MultiplEYE Final Conference
August 20-21, 2026
Skopje, North Macedonia

A Living Overview of Eye-Tracking-while-Reading Datasets with Open Library Support

Jana Hofmann[1], Deborah N. Jakobi[*1], David R. Reich[1,2], Paul Prasse[2], Lena S. Bolliger[1], Lena A. Jäger[1,3]

[1]Department of Computational Linguistics, University of Zurich
[2]Department of Computer Science, University of Potsdam
[3]Institute of Informatics, University of Zurich

*contact: deborahnoemie.jakobi@uzh.ch

Eye-tracking-while-reading corpora are a valuable resource for a wide range of disciplines and applications. These applications span from investigating the cognitive processes involved in reading to machine-learning-based uses, such as gaze-driven assessments of reading comprehension (e.g., Rayner *et al.*, 2006; Ahn *et al.*, 2020; Reich *et al.*, 2025). In recent decades, both the number and size of eye-tracking-while-reading datasets have grown, along with increasing diversity in terms of stimulus languages, participants' linguistic backgrounds, and the inclusion of psychometric or demographic information. However, the distribution of data across different disciplines, combined with the absence of common data-sharing standards, has resulted in many existing datasets being difficult to reuse due to limited interoperability.

To overcome the lack of transparency and clarity with regards to existing datasets and their features across different disciplines, we present a living dataset review with open library support which can be found at the following link: `https://t.uzh.ch/1Yh`. The purpose of this review is to present existing datasets including a wide range of different features. These features include the number of participants and items, description of the stimuli, information on available data formats, and many more. The living nature of the review allows for adding new datasets as they are created by their authors. In addition, already added datasets can be edited should there be incomplete or missing information by sending an edit request. In addition, all publicly available datasets have been integrated into the Python package *pymovements* (Krakowczyk *et al.*, 2023) which offers an eye-tracking datasets library (Krakowczyk *et al.*, 2025).

The living overview as well as the *pymovements* integration simplifies sharing of new datasets, makes existing datasets more visible and their features transparent, and therefore strengthens the FAIR (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable) principles (Wilkinson *et al.*, 2016) and encourages reproducing and replicating studies as good scientific practices.

References

Ahn S., Kelton C., Balasubramanian A., Zelinsky G. (2020). Towards predicting reading comprehension from gaze behavior. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA '20 short papers, pp. 1–5. Association for Computing Machinery. doi:10.1145/3379156.3391335.

Krakowczyk D. G., Reich D. R., Chwastek J., Süss A., Prasse P., Jakobi D. N., Turuta O., Kasprowski P., Jäger L. A. (2023). pymovements: A Python package for eye movement data processing. In *Proceedings of the ACM Symposium on Eye-Tracking Research and Applications*, ETRA '23, pp. 1–2. Association for Computing Machinery.

Krakowczyk D. G., Reich D. R., Säuberli A., Škrjanec I., Cretton I. C. R., Jakobi D. N., Nisioi S., Prasse P., Jäger L. A. (2025). The more the merrier: Boost your dataset visibility and discover eye-tracking datasets with pymovements. In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications*, ETRA '25, pp. 1–3. Association for Computing Machinery. doi:10.1145/3715669.3726810.

Rayner K., Chace, Kathryn H. S. T. J., Ashby J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading 10*(3), 241–255. doi:10.1207/s1532799xssr1003_3.

Reich D., Shubi O., Jäger L., Berzak Y. (2025). Eye Tracking and NLP. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics*, pp. 2. Association for Computational Linguistics. doi:10.18653/v1/2025.acl-tutorials.2.

Wilkinson M. D., Dumontier M., Aalbersberg I. J., Appleton G., Axton M., Baak A., Blomberg N., Boiten J.-W., da Silva Santos L. B., Bourne P. E., Bouwman J., Brookes A. J., Clark T., Crosas M., Dillo I., Dumon O., Edmunds S., Evelo C. T., Finkers R., Gonzalez-Beltran A., Gray A. J. G., Groth P., Goble C., Grethe J. S., Heringa J., 't Hoen P. A. C., Hooft R., Kuhn T., Kok R., Kok J., Lusher S. J., Martone M. E., Mons A., Packer A. L., Persson B., Rocca-Serra P., Roos M., van Schaik R., Sansone S.-A., Schultes E., Sengstag T., Slater T., Strawn G., Swertz M. A., Thompson M., van der Lei J., van Mulligen E., Velterop J., Waagmeester A., Wittenburg P., Wolstencroft K., Zhao J., Mons B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data 3*(1), 160018. doi:10.1038/sdata.2016.18.