thousand participants across over 60,000 items (30,000 words and 30,000 pseudowords), resulting in nearly two million trials. Stimuli are extensively matched on orthographic, phonological, and sublexical properties and include rich item-level information, such as letter, phoneme, and phone counts, neighborhood density measures, and detailed transition frequency statistics. Together, these features make the GLP a robust resource for investigating lexical, orthographic, phonological, and semantic processing in German.

Results: Data collection started in December 2025 and is ongoing (Figure 1). We will present the preliminary findings from the GLP, including response time and accuracy effects of key lexical, orthographic, and phonological predictors. In addition, we will illustrate how this dataset—and large-scale open-access lexical resources more generally—can be used to advance research on word processing.

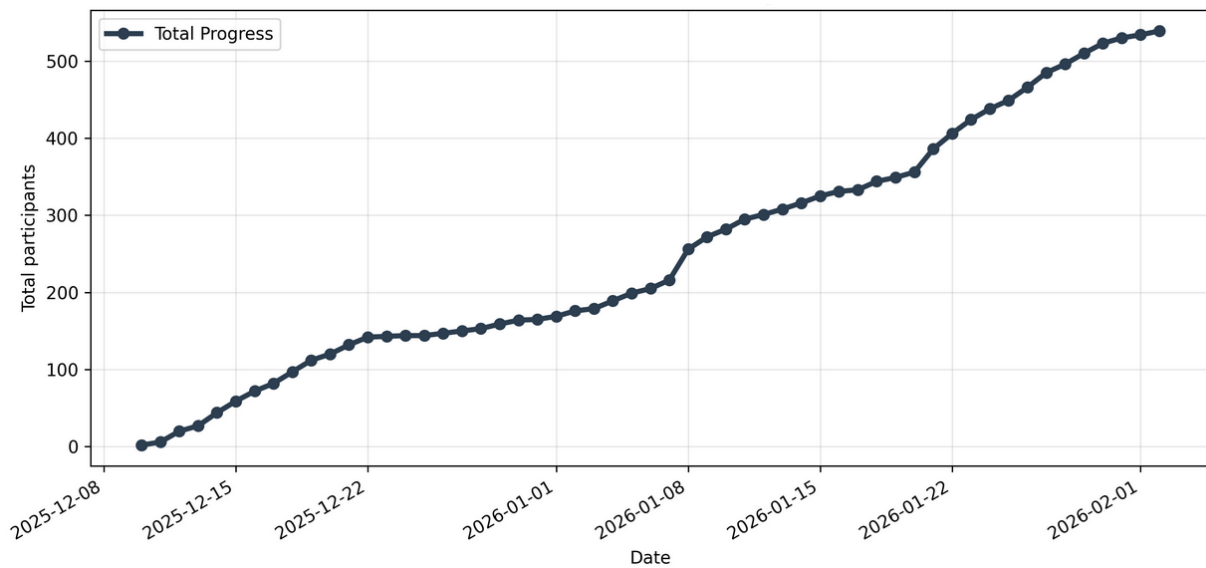


Figure 1: Between December 2025 and February 2026, 539 of 1,440 planned datasets (37.4%) were collected across all testing sites.

References

- [1] Keuleers, E., & Balota, D. (2015). Megastudies, crowdsourcing, and large datasets in psycholinguistics: An overview of recent developments. *Quarterly Journal of Experimental Psychology*, 68, 1457–1468. <https://doi.org/10.1080/17470218.2015.1051065>
- [2] Amenta, S. (2025). The Italian Crowdsourcing Project: Visual word recognition times for 130,495 Italian words. *Behavior Research Methods*, 57(1), 1–18. <https://doi.org/10.3758/s13428-024-02548-4>
- [3] David, A., & Balota. (2007). The English lexicon project. *Behavior research methods*, 39(3), 445–459.
- [4] Brysbaert, M., & Keuleers, E. (2025). *The Dutch Lexicon projects*. Elsevier.
- [5] Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, 44(1), 287–304. <https://doi.org/10.3758/s13428-011-0118-4>
- [6] Schröter, P., & Schroeder, S. (2017). The Developmental Lexicon Project: A behavioral database to investigate visual word recognition across the lifespan. *Behavior Research Methods*, 49(6), 2183–2203. <https://doi.org/10.3758/s13428-016-0851-9>
- [7] Wei, P., Sze, S. J. R., Liow, M., & Yap. (2014). The Chinese Lexicon Project: A repository of lexical decision behavioral responses for 2,500 Chinese characters. *Behavior Research Methods*, 46(1), 263–273.
- [8] TRUST Network. (2026). The TRUST Network: Best practice guides for single-word reading research in German [Accessed: 2026-06-26]. *GeWoNN*.