



MultipleYE
FINAL CONFERENCE

BOOK OF ABSTRACTS

Advancing multilingual eye-tracking-while-reading research

20-21 August, Skopje, North Macedonia



"Ss. Cyril and Methodius" University in Skopje
**FACULTY OF COMPUTER
SCIENCE AND ENGINEERING**



MultipleYE Final Conference in Skopje, North Macedonia, 20-21 August 2026



Conference Venue Faculty of Computer Science and Engineering
Location ul. Rudzer Boshkovikj 16 <https://maps.app.goo.gl/DnSbMQ6P3DLkZ9Ph8>

Conference Room FINKI Amphitheater [ФИНКИ / FINKI - 360° VR Tour](#)
Although unmarked, it is next to the FINKY Dean's office (left). See the room in the main menu → Amphitheatres → FINKI Amphitheater

Official Program

20 Aug

8:30 - 9:00 Registration and Coffee

9:00-9:30 Opening Remarks (Local Organizer & MultipleYE chairs)

9:30 - 10:30 Invited Talk 1

Elizabeth R. Schotter: **Reading efficiency: Optimal coordination between eye movements and distributed language processing**

10:30 - 11:00 Coffee Break

11:00 - 12:15 MultipleYE Session 1

Lena Ann Jäger: **Introduction to the MultipleYE data collection**

Oleksandra Kuvshynova & Mircea Marin: **Multilingual annotation of the MultipleYE corpus texts**

Maja Stegenwallner-Schütz: **The MultipleYE Assessment of Text Comprehension during Natural Reading**

12:15 - 13:00 Short Talks

12:15 Cui Ding, Ethan Gottlieb Wilcox, Lena Ann Jäger: **Individual differences of spillover in naturalistic reading**

12:30 Li Kloostera, Rick Nouwen, Jakub Dotlacil: **Coherence illusions in discourse processing: eye-tracking and acceptability studies on presupposition resolution with 'again'**

12:45 Candace M. van der Stelt, Elise Bossenbroek, Kelly Knollman-Porter, Karen Hux, Sarah E. Wallace: **Effects of text simplification on reading comprehension accuracy, perceived difficulty, and reading fluency in post-stroke aphasia**



13:00 - 14:30 **Lunch**

14:30 - 15:30 **MultiplEYE Session 2**

Eva Pavlinušić Vilus: **The MultiplEYE Psychometric Test Battery**

Carlson Büth: **The MultiplEYE Pipeline: An Open Source, Adaptable Eye-Tracking-while-Reading Preprocessing Pipeline Including a Tutorial**

Marie-Luise Müller: **EyeStore - Facilitating Discovery and Reuse for Eye-Tracking-While-Reading Data**

15:30 - 16:30 **Invited Talk 2**

Noam Siegelman: **Reading as statistical learning: New insights from the Multilingual Eye-Movement Corpus (MECO)**

16:30 - 17:30 **Poster Session and Coffee Break (7 posters)**

19:00 **Social Event / Dinner**

Restaurant Lira

Location: Ul. Nikola Tesla 11 <https://maps.app.goo.gl/mG1CxVxts1vmDeT76>

21 Aug

8:30 - 9:00 **Registration and Coffee**

9:00 - 10:00 **Invited Talk 3**

Ekta Sood: **From human gaze to behavior-grounded AI**

10:00 - 10:45 **Short Talks**

10:00 Daniel G. Krakowczyk, Carlson Moses Büth, David Robert Reich, Deborah Noemie Jakobi, Paul Prasse, Andreas Säuberli, Anastassia Shaitarova, Lena Ann Jäger: **pymovements: Building blocks for reproducible eye-tracking pipelines**

10:15 Reinhold Kliegl, Maximilian M. Rabe: **Integrating statistical and computational modeling of eye movement control in reading with a linked mixture model**

10:30 Anna Bondar, David Robert Reich, Lena Ann Jäger: **Mixture of Readers (MoRe): Enabling scanpath generation for different reader groups**

10:45 - 11:45 **Poster Session and Coffee Break (7 posters)**

11:45 - 12:30 **Short Talks**

11:45 Annika Kängsepp, Maarja-Liisa Pilvik, Kaidi Lõo, Külli Habicht, Liina Lindström: **Reading standard and nonstandard parallel forms in Estonian**

12:00 Johanne S. K. Nedergaard, Solvej W. Jaeger, Ivalo I. Mathiassen: **Tracking reading in Kalaallisut: The morpheme as the basic unit of analysis**

12:15 Maroš Filip, Michaela Svoboda, Kateřina Chládková: **Eye-tracking of sign language: the effect of learning experience**

12:30 - 14:00 **Lunch Break**



14:00 - 15:00 **Invited Talk 4**

Jochen Laubrock: **The predictive reader: What large language models reveal about reading across languages and development**

15:00 - 16:00 **Poster Session and Coffee Break (7 posters)**

16:00 - 16:30 **Closing Remarks**

Poster sessions:

Day 1: 16:30 - 17:30 **Poster Session and Coffee Break (6 posters)**

1. Kristina Cergol, Ana Matić, Eva Pavlinušić Vilus, Marijan Palmović: **Looking ahead in L1 and L2**
2. Jana Mara Hofmann, Deborah Noemie Jakobi, David Robert Reich, Paul Prasse, Lena Sophia Bolliger, Lena Ann Jäger: **A living overview of eye-tracking-while-reading datasets with Open library support**
3. Jan Brasser, Vera Demberg, Lena Ann Jäger, Maja Stegenwallner-Schütz: **Which reading measures are the strongest predictors of reading comprehension?**
4. Daniele P. Radicioni, Patrizia Paggio, Alessandro Corona Mendoza, Matteo Delsanto, Alberto Parola, Viviana Patti, Calogero Jerik Scozzaro: **Genre effects on eye movements in Italian reading: Evidence from the MultiPLEYE IT dataset**
5. Adrielli Tina Lopes Rego, Joshua Snell, Martijn Meeter: **Capturing comprehension in the reading brain: Effects from language model-derived propositions on eye movements**
6. Thyra Krosness, Not Battesta Soliva, Lena Ann Jäger: **From the Alps to the Diaspora: A work-in-progress on reading profiles across Swiss minority and heritage languages**

Day 2: 10:45 - 11:45 **Poster Session and Coffee Break (6 posters)**

8. Cengiz Acarturk, Anna Wilkosz, Melike Çağlayan, Jakub Jan Romanowski: **Brain activity correlations in reading: An ongoing study with fNIRS and eye-tracking coregistration**
9. Hanne B. Søndergaard Knudsen, Kristine Jensen de Lopez, Nik Kharlamov: **Reading comprehension, executive functions and vocabulary in Danish adult L1 speakers**
10. Lamija Zahirović, Lea Popović, Anastasija Žarevac, Eva Viktorija Zajc, Katarina Lasković, Kristian Elersič, Anja Podlesek: **Working memory capacity and eye movement patterns in reading: A correlational analysis of regressions in the Slovenian MultiPLEYE dataset**
11. Onur Keleş, Sezen Bektaş Yüksel, Servet Arda Kayhan, Feyza Budan, İris Beyza Deveci, Nazik Dinctopal Deniz: **Effects of context length and text genre on Turkish language models' prediction of human reading: A lossy-context surprisal analysis**
12. Rares-Alexandru Roscan, Oleksandra Kuvshynova, Mircea-Mihai Marin, Sergiu Nisioi: **Investigating the effect of text simplification on reading comprehension in advanced L2 English speakers**
13. Daria Horga, Antonia-Maria Popeanga: **Eye-movement patterns in Romanian text reading: A linguistic perspective**

Day 2: 15:00 - 16:00 **Poster Session and Coffee Break (7 posters)**



15. Thyra Krosness, Annette Lichtenauer, Claire Reymond, Lena Ann Jäger, Simone Girard-Groeber: **A work-in-progress towards guidelines for designing Easy Language multiple choice reading comprehension items**
16. Qingyang Tang, Bonnie Wing-Yin Chow, Alex Perovic: **Irony processing in bilingual adults with autism: Exploring the pragmatic-cognitive interface through eye-tracking**
17. Hatice Ozata, Filiz Rızaoğlu: **Reading goals in eye-tracking research: A comparative analysis of purpose-driven processing in L1 and L2 readers**
18. Cengiz Acarturk, ESMEHAN ÖZER, Belgin Üstün Güllü, Halime Betül Kırmızı, Kardelen Çiftci: **Eye movements and reading comprehension in neurodivergent readers: A study on children with formal ADHD diagnosis**
19. Akshay Mendhakar, Helmut Leder: **Evatrack: a flexible framework for standardised eye-tracking research in reading research**
20. Isabelle Caroline Rose Cretton, David Robert Reich, Lena Ann Jäger: **Surprisal by design: A principled pipeline for human-LLM stimulus pairs in eye-tracking reading experiments**
21. Jurgita Motiejūnienė, Ramunė Kasperė, Vilma Linkevičiūtė, Valdas Grigaliunas, Paulius Budreika: **Methodological and accessibility challenges in eye-tracking research with Deaf and hard of hearing**



Organized and funded by COST Action CA21131 *Enabling multilingual eye-tracking data collection for human and machine language processing research*



Invited talks

Invited Talk 1

Reading efficiency: Optimal coordination between eye movements and distributed language processing

Elizabeth R. Schotter
University of South Florida

To read successfully, a person must visually perceive words, allocate attention to recognize their meanings, and plan eye movements to the next location in the text. Reading *efficiently* (quickly with good understanding) requires coordinating these processes adaptively in real time, and processing information at central fixation (i.e., in foveal vision) as well as surrounding areas (i.e., in parafoveal vision). In this talk, I will summarize what we have learned about reading from studies using eye tracking paradigms combined with visual manipulations of the text (i.e., gaze-contingent moving window/mask studies). For the most part, readers prioritize foveal processing for word recognition and comprehension, and use parafoveal processing to support reading speed. I will also describe the additional insights we have gained from studies of deaf readers, who have been shown to read more efficiently than skill-matched hearing readers, because they distribute visual attention more broadly across the text than their hearing peers, and use parafoveal information to both support reading speed and comprehension. These findings highlight the flexibility of the reading system and its adaptiveness to different sensory and linguistic experiences that precede learning to read.

Invited Talk 2

Reading as statistical learning: new insights from the Multilingual Eye-Movement Corpus (MECO)

Noam Siegelman
Hebrew University of Jerusalem

In recent years, the study of reading has become increasingly grounded in theories of statistical learning. This shift has led to a reconceptualization of reading acquisition as “an exercise in statistical learning,” whereby learners become attuned to the statistical structure of their writing system. In this talk, I first review this theoretical framework and its core empirical findings, which to date have largely focused on single-word recognition. I then present a series of recent studies examining how naturalistic reading behavior, as measured via eye movements, is shaped by the statistical properties of different writing systems. These studies draw on the Multilingual Eye-Movement Corpus (MECO), a large-scale dataset comprising over 1,200 readers across dozens of sites worldwide. Across studies, I highlight three key insights: (1) writing systems encompass an array of rich, quasi-regular statistical structures; (2) readers engage in computations that are more complex and nuanced than previously documented; and (3) individual differences in reading are tied to a sophisticated sensitivity to the structure of one’s writing system. I conclude by arguing that reading constitutes an ecological “statistical learning experiment,” offering a powerful lens for constraining and refining theories of statistical learning more broadly.



Invited Talk 3 **From human gaze to behavior-grounded AI**

Ekta Sood
University of Colorado Boulder

Human gaze and eye-tracking data have been central to psycholinguistic research on how people read, allocate attention, engage, and reason with language. This raises a natural question: can these behavioral signals be used not only to study human language processing, but also to build systems that better support language-based human-AI interaction? Despite this potential, using gaze in modern natural language processing (NLP) remains challenging because eye-tracking data is scarce, gaze varies across tasks and domains, and human attention does not always align cleanly with neural attention, feature attribution, or model performance. In this talk, I will present three lines of work within this research theme: (1) comparing human gaze with neural attention in machine reading comprehension and multimodal visual question answering; (2) using computational models of reading and human eye-tracking data to generate gaze-based signals that improve NLP tasks under data scarcity; and (3) learning gaze embeddings through self-supervised learning and multimodal gaze-text alignment to support cognitive-state prediction and future behavior-grounded systems for human-AI teaming.

Invited Talk 4 **The predictive reader: what large language models reveal about reading across languages and development**

Jochen Laubrock
University of Potsdam

The human reading system operates predictively, continuously anticipating upcoming input to guide eye movements and facilitate word recognition. Cloze predictability is the standard word expectancy measure, but collecting cloze norms is expensive and therefore limited to a small number of curated stimulus sets. We argue that surprisal estimates derived from large language models (LLMs) trained on next-token prediction provide a scalable alternative. Across multiple corpora and languages, LLM-derived predictability correlates strongly ($r \approx .70$) with cloze norms and emerges as a strong linguistic predictor of fixation durations, surpassing word frequency and word length (Chandra et al., 2023). We present converging evidence from several lines of research. First, a cross-linguistic extension to Chinese demonstrates that the granularity of tokenization is critical, because Chinese characters correspond to morphemes (Yan and Laubrock, 2026). Second, re-analyses of data from our longitudinal study of perceptual span development reveal that the use of predictive information during reading increases with developing expertise (Meixner et al., 2022; Meixner and Laubrock, 2024). Third, gaze-contingent boundary experiments crossing LLM-derived predictability with preview type (identical, semantically related, nonword) show that predictability amplifies the identical-preview benefit in early eye-movement measures. We also replicate semantic preview benefit in earlier measures in German across several experiments. Together, these findings establish LLM predictability as a scalable tool for multilingual reading research and illuminate the cognitive mechanisms by which prediction influences the reader's eye movements.



Abstracts

MultiPEYE Corpus Session 1

The MultiPEYE Eyetracking-While-Reading Corpus

Lena Jäger^{*1,2}

¹Department of Computational Linguistics Department of Informatics, University of Zurich

²Department of Informatics, University of Zurich

*contact: lenaann.jaeger@uzh.ch

This presentation will provide an overview of the MultiPEYE Corpus (Jäger *et al.*, 2026; Jakobi *et al.*, 2025), collected within the MultiPEYE COST Action. The corpus is a multilingual eye-tracking-while-reading dataset based on parallel stimulus texts representing four text types, namely argumentative, institutional, literary, and popular-science writing. This design enables systematic comparisons of reading behaviour across languages, scripts and text types.

Each text is followed by a reading-comprehension assessment targeting different levels of comprehension. The assessment was developed through an iterative process involving both in-lab and online data collection, followed by Item-Response Theory analyses. The eye-tracking data are complemented by extensive sociolinguistic and demographic information and a comprehensive psychometric test battery covering constructs central to psycholinguistic research, including verbal and non-verbal working memory and cognitive control, vocabulary knowledge, language aptitude, and automatized phonological retrieval.

Data collection follows a standardized experimental setup and procedure across all participating sites, including systematic metadata documentation and data-quality reporting. For data preprocessing, we developed a fully transparent and reproducible open-source pipeline in accordance with the FAIR (Findability, Accessibility, Interoperability, Reusability) principles (Wilkinson *et al.*, 2016). Alongside the corpus, we release general-purpose open-source software packages for administering eye-tracking-while-reading experiments and psychometric tests, as well as for processing and analysing the resulting data. Although developed in the context of the MultiPEYE Corpus, these tools are designed to be used in reading research more broadly.

Data collection is currently under way for 35 languages from different language families, with a target sample of 100 native speakers per language. Collection has already been completed for 14 languages. The data will be made available through a filterable user interface, with Zenodo serving as the archival backend. Data collection will continue beyond the initial release, allowing further languages to be added to the corpus over time.

References

- Jäger L. A., Hollenstein N., Matić Škorić A., Jakobi D. N., Stegenwallner-Schütz M., Ding C., Pavlinušić Vilus E., Kasperé R., Müller M.-L. (2026). *MultiPEYE: Enabling multilingual eye-tracking data collection for human and machine language processing research*. PsychArchives. Preregistration. [doi:10.23668/psycharchives.21607](https://doi.org/10.23668/psycharchives.21607).
- Jakobi D. N., Stegenwallner-Schütz M., Hollenstein N., Ding C., Kasperé R., Matić Škorić A., Pavlinusic Vilus E., Frank S., Müller M.-L., Jensen de López K. M., Kharlamov N., B. Søndergaard Knudsen H., Berzak Y., Lion E., Sekerina I. A., Acarturk C., Ansari M. F., Harezlak K., Kasprowski P., Bautista A., Beinborn L., Bondar A., Boznou A., Bradshaw L., Hofmann J. M., Krosness T., Soliva N. B., Çepani A., Cergol K., Došen A., Palmovic M., Çerpja A., Chirino D., Chromý J., Demberg V., Škrjanec I., Deniz N. D., Fajardo D. I., Giménez-Salvador M., Minguez-López X., Filip M., Freibergs Z., Gomes J., Janeiro A., Luegi P., Verissimo J., Gramatikov S., Hasenäcker J., Haveriku A., Kote N., Kamal M. M., Kunlineddzierska H., Klimek-Jankowska D., Kosutar S., Krakowczyk D. G., Krejtz I., Łockiewicz M., Lõo K., Motiejūnienundefined J., Nasir J. A., Nedergård J. S. K., Özkan A., Preininger M., Pungă L., Reich D. R., Tschirner C., Rot , Säuberli A., Solé-Casals J., Strati E., Svoboda I., Trandafil E., Varlokosta S., Vulchanova M., Jäger L. A. (2025). MultiPEYE: Creating a multilingual eye-tracking-while-reading corpus. In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications*, ETRA '25, pp. 1–11. Association for Computing Machinery. [doi:10.1145/3715669.3726843](https://doi.org/10.1145/3715669.3726843).
- Wilkinson M. D., Dumontier M., Aalbersberg I. J., Appleton G., Axton M., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3(1), 160018. [doi:https://doi.org/10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).



Abstracts

u p Corpus S ss o

The MultiplEYE Pipeline: An Open Source, Adaptable Eye-Tracking-while-Reading Preprocessing Pipeline Including a Tutorial

Deborah N. Jakobi¹, Carlson Bith¹, Jana Hofmann¹, Daniel Krakowczyk³, Anastassia Shaitarova¹, Andreas Säuberli¹, Antonia Popp¹, Eva Pavlinušić Vilus¹, Lena A. Jäger^{1,2}

¹Department of Computational Linguistics, University of Zürich

²Institute of Informatics, University of Zürich

³Institute of Computer Science, University of Potsdam

*contact: deborahnoemie.jakobi@uzh.ch

Eye-tracking-while-reading data are a commonly used resource across different fields (e.g, Mézière *et al.*, 2023; Reich *et al.*, 2025). Despite many use cases there is a lack of transparency regarding many aspects in the life cycle of these data. This lack is particularly severe as eye-tracking data are not used as-is but undergo several preprocessing steps. The steps range from parsing the collected gaze samples from the original data file, detecting gaze events such as fixations and saccades, or aggregating fixation data further into word-level measures. None of these steps are trivial and highly depend on the exact characteristics of data (single-line sentences, multi-page, different writing systems, etc.) and its use cases (single word analyses, inference at the text level, etc.). Researchers need to make many decisions regarding the applied methods, its parameters, or the exact implementation of each of the preprocessing steps which highly impact the resulting data and its analyses. This lack of standards and incompletely shared information leads to preprocessed data which cannot easily be reused, undiscovered errors in preprocessing implementations as it cannot be checked by other researchers, and eventually unreproducible analyses.

As a response to this gap, we present the MultiplEYE pipeline which is an open-source preprocessing pipeline in Python. While initially developed for the MultiplEYE data, the code is accompanied by an extensive tutorial to explain the preprocessing and make all decisions entirely understandable and transparent. The pipeline covers data preparation, preprocessing steps such as cleaning, event detection, AOI mapping, and scanpath generation, as well as the integration of questionnaires, psychometric tests, and metadata. Furthermore, the code is developed in such a way that it serves as an example for preprocessing similar data. This is partially due to the fact that the implementation is built on the existing eye-tracking preprocessing Python package *pymovements* (Krakowczyk *et al.*, 2023), and partially due to the pipeline being structured into easy to understand components which tackle one task of the preprocessing pipeline at a time and integrate the respective *pymovements* functions in a clearly documented way. A review app additionally supports session-level data inspection and quality control.

In sum, the pipeline makes two important contributions to the field of eye-tracking-while-reading research. First, it presents a transparent, well-documented implementation for a huge, multilingual, multi-script, multi-device dataset which serves as an example for any preprocessing pipeline of the same kind. Second, it includes an extensive step-by-step tutorial to go through the preprocessing pipeline which not only increases knowledge about often neglected and unpublished preprocessing decisions, but also sets the basis for developing the pipeline further in the same style for a custom dataset.

References

- 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 2023, pp. 1–2. Association for Computing Machinery.
- Mézière D. C., Yu L., Reichle E. D., von der Malsburg T., McArthur G. (2023). Using eye-tracking measures to predict reading comprehension. *Reading Research Quarterly* 58(3), 425–449. doi:10.1002/rrq.498.
- Reich D. R., Shubi O., Jäger L., Berzak Y. (2025). *ACL 2025 tutorial: Eye tracking and NLP*. URL: <https://acl2025-eyetracking-and-nlp.github.io/>.



Abstracts

MultiPLEYE Corpus Session 2

EyeStore - Facilitating Discovery and Reuse for Eye-Tracking-While-Reading Data

Marie-Luise Müller¹, Gorka Fraga Gonzalez², Anna Bondar², Cui Ding², Nora Hollenstein², Deborah N. Jakobi², Ana Matic³, Maja Stegenwallner-Schütz⁴, Beliz Sertcan Gökmen², Eva Pavlinušić Vilus², Lena A. Jäger²

¹Archiving and Publication Services, Leibniz Institute for Psychology (ZPID)

²Department of Computational Linguistics, University of Zurich

³Faculty of Education and Rehabilitation Sciences, University of Zagreb

⁴Department of Special Education/Language Development, University of Koblenz

*contact: mm@leibniz-psychology.org

Eye-tracking is a central method for studying reading and language processing, yet sharing and reusing eye-tracking-while-reading data remains challenging. Datasets are often stored in general-purpose repositories with heterogeneous metadata and limited possibilities for systematic discovery and comparison (Godwin *et al.*, 2025; Jakobi *et al.*, 2026). This presentation introduces *EyeStore*, discovery platform and interface for eye-tracking-while-reading datasets that follows the FAIR (Findability, Accessibility, Interoperability, Reusability) principles (Wilkinson *et al.*, 2016). Implemented by the Language & Information Research Lab (LiRI) at the University of Zurich, EyeStore provides structured metadata and tools for searching, filtering, and exploring datasets according to relevant experimental, linguistic, and technical characteristics. The current implementation supports the multilingual corpus developed within the MultiPLEYE COST Action ^{1,2}, with datasets deposited in the MultiPLEYE community on Zenodo. We discuss how EyeStore complements established repositories by making eye-tracking data easier to discover, explore, and reuse.

References

- Godwin H. J., Dewis H., Darch P. T., Hout M. C., Ernst D., Broadbent P., Papesh M., Wolfe J. M. (2025). A sharing practices review of the visual search and eye movements literature reveals recommendations for our field and others. *Behavior Research Methods* 57(9), 235–. URL: <https://doi.org/10.3758/s13428-025-02759-3>, doi:10.3758/s13428-025-02759-3.
- Jakobi D. N., Reich D. R., Prasse P., Hofmann J. M., Bolliger L. S., Jäger L. A. (2026). *Eye-Tracking-while-Reading: A Living Survey of Datasets with Open Library Support*. URL: <https://arxiv.org/abs/2602.19598>, arXiv:2602.19598.
- Wilkinson M. D., Dumontier M., Aalbersberg I. J., Appleton G., Axton M., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3(1), 160018. doi:<https://doi.org/10.1038/sdata.2016.18>.

¹<https://multipleye.eu>

²www.cost.eu/actions/CA21131/



Abstracts

MultiplEYE Corpus Session 2

The MultiplEYE Assessment of Text Comprehension during Natural Reading

Maja Stegenwallner-Schütz^{*1}

¹University of Koblenz

*contact: stegenwa@uni-koblenz.de

Eye-tracking-while-reading research has contributed substantially to our understanding of how readers process language in text, offering insights into reading as it unfolds. Relating online processing to the outcome of comprehension typically requires a comprehension assessment. Inspired by reading assessments that conceptualize literacy as a broad ability that encompasses the use of reading across contexts (Gehrer *et al.*, 2013), we developed a reading assessment based on the MultiplEYE Text Corpus. It comprises comprehension questions that probe multiple levels of comprehension, spanning levels of text representation from the surface textbase to the full situation model (Kintsch, 1998).

The MultiplEYE reading assessment comprises ten texts representing different text types, each followed by six comprehension questions: two local, two bridging, and two global questions (see Table 1). It is administered in a computerized format in which readers cannot return to previously read pages and questions are presented only after completion of the text. The assessment therefore targets *learning from the text*, as readers cannot preview the questions or use them to guide their reading. The comprehension questions are preceded by ratings of text familiarity and difficulty on five-point Likert scales. Each question is presented with four plausible answer options in randomized order. One option is correct. The three distractors are systematically constructed according to their textual support, with the relevant text spans annotated following the STARC scheme (Berzak *et al.*, 2020). In the local and bridging questions, distractor A reflects a plausible misunderstanding of the *target span*, distractor B draws on information from a different part of the text, annotated as the *distractor span*, and distractor C has no textual support, serving as a check on whether participants engaged with the text. For global questions, distractors A and B are drawn from within the target span, which encompasses the entire text.

The assessment underwent rigorous iterative validation across multiple experiments. Initial experiments evaluated the plausibility and textual support of answers, followed by multiple rounds of online testing with and without the accompanying text. These studies evaluated the construction of distractors and used Item Response Theory to calibrate item difficulty and assess measurement properties. Items not meeting the standards were revised and retested. Only validated items were retained in the final assessment.

Crucially, the assessment enables researchers to directly link eye movements during natural reading in the *MultiplEYE eye movement corpus* to individual differences in comprehension. Comprehension scores provide a measure of reading ability and can be related to fine-grained eye-movement behavior on text spans relevant to comprehension accuracy, including spans associated with questions of varying difficulty. The questions have been translated in parallel across MultiplEYE languages, creating a multilingual reading assessment that allows researchers to study how reading behavior predicts comprehension across languages, readers, and text types.

Table 1: Question types and characteristics

Question Type	Cognitive Process	Textual Support	Literalness
Local	Information retrieval	Up to two clauses	Literal
Bridging	Information integration	Several, not necessarily consecutive, clauses	Non-literal
Global	Reflection	Entire text	Non-literal

References

- Berzak Y., Malmaud J., Levy R. (2020). STARC: Structured annotations for reading comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics.
- Gehrer K., Zimmermann S., Artelt C., Weinert S. (2013). NEPS framework for assessing reading competence and results from an adult pilot study. *Journal for Educational Research Online* 5(2), 50–79.
- Kintsch W. (1998). *Comprehension: A paradigm of cognition*. Cambridge University Press.



Abstracts

MultiPLEYE Corpus Session 2

MULTIPLEYE PSYCHOMETRIC TEST BATTERY

Eva Pavlinušić Vilus^{*1} and Ana Matic^{1,2}

¹Department of Computational Linguistics, University of Zurich

²Department of Speech and Language Pathology, University of Zagreb

*contact: eva.pavlinusicvilus@uzh.ch

The MultiPLEYE psychometric test battery is a digital, fully automated, multilingual, and openly available tool designed to assess cognitive abilities relevant to research on reading and language processing. The battery includes measures of working memory capacity, rapid automatized naming, cognitive control, language aptitude, and vocabulary knowledge. All tasks are administered through a user-friendly interface in a standalone Python environment. The standardized procedure minimizes experimenter-related variability, while responses, reaction times, and audio recordings are automatically saved to local output files.

The implementation is accompanied by a standardized data collection protocol and an automated processing pipeline that generates participant-level scores for each test. So far, materials have been prepared for 28 languages. Psychometric data are currently being collected in more than 20 languages, with the aim of evaluating the psychometric properties of the battery. In this talk, we present the design and implementation of the MultiPLEYE psychometric test battery, the accompanying data collection and processing infrastructure, and its potential as a standardized resource for cross-linguistic research on cognitive factors underlying reading and language processing.



Individual differences of spillover in naturalistic reading

Cui Ding¹, Ethan Gottlieb Wilcox², Lena Ann Jäger¹

¹Department of Computational Linguistics, University of Zurich, ²Department of Linguistics, Georgetown University, Washington, DC

*contact: cui.ding@uzh.com

Background: Spillover effect happens when the processing difficulty of one word affects reading times on the following, but its causes and stability across individuals are still unknown. Classic reading models, such as E-Z Reader and SWIFT, link spillover to oculomotor planning or parallel processing, assuming it remains stable across individuals. Sentence processing models, e.g., surprisal theory, often treat it as a nuisance. Some recent studies suggest that spillover may reflect deeper cognitive processes, but it is still unclear whether it reflects a stable reader trait, a temporary state, or an interaction of both. This project investigates the stability of individual spillover effects across reading contexts, examines their modulation by cognitive abilities, and tests whether they are associated with overall reading patterns.

Method: We used six eye-tracking datasets that allow us to compare each participant's spillover effects across different reading contexts. In InDiCo (German) (Haller, 2026), participants read in two sessions, with a two-week interval between them. In PoTeC (German) (Jakobi, 2025), participants read both within and outside their expertise domain. In OneStop (English) (Berzak, 2025), participants did both normal naturalistic reading and repeated reading. In GECO-NL (Cop, 2017) and GECO-ZH (Sui, 2023), Dutch and Chinese participants read in both their native language (L1) and English (L2). In HKC (Chinese) (Wu and Kit, 2023), participants read both single isolated sentences and full paragraphs. For each participant, we estimated spillover effects of word length, lexical frequency, surprisal, and their stability using a Bayesian two-task hierarchical model (Rouder and Haaf, 2019). We then tested how these spillover effects are modulated by cognitive abilities and related to skipping and regression reading patterns.

Results & Discussion: Spillover effects showed different levels of stability across reading contexts. Word length spillover was the most stable, while frequency and surprisal spillover were less consistent. People showed different spillover patterns depending on the reading contexts—they did not adopt the same strategy across situations. Spillover effects were more stable across sessions, domains, and when the L1 and L2 were similar (e.g., Dutch–English), but less stable across distant languages (e.g., Chinese–English), reading regimes, and context lengths. Spillover effects are not uniform across readers and are modulated by reading fluency. Low-fluency readers showed resource-driven spillover, with stronger effects of low-level difficulty such as word length and frequency. In contrast, high-fluency readers showed integrative spillover, driven mainly by surprisal. These individual differences were systematically related to the overall reading patterns: greater sensitivity to previous word length and frequency was associated with less skipping and more regressions, while greater sensitivity to previous word surprisal was associated with more skipping and fewer regressions. These results highlight the importance of individual difference approaches for understanding spillover and challenge population-level analyses that obscure meaningful heterogeneity in reading.

References

- Berzak Y. e. a. (2025). OneStop: A 360-participant english eye-tracking dataset. *Scientific Data* 12, 1995.
Cop U. e. a. (2017). Presenting GECO. *Behavior Research Methods* 49, 602–615.
Haller P. e. a. (2026). Replicate me if you can. *Cognitive Science* 50, e70121.
Jakobi D. N. e. a. (2025). PoTeC: A german eye-tracking-while-reading corpus. *Behavior Research Methods* 57, 211–248.
Rouder J. N., Haaf J. M. (2019). Psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review* 26, 452–467.
Sui L. e. a. (2023). GECO-CN: Chinese-English bilingual sentence reading. *Behavior Research Methods* 55, 2743–2763.
WuY., Kit C. (2023). Hong Kong corpus of Chinese sentence and passage reading. *Scientific Data* 10, 899.



Coherence illusions in discourse processing: eye-tracking and acceptability studies on presupposition resolution with 'again'

Li Kloostra, Rick Nouwen, Jakub Dotlacil

Department of Language, Literature and Communication, Institute for Language Sciences

*contact: l.kloostra@uu.nl

Background: Language processing frequently requires the resolution of dependency relations between different linguistic items. Memory research has primarily focused on syntactic dependencies, where matching distractors can elicit "grammaticality illusions" and facilitate the processing of ungrammatical structures (Wagers *et al.*, 2009; Phillips *et al.*, 2011). Both misrepresentation accounts, e.g., Marking and Morphing (Eberhard *et al.*, 2005), and misretrieval accounts, e.g., Cue-Based Retrieval (Lewis and Vasishth, 2005), predict this facilitatory effect in ungrammatical contexts; however, they diverge in their predictions for grammatical contexts. To date, the empirical evidence for these grammatical contexts remain weak or inconsistent. Moreover, little is known about interference effects beyond the syntax domain. To address this gap, our study investigates whether interference effects and "coherence illusions", comparable to those found in syntax, arise during dependency resolution at the discourse level. In particular, we focus on presupposition resolution triggered by 'again', following recent research on presuppositions with 'too' (Schmitz *et al.*, 2025).

Method: We conducted two eye-tracking studies, along with acceptability studies for each, on short Dutch discourses. Each discourse described two events that could either match or mismatch the presupposition of 'again'; one of them (distractor) appeared in the scope of negation (EX1) or linked to a different discourse referent (EX2) making that event inappropriate for resolution.

Results & Discussion: Bayesian analyses of both eye-tracking data and offline acceptability revealed that when the target mismatched (incoherent discourse), matching distractors facilitated processing and increased acceptability. These results are in line with predictions from both Marking and Morphing and Cue-Based Retrieval and favor the notion of 'coherence illusions' in discourse. In target-match (coherent) conditions, the pattern is more complex: both studies showed a similar speed-up effect of matching distractors. This favors the MM account, although it would have to be generalized to discourse structures. The offline acceptability data aligned mostly with CBR, though comprehension data (EX2) suggest that stylistic preference may have been a confound. Critically, both eye-tracking experiments revealed a novel negative interaction in regression measures: readers regressed more whenever only the target or only the distractor matched the presupposition, regardless of coherence. Speeded acceptability and comprehension data of EX2 further suggested that facilitation effects were driven by trials where readers confused discourse referents. We interpret this as evidence for failed binding in memory: when two events share features but differ in referent or polarity, readers lose track, triggering confusion. Our results demonstrate that discourse-level representations are susceptible to interference effects, coherence illusions, and the binding problem. Moreover, we show that presupposition resolution is a productive domain for studying memory in language and that discourse processing can be used as a means to advance memory models beyond syntax.

References

- Eberhard K. M., Cutting J. C., Bock K. (2005). Making syntax of sense: number agreement in sentence production. *Psychological review* 112(3), 531.
- Lewis R. L., Vasishth S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29, 375–419.
- Phillips C., Wagers M. W., Lau E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. *Experiments at the Interfaces* 37, 147–180.
- Schmitz T., Nouwen R., Dotlacil J. (2025). Memory retrieval in discourse: Illusions of coherence during presupposition resolution. *Journal of Memory and Language* 143, 104637.
- Wagers M. W., Lau E. F., Phillips C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of memory and language* 61(2), 206–237.



Effects of text simplification on reading comprehension accuracy, perceived difficulty, and reading fluency in post-stroke aphasia

Candace M. van der Stelt¹, Elise Bossenbroek², Kelly Knollman-Porter², Karen Hux³, Sarah E. Wallace¹

¹Department of Communication Science and Disorders, University of Pittsburgh, ²Department of Speech Pathology and Audiology, Miami University, ³Psychology Department, Quality Living, Inc.

*contact: cav122@pitt.edu

Background: Acquired reading impairments can limit quality of life following stroke (Knollman-Porter et al., 2015). Text simplification is a possible mechanism to increase access to and independence with written materials for people with aphasia (PWA). Shortening words and sentences is one approach to text simplification and is measurable through Flesch-Kincaid grades levels (FKGLs; Kincaid et al., 1975). Whether PWA benefit from passages with low FKGLs is unknown. This study investigated the effect of four FKGLs on reading comprehension accuracy, perceived difficulty, and reading fluency (i.e., speed and eye-tracking metrics) in PWA compared to demographically matched neurotypical adults (NAs).

Method: Twenty PWA and 20 NAs read 20 passages, five at each of four distinct FKGLs (1st, 4th, 8th, and 12th). Number of words between passages were comparable (70-77 each).

Participants read each passage one time, answered five comprehension questions, and rated perceived difficulty. A Tobii Dynavox® Pro Spectrum eye tracker captured participants' eye-movements. To measure reading fluency, we calculated the total reading time and extracted eye-fixation metrics (fixation probability, number of fixations, number of visits, and average duration of visits). We analyzed each outcome measure in separate mixed effects models, followed by within- and between-group simple effects analyses using Tukey's HSD to correct for multiple comparisons.

Results & Discussion: Comprehension Accuracy: A significant main effect of group ($p < 0.001$) indicated lower odds of correct responses by PWA vs. NAs. No significant effects of FKGLs emerged for either group. Perceived Difficulty: A significant main effect of group showed that PWA perceived passages to be more difficult than did NAs ($p < .001$). Both groups rated 1st, 4th, and 8th grades as less difficult than 12th grade. No significant differences were observed between ratings for the lower three grades. Reading Fluency: First and 4th FKGLs reduced reading time for both groups (all $p < .001$). PWA read more slowly than NAs at all FKGLs (all $p < .001$). No significant group x grade-level interactions indicated that text simplification did not close the reading-time gap between groups. PWA had higher fixation probability, more fixations and visits, and longer visit durations than NAs across all grade-levels (all $p < .001$). Text simplification reduced number of fixations and average duration of visits for both groups, while fixation probability only decreased for NAs. Number of visits remained relatively consistent across grade-levels for both groups.

Text simplification measured via the FK formula provided some benefit to PWA, specifically by reducing perceived difficulty and reading time. This reflects fewer fixations and shorter visits per word, likely due to shorter words in lower FKGLs (Kliegl et al., 2004). However, lower FKGLs did not reduce group differences in accuracy, reading time, fixation probability, or number of visits. These results suggest that PWA fixate on and re-visit most passage words, regardless of FKGL, to maintain accuracy at the cost of reading time.

References

- Kincaid, J. P., Fishburne Jr., R. P., Richard, L. C., & Brad, S. (1975). *Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel.*
- Kliegl, R., Grabner, E., Rolfs, E., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, 16, 262–284
- Knollman-Porter, K., Wallace, S. E., Hux, K., Brown, J., & Long, C. (2015). Reading experiences and use of supports by people with chronic aphasia. *Aphasiology*, 29(12), 1448–1472.



pymovements: building blocks for reproducible eye-tracking pipelines

Daniel G. Krakowczyk,^{1,2} Carlson Moses Büth¹, David Robert Reich^{1,2}, Deborah N. Jakobi¹, Paul Prasse², Andreas Säuberli³, Anastassia Shaitarova¹, Lena Ann Jäger¹

¹Department of Computational Linguistics, University of Zurich, Switzerland, ²Department of Computer Science, University of Potsdam, Germany, ³Center for Information and Language Processing, LMU Munich, Germany

*contact: danielgeorg.krakowczyk@uzh.ch

pymovements is an open-source Python package for processing and analyzing eye-tracking data, designed to address fragmentation in preprocessing and analysis practices. Preprocessing pipelines are frequently implemented independently across laboratories, resulting in inconsistent filtering, event detection, and data handling procedures. This variability reduces reproducibility, complicates cross-study comparisons, and increases barriers for new researchers. *pymovements* mitigates these issues by providing a unified, transparent, and extensible framework that standardizes core workflows. The project has been in active development since 2022 and is closely linked to the MultipleYE COST Action as a foundation for its data preprocessing pipeline.

The package supports parsing of widely used eye-tracking formats, including EyeLink and BeGaze, and provides access to more than 35 publicly available eye-tracking datasets. It includes functionality for filtering and transformation, data quality assessment and data cleaning, as well as detection of oculomotor events such as fixations, saccades, and blinks. Together, these components establish a consistent basis for downstream analysis.

On top of these core primitives, *pymovements* provides tools for computing derived measures used in psycholinguistic and reading research, including stimulus-aligned analyses for textual data and standardized visualization tools such as heatmaps, scanpaths, and saccadic main-sequence plots, facilitating both exploratory data analysis and methodological evaluation.

The package is developed under continuous integration with automated testing, mandatory code review, and comprehensive documentation. It includes initial support for BIDS-compliant [Gorgolewski et al. \(2016\)](#) formats and adheres to FAIR principles [Wilkinson et al. \(2016\)](#), promoting findability, accessibility, interoperability, and reusability of eye-tracking data and analysis workflows. Released under the MIT License, *pymovements* is developed collaboratively by a distributed community of more than 45 contributors from multiple research laboratories. We actively encourage contributions that may take the form of documentation improvements, dataset integrations, bug reports, and feature enhancements.

Overall, *pymovements* provides a modular and extensible infrastructure for reproducible eye-movement re-search. By integrating essential components of eye-tracking analysis into a coherent and standardized framework, it reduces methodological heterogeneity, improves transparency, and supports robust, reproducible re-search practices across the field.

References

- Gorgolewski K. J., Auer T., Calhoun V. D., Craddock R. C., Das S., Duff E. P., Flandin G., Ghosh S. S., Glatard T., Halchenko Y. O., Handwerker D. A., Hanke M., Keator D., Li X., Michael Z., Maumet C., Nichols B. N., Nichols T. E., Pellman J., Poline J.-B., Rokem A., Schaefer G., Sochat V., Triplett W., Turner J. A., Varoquaux G., Poldrack R. A. (2016, June). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data* 3(1). doi:10.1038/sdata.2016.44.
- 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., Groth P., Goble C., Grethe J. S., Heringa J., 't Hoen P. A., 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, March). The fair guiding principles for scientific data management and stewardship. *Scientific Data* 3(1). doi:10.1038/sdata.2016.18



Integrating statistical and computational modeling of eye movement control in reading with a linked mixture model

Reinhold Kliegl¹, Maximilian M. Rabe²

¹Department of Sports and Health Sciences, University of Potsdam, ²Department of Psychology, University of Potsdam

*contact: reinhold.kliegl@uni-potsdam.de

Background: Research on eye-tracking-while-reading corpora oscillates between two paradigms: statistical modeling of aggregated summary statistics (e.g., go-past times) and computational modeling of fixation sequences (e.g., SWIFT, E-Z Reader). Both approaches possess limitations; statistical models often rely on the questionable assumption that every fixation is intentional, while computational models frequently impose rigid processing assumptions and ignore the hierarchical nature of participant data. As large-scale, multilingual datasets like the MultipleYE corpus emerge, there is a critical need for flexible modeling frameworks that can handle diverse linguistic structures without sacrificing statistical rigor.

Method: We present a highly extendable linked mixture model designed to bridge these traditions. Unlike traditional summary-statistic approaches, this model is fitted directly to raw fixation locations and durations. It accounts for stochasticity in target selection and saccadic error, similar to computational models, but avoids restrictive lexical processing assumptions. Crucially, the framework utilizes a hierarchical structure, allowing for highly customizable fixed and random effects. This makes it uniquely suited for large-scale corpora where individual differences and linguistic variability must be explicitly modeled.

Results & Discussion: Applying the model to the Potsdam Sentence Corpus, we already recovered some well-established oculomotor patterns, including fixation durations, skipping probabilities, and saccade amplitudes with the first attempt; there are also some puzzling results currently under investigation. Importantly, the model setup enabled the explicit estimation of how categorical and continuous factors (such as reader age) influence and co-vary across individuals.

The proposed linked mixture model offers a robust alternative for the analysis of reading data. In the context of the MultipleYE initiative, this methodology provides a pathway to move beyond “average” reading behaviors. By accounting for the interplay between saccadic execution and fixation durations within a unified statistical framework, this approach is particularly relevant for crosslinguistic comparisons where language-specific constraints may interact differently with the oculomotor system. We look forward to explore how this model can be applied to the newly released MultipleYE corpus to enhance our understanding of reading behavior across the 25 documented languages.



Mixture of Readers (MoRe): Enabling scanpath generation for different reader groups

Anna Bondar¹, David Robert Reich^{1,2}, Lena Ann Jäger¹

¹Department of Computational Linguistics, University of Zurich, ²Department of Computer Science, University of Potsdam

*contact: anna.bondar@uzh.ch

Eye movements in reading provide a rich behavioural signal that reflects moment-to-moment cognitive processing. Because reading behaviour depends not only on the text, but also on reader characteristics such as language background, reading proficiency, and cognitive profile, scanpaths vary systematically across reader groups. At the same time, current generative models of eye movements are often trained on heterogeneous populations with a single shared set of parameters. While such training can improve overall performance, it may also encourage the model to prioritise what is common across groups and to underrepresent patterns that are specific to individual populations, effectively producing an averaged scanpath. Modelling group-specific scanpaths is important both for basic research and for downstream applications: preserving population-specific variation can improve tasks that aim to infer reader characteristics from gaze data and can make synthetic scanpaths more useful as cognitively informative signals for NLP. In this work, we present Mixture of Readers (MoRe), an extension of the state-of-the-art scanpath generation model Eyettention (Deng *et al.*, 2023) for group-specific scanpath generation. For this we leverage a Mixture-of-Experts architecture (Shazeer *et al.*, 2017), originally proposed as a way of introducing sparse conditional computation into neural networks by replacing fully shared dense processing with a set of specialised experts, only some of which are activated for each input. We evaluate the model on CopCo (Hollenstein *et al.*, 2022), a Danish eye-tracking corpus comprising 58 readers and 2,032 sentence-level instances from three populations: L1 non-dyslexic readers, L1 readers with dyslexia, and L2 non-dyslexic readers. MoRe introduces reader-group-conditioned routing into different scanpath-processing components of Eyettention. In this way, rather than processing scanpaths from all reader groups with a single shared set of parameters, MoRe allows the model to select different weights depending on the reader group or on characteristics of the scanpath. We experiment with introducing mixture-of-experts routing into the computation of the cross-attention query and into the decoder, that is, the prediction head responsible for generating the next fixation. We also investigate full routing, in which both components are routed. Furthermore, we compare several routing strategies: hard routing, where the expert is selected deterministically based on the reader-group label, and soft routing, where a learned gate assigns weights over experts, the input is processed by all experts, and the final output is computed as a weighted combination of their outputs. In addition, we examine the use of a shared expert, which is activated independently of the routed group and is intended to capture patterns common across groups, while allowing the group-specific experts to focus on differences in reading behaviour. We conduct experiments under both new-sentence and new-reader data splits, considering training settings that include all reader populations, L1 non-dyslexic and dyslexic readers, and L1 and L2 non-dyslexic readers. MoRe consistently improves over the Eyettention baseline across all data splits. The strongest performance is achieved by configurations using either soft decoder routing or soft full routing with shared experts. These findings suggest that group-conditioned routing is particularly beneficial when the model is given the flexibility to select experts based on scanpath characteristics. Overall, MoRe provides a way to exploit shared regularities in reading behaviour across groups while still generating more population-sensitive synthetic scanpaths and achieving stronger overall performance.

References

- Deng S., Reich D. R., Prasse P., Haller P., Scheffer T., Jäger L. A. (2023). Eyettention: An attention-based dual-sequence model for predicting human scanpaths during reading. *Proc. ACM Hum.-Comput. Interact.* 7.
- Hollenstein N., Barrett M., Björnsdóttir M. (2022). The Copenhagen corpus of eye tracking recordings from natural reading of Danish texts. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, Marseille, France, pp. 1712–1720.
- Shazeer N., Mirhoseini A., Maziarz K., Davis A., Le Q., Hinton G., Dean J. (2017). Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *International Conference on Learning Representations*.



Reading standard and nonstandard parallel forms in Estonian

Annika Kängsepp, Maarja-Liisa Pilvik, Kaidi Lõo, Külli Habicht, Liina Lindström

Institute of Estonian and General Linguistics, University of Tartu

*contact: kaidi.loo@ut.ee

Background: Estonian has a rich inflectional system in which word forms can be created via agglutination, fusion, or a combination of both. As a result, parallel inflected forms, i.e., multiple forms marking the same grammatical case, are very common (e.g., Viht and Habicht 2019). Corpus studies have indicated that alongside standard parallel forms (e.g., *vaidlusi* vs. *vaidluseid* 'argument' PL.PART), also nonstandard variants i.e., forms that diverge from the prescribed norm (e.g., *imet* vs. *ime* 'miracle' SG.PART, where *imet* is the nonstandard form) are widespread (see e.g., Siiman 2019; Kängsepp 2024; Aigro et al. 2026). Despite their frequency, little is known about how such morphological variation is processed in real time. Moreover, the role of individual differences in morpho-logical processing remains underexplored. To address this, we conducted an eye-tracking sentence reading study examining the effects of frequency, normative status and individual differences on Estonian morphological processing.

Method: Eighty-four native speakers of Estonian (47 female, 34 male, 2 non-binary, 1 preferred not to say; 18-59 years old, mean 35 years) participated in the experiment. The participants first completed a demographic questionnaire and a Big Five personality inventory which was followed by the reading task. Participants were instructed to silently read sentences one by one at their own pace, while their eye movements were recorded with an Eyelink Duo eye tracker (SR Research). Two experimental lists were constructed such that each participant saw only one version of each sentence, with the versions differing only in the inflected form. Each participant read a total of 160 sentences (108 experimental, 52 filler sentences). The entire procedure lasted about 60 minutes.

Results & Discussion: The data was analyzed using generalized additive regression. Preliminary analyses indicated that form frequency, but not normative status (standard vs. non-standard), affects reading during the first-pass reading. Participants were, however, sensitive to the nonstandard status of word forms during later stages of reading: nonstandard forms were read longer in total. Regarding individual differences, older female participants with a more open-minded personality read faster than younger male participants with less open-minded personality. Similar pattern was observed in reading both standard and non-standard forms, and only at later stages of reading.

The current study suggests a difference between early and later readings. Early reading is influenced by frequency, while sensitivity to prescriptive norms emerges during later and possibly more reflective reading stages. Overall, the findings support usage-based accounts of morphology, while also suggesting the existence of normative knowledge in real-time reading. Our findings highlight the importance of combining corpus evidence with real-time experimental methods to gain additional insights on morphological variation.

References

- Aigro M., Kängsepp A., Todesk T. (2026). Increasing transparency: Usage of nonstandard inflections by Estonian adults. *Journal of Linguistics*. to appear.
- Kängsepp A. (2024). Ka "a" ndevormide varieerumine ja seda mõjutavad tegurid kirjalikus keeles indefiniitpronoomeni keegi na "itel. *Keel ja Kirjandus* 11, 1016–1037. doi:10.54013/kk803a3.
- Siiman A. (2019). *Vormikasutuse varieerumine ning selle po "hjused osastava ja sissee "tleva ka "a "nde na "itel*. PhD Dissertation. Tartu. URL: <http://hdl.handle.net/10062/64787>.
- Viht A., Habicht K. (2019). *Eesti keele so "namuutmine*. Eesti keele varamu IV. Tartu: University of Tartu Press. URL: <http://hdl.handle.net/10062/76416>.

Tracking reading in Kalaallisut: The morpheme as the basic unit of analysis

Johanne S. K. Nedergaard¹, Solvej W. Jaeger¹, Ivalo I. Mathiassen²

¹Department of Nordic Studies and Linguistics, University of Copenhagen,²Institute of Learning, Ilisimatusarfik

*contact: jskn@hum.ku.dk

Polysynthetic languages, such as Kalaallisut (West Greenlandic), are characterised by rich inflectional and derivational morphology where a complex word is often equivalent to a sentence in an analytic language like English. While it is important to include typologically diverse languages in large multilab collaborations such as MultiPLEYE (Blasi *et al.*, 2022), transferring the measures and analysis methods used for e.g. eye-tracking while reading English is not straightforward.

The key difference when studying reading in Kalaallisut, compared to Indo-European languages, is that the word cannot be treated as the fundamental unit of analysis. Although little is known about how polysynthetic languages are read (but see Mathiassen, 2025; Jaeger *et al.*, 2026), a morpheme-based approach is likely more appropriate. To give a simple illustration of this, we here show a comparison between two word-based and morpheme-based measures of the MultiPLEYE texts in Kalaallisut and in English (see Figure 1).

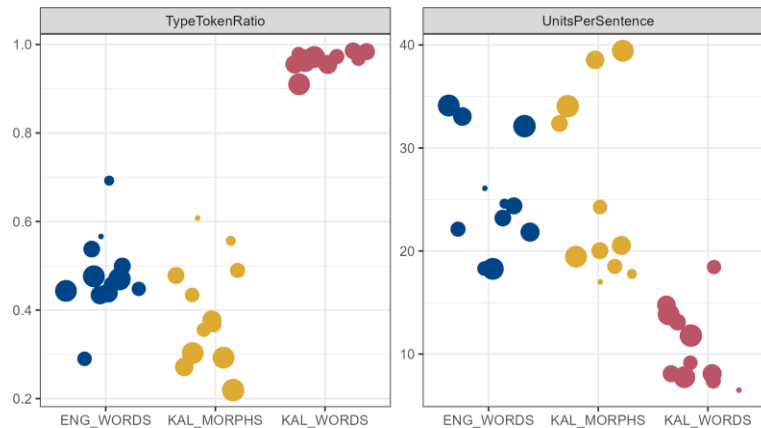


Figure 1: Comparison of type-token ratio (left) and units per sentence (right) for English words and for Kalaallisut morphemes and Kalaallisut words respectively. Each dot represents one of the MultiPLEYE texts with the dot size scaled by text length.

As is clear from Figure 1, the morpheme-based measures are much more similar to the English word-based measures than the Kalaallisut word-based ones are. Morpheme-based analysis will enable us to answer several interesting, fundamental questions about how people read a polysynthetic language, such as: (how) do readers decompose multi-morphemic words? Could different decomposition strategies correlate with reading speed or reading comprehension?

References

- Blasi D. E., Henrich J., Adamou E., Kemmerer D., Majid A. (2022). Over-reliance on english hinders cognitive science. *Trends in cognitive sciences* 26(12), 1153–1170.
- Jaeger S. W., Nedergaard J. S. K., Nielsen M., Trondhjem N. B., Hendriksen M. A., Joergensen A. H., Boye K. (2026, April). Eye tracking in a polysynthetic language: Evidence for a distinction between lexical and grammatical suffixes. Retrieved from osf.io/preprints/psyarxiv/qa7ur_v1.
- Mathiassen I. I. (2025, December). *Elevers stavelser- og morfologiske afkodningsstrategier på yngstetrinnet - En undersøgelse af læseundervisningen i didaktisk design-perspektiv i faget kalaallisut*. Phd thesis, Ilisimatusarfik, University of Greenland, Nuuk, Greenland. Available at <https://da.uni.gl/media/xmzl4tqe/ivalos-phdafhandling-2025.pdf>.



Eye-tracking of sign language: the effect of learning experience

Maroš Filip^{1,2}, Michaela Svoboda^{1,2}, Kateřina Chládková^{1,2}

¹Faculty of Arts, Charles University, ²Institute of Psychology, Czech Academy of Sciences

*contact: filip@praha.psu.cas.com

Eye-tracking (ET) research on sign language indicates that experienced or native signers tend to fixate primarily on the face while observing sign language, whereas non-signers focus more on the hands in the signing space (Sander *et al.*, 2025).

However, findings regarding sign language learners remain somewhat ambiguous. Emmorey *et al.* (2009) re-reported that beginner sign language learners tend to fixate more on the signer's mouth, while experienced native signers focus predominantly on the signer's eyes and surrounding facial regions. In contrast, other studies have found no significant differences in face fixation time between groups (Agrafiotis *et al.*, 2003; Muir and Richardson, 2005).

We have initiated a longitudinal study tracking learners of Czech Sign Language (CSL; N = 19) to examine whether their visual attention patterns change over time and to compare them with a control group of naive participants (N = 21) with no prior CSL experience. The study involves co-registration of ET and electroencephalography (EEG; not reported here). During each session, participants watch 28 short CSL videos (Fig. 1) while completing an attention task requiring recognition of excerpts from the videos.

In total, three sessions are planned: the first shortly after participants begin learning CSL (together with the naive control group), the second approximately six months later, and the third after one year. At the conference, we plan to present data from the first two sessions; however, until this point, the data are not yet processed (first session) or collected (second session). We hypothesize that by the second session, the participants will exhibit increased fixation on the face compared to both their first session and the naive control group.

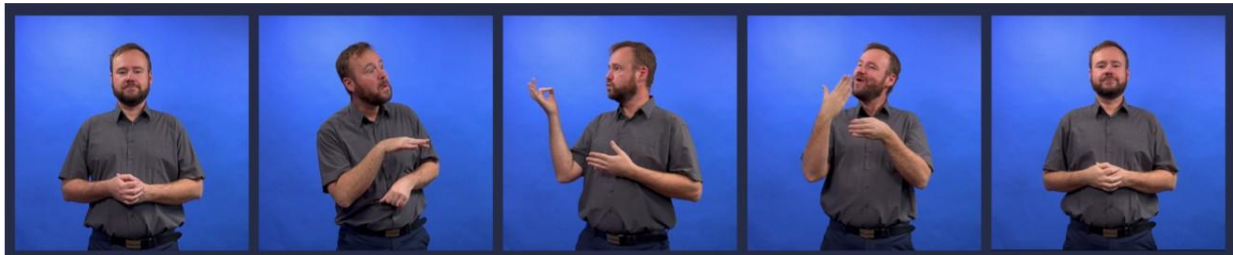


Figure 1: Example of video stimuli.

References

- Agrafiotis D., Canagarajah N., Bull D. R. (2003). Perceptually optimised sign language video coding. In *Proceedings of the 10th IEEE International Conference on Electronics, Circuits and Systems (ICECS 2003)*, Volume 2, pp. 623–626. IEEE.
- Emmorey K., Thompson R., Colvin R. (2009). Eye gaze during comprehension of american sign language by native and beginning signers. *Journal of Deaf Studies and Deaf Education* 14(2), 237–243. doi:10.1093/deafed/enn037.
- Muir L. J., Richardson I. E. (2005). Perception of sign language and its application to visual communications for deaf people. *Journal of Deaf Studies and Deaf Education* 10(4), 390–401. doi:10.1093/deafed/eni037.
- Sander J., Meister N., Finkbeiner T., Rowland C. F., Steinbach M., Friederici A., *et al.* (2025). Deaf signers adapt their eye gaze behaviour when processing an unknown sign language. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Volume 47. URL: <https://escholarship.org/uc/item/018384xx>.



Looking ahead in L1 and L2

Kristina Cergol^{1,4}, Ana Matic^{2,4}, Eva Pavlinušić Vilus^{3,4}, Marijan Palmović^{2,4}

¹Faculty of Teacher Education, University of Zagreb, ²Faculty of Education and Rehabilitation Sciences, University of Zagreb, ³Department of Computational Linguistics, University of Zurich⁴, Laboratory for Psycholinguistics Research, University of Zagreb

*contact: kristina.cergol@ufzg.hr

Background: How does the brain generate predictions during reading, and does this process work the same way in a first and a second language? This study addresses these questions by examining the effects of surprisal and contextual entropy on eye movements during naturalistic reading in L1 Croatian and L2 English, using a within-subject design with 45 proficient Croatian-English bilingual readers. Surprisal — the information-theoretic measure of how unexpected a word is given its context — and contextual entropy — the degree of uncertainty about what word will be presented next — capture distinct, complementary aspects of predictive processing, and their dissociation in bilingual reading reveals something that neither of these two measures could separately show.

Method: Surprisal was computed using a hybrid 5-gram language model interpolating local contextual probabilities from the text with corpus-based unigram frequencies (weighted 30/70 respectively), while contextual entropy was estimated at each word position as the Shannon entropy over the distribution of possible continuations derived from the same model, with add-k smoothing applied to handle sparse contexts.

Results & Discussion: The results of the present study reveal a clear L1–L2 asymmetry. In L1 Croatian, readers slowed down for less predictable words, replicating the classic surprisal effect and confirming that fluent native reading is guided by active, anticipatory prediction. In L2 English, this effect was absent or even reversed: the same readers showed no sensitivity to word predictability on first fixation, suggesting that predictive processing does not operate as a default strategy in the non-native language — even if readers are of a relatively high proficiency. Critically, entropy effects diverged even more sharply: contextual uncertainty seemed to bear no influence on L1 reading, while higher entropy in L2 was associated with faster reading. This counterintuitive finding points to a qualitative shift in the processing strategy: when the linguistic context is uninformative and predictions are unlikely to be accurate, L2 readers appear to abandon predictive attempts altogether and fall back to the more efficient bottom-up, word-by-word processing. These findings seem to speak to fundamental questions in bilingualism research. They demonstrate that the architecture of predictive language comprehension is not simply scaled down in the second language — it is reorganised. Surprisal and entropy dissociate across languages in a way that reveals the strategic, resource-sensitive nature of prediction: L2 readers do not merely predict less, they predict differently, adapting when and how much they invest in anticipation based on the expected payoff. The present findings thus have implications for theories of L2 sentence processing, for understanding how individual differences in proficiency shape predictive capacity, and more broadly for predictive coding accounts of language that assume a single hierarchical generative model operates across all languages a speaker knows.



A living overview of eye-tracking-while-reading datasets with Open library support

Jana Mara Hofmann¹, Deborah Noemie Jakobi¹, David Robert Reich^{1,2}, Paul Prasse², Lena Sophia Bolliger¹, Lena Ann Jäger^{1,3}

¹Department of Computational Linguistics, University of Zurich, ²Department of Computer Science, University of Potsdam, ³Institute 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 (Findable, Accessible, Interoperable, Reusable) 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., Su" ss A., Prasse P., Jakobi D. N., Turuta O., Kasprowski P., Ja" ger 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., Sauberli A., Skrjanec I., Cretton I. C. R., Jakobi D. N., Nisioi S., Prasse P., Jager 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., Jager 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.



Which reading measures are the strongest predictors of reading comprehension?

Jan Brassler¹, Vera Demberg², Lena Ann Jäger¹, Maja Stegenwallner-Schütz³

¹Department of Computational Linguistics, University of Zurich, ²Department of Computer Science, Saarland University, ³Institute of Special Education, University of Koblenz

*contact: jan.brassler@uzh.ch

Background: Reading measures derived from eye movements are commonly categorized into first-pass, second-pass, and global measures, reflecting initial word encounters, rereading, or aggregated scanpath characteristics across the entire text. Previous studies have demonstrated that eye movements reflect how well readers comprehend text (Shubi *et al.*, 2026; Mézière *et al.*, 2024), but mostly focused either on individual measures or on undifferentiated sets of measures, without necessarily considering the category of measures. Moreover, comprehension questions differ in the representations they probe, ranging from the surface text base to the fully enriched situation model. In this study, we determine which combination of first-pass, second-pass, and global measures best predicts accuracy across question types.

Method: We analyze the eye movements of 71 native German-speaking adults reading expository and prose texts, and model the readers' comprehension accuracy using different combinations of reading measures as predictors, as part of the MultiplEYE project. Comprehension questions vary with respect to the level of text comprehension they target, spanning local (inquiring explicit information), bridging (requiring inferences), and global question types (requiring reflecting and assessing information). Measures were calculated exclusively from eye movements on annotated critical spans containing the information relevant to each question. Using Bayesian hierarchical models, we tested all combinations comprising exactly one measure from each category (four first-pass, two second-pass, and five global) to predict comprehension accuracy, and evaluated model performance (five-fold cross validation). Data were split into an exploratory set ($N = 50$) and an independent validation set ($N = 21$) to test the replicability of the identified effects.

Results & Discussion: The top-performing model contains average first-pass gaze duration (first-pass measure), regression rate (second-pass measure), and wrap-up effects (global measure). Specifically, shorter first-pass gaze durations predicted higher accuracy when responding to local and global questions, while higher regression rates were associated with higher accuracy on bridging questions. Despite wrap-up effects appearing in all top-performing models, wrap-up effects were not associated with comprehension accuracy. Validation on an independent sample failed to replicate the observed effects, likely due to limited sample size; ongoing data collection will bring the validation sample to up to 50 readers.

Increased first-pass gaze duration is a well-established index of processing difficulty, specifically regarding lexical access or syntactic integration. Although processing difficulty can in principle compromise the representation of text information at all levels, this study only confirms this for explicit information derived from the text-base and full situation models. Similarly, regression rates reflect difficulty in sentence processing leading readers to reread the text. Since regression rates are associated with improved comprehension performance, they signal successful rereading attempts. The role of wrap-up effects remains unclear and the failure to replicate in the validation models limits current generalizability of the findings. Taken together, the systematic combination of reading measures allows us to distinguish whether certain measures reflect mere processing difficulty or more mindful rereading strategies. This study points to the potential of modelling comprehension accuracy in relation to different types of questions, but the reliability of these associations remains to be established.

References

- Mézière D. C., Yu L., McArthur G., Reichle E. D., von der Malsburg T. (2024). Scanpath regularity as an index of reading comprehension. *Scientific Studies of Reading* 28(1), 79–100. doi:10.1080/10888438.2023.2232063.
- Shubi O., Reich D. R., Klein K. G., Angel Y., Prasse P., Jäger L. A., Berzak Y. (2026). Eyebench: Predictive modeling from eye movements in reading. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.



Genre effects on eye movements in Italian reading: Evidence from the MultiplEYE IT dataset

Daniele P. Radicioni¹, Patrizia Paggio², Alessandro C. Mendoza², Matteo Delsanto¹, Alberto Parola², Viviana Patti¹, Calogero J. Scozzaro¹

¹Computer Science Department, University of Turin, ²Department of Nordic Studies and Linguistics, University of Copenhagen

*contact: daniele.radicioni@unito.it

The MultiplEYE texts have been chosen to ensure coverage of different genres and text types (Kaspere *et al.*, 2026). We know, in fact, that eyetracking measures of reading vary depending on these features (Bolliger and Jaeger, 2025). In this study we have analysed the four genres represented by the texts used to construct stimuli for the Italian collection to see whether they differ in style, lexical difficulty and syntactic complexity. For each text we have extracted type-token ratio (TTR), mean sentence length, mean word length and mean dependency distance (MDD). The measures show variability across the genres: the institutional texts display low lexical difficulty and high syntactic complexity; the popular scientific ones have the longest sentences and show great TTR variability; the literary texts have relatively high TTR and MDD; the argumentative ones score high on lexical and low on syntactic difficulty. We expect these differences to correspond to differences in the eyetracking measures. To test our expectations, we have analysed the data collected in Turin to extract the distribution of eyetracking measures reflecting early as well as later and cumulative processing. This analysis shows noticeable differences across the four genres. In particular, the trends we observe seem to indicate that genre has a limited influence on the earliest stage of processing (FPRT). By contrast, clearer genre-related variation emerges in the later and more cumulative measures (RRT, RPD inc and TFT). At the same time, the sizeable variation in all the measures make it necessary to move beyond descriptive summaries and assess these apparent differences through inferential modelling that accounts for subject- and item-level variation. The analysis of genre effects indicates that reading times vary systematically across genres, but that these differences are not equally visible in all eye-tracking measures. Using mixed-effects regression models with random intercepts for subjects and items, we tested whether genre explains variation in reading behavior beyond baseline heterogeneity associated with individual readers and textual items. Literary texts were treated as the reference category, and separate models were fitted for first-pass reading time (FPRT), re-reading time (RRT), inclusive regression-path duration (RPD inc), and total fixation time (TFT). The results show that genre effects are weakest in the earliest stage of processing and become more pronounced in later, more integrative measures. These findings suggest that genre affects not so much immediate word-by-word uptake as the amount of later rereading, revision, and cumulative effort required for interpretation.

Our results are consistent with recent literature showing that genre shapes eye-movement behavior at more global and strategic levels. Markevich and Streltsova (2025) report genre-driven differences in scanpath organization across narrative, expository, and poetic texts, highlighting the role of genre in shaping reading strategies beyond local fixation behavior. Our results are also compatible with the work by Bolliger and Jaeger (2025), who show that text type modulates reading behavior and interacts with lexical predictors such as surprisal, word length, and frequency. Taken together, these studies converge with our findings in showing that genre is a determinant of reading behavior, while our results further suggest that its clearest effects emerge in rereading- and integration-sensitive measures rather than in early first-pass processing.

References

- Bolliger L. S., Jager L. A. (2025). Genre matters: How text types interact with decoding strategies and lexical predictors in shaping reading behavior. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, Suzhou, China, pp. 7459–7476.
- Kaspere R., Bondar A., ... (2026). The MultiplEYE text corpus: Towards a diverse and ever-expanding multilingual text corpus. LREC 2026: Language Resources and Evaluation Conference.
- Markovich M., Streltsova A. (2025). The influence of text genre on eye movement patterns during reading. *Journal of Eye Movement Research* 18(6), 60.

Capturing comprehension in the reading brain: Effects from language model-derived propositions on eye movements

Adrielli Tina Lopes Rego¹, Joshua Snell², Martijn Meeter¹

¹Education Sciences, Vrije Universiteit Amsterdam, ²Applied and Experimental Psychology, Vrije Universiteit Amsterdam

*contact: a.t.lopesrego@vu.nl

It is well known that reading slows down when text comprehension is difficult. Yet incorporating higher-order linguistic information remains a challenge for models of reading, as this requires representing meaning as it is dynamically derived from text. Based on Kintsch's Construction-Integration theory of discourse processing (Kintsch and Van Dijk, 1978; Kintsch, 1988), here we propose to approximate relations that readers may infer during reading by using propositional triplets – structured representations of the form subject, relation, and object, such as “Google, test, car” in the sentence “Google has begun testing the electric car”. Drawing from natural language processing techniques for information retrieval, we extracted triplets from English texts in three eye-movement corpora - Provo, (Luke and Christianson, 2018), MECO, (Siegelman *et al.*, 2022), and OneStop (Berzak *et al.*, 2025) - using GPT-4, and measured the effect of triplet activation on word reading times. Across corpora, triplet formation at a word was associated with longer reading times at the previous word, beyond the effects of position in sentence, position in text, length, frequency, and surprisal. By offering a way to represent text comprehension through linguistic relationships, our approach may encourage further research on integrating semantic-level processing into models of reading behavior.

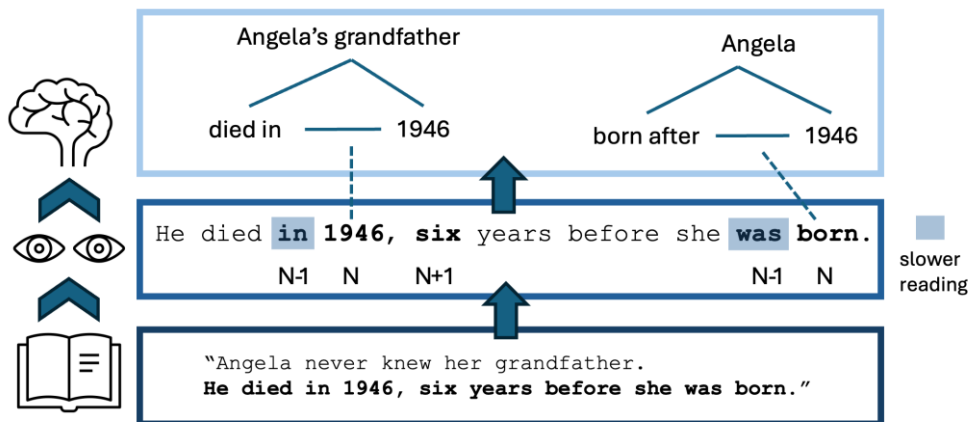


Figure 1: Words preceding the formation of a propositional triplet are read more slowly.

References

- Berzak Y., Malmaud J., Shubi O., Meiri Y., Lion E., Levy R. (2025). Onestop: A 360-participant english eye tracking dataset with different reading regimes. *Scientific Data*.
- Kintsch W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological review* 95(2), 163.
- Kintsch W., Van Dijk T. A. (1978). Toward a model of text comprehension and production. *Psychological review* 85(5), 363.
- Luke S. G., Christianson K. (2018). The provo corpus: A large eye-tracking corpus with predictability norms. *Behavior research methods* 50(2), 826–833.
- Siegelman N., Schroeder S., Acarturk C., Ahn H.-D., Alexeeva S., Amenta S., Bertram R., Bonandrini R., Brysbaert M., Chernova D., *et al.* (2022). Expanding horizons of cross-linguistic research on reading: The multilingual eye-movement corpus (mecoc). *Behavior research methods* 54(6), 2843–2863.



From the Alps to the Diaspora: A work-in-progress on reading profiles across Swiss minority and heritage languages

Thyra Krosness^{1,2}, Not Battesta Soliva¹, Lena Ann Jäger¹

¹Department of Computational Linguistics, University of Zurich, ²FHNW School of Social Work, University of Applied Sciences and Arts Northwestern Switzerland

*contact: thyrasigynhedda.krosness@uzh.ch

It remains unclear whether native speakers of minority language readers exhibit eye movements while reading that is similar to L1 readers of majority languages, heritage speakers of minority languages, or show their own reading profile. Majority language speakers encounter written input across the full range of linguistic domains. Minority language speakers, by contrast, face systematic domain restriction: written use is often limited to informal communication, while the majority language mediates all formal and literacy-intensive activities. A structurally comparable restriction characterizes heritage speakers, whose home language, though potentially a majority language elsewhere, is in its written form restricted to informal, and family contexts. Previous eye-tracking studies reveal that heritage readers show longer mean fixations, lower skipping, higher regressive saccade rates and enlarged frequency effects compared to monolingual adults (Parshina *et al.*, 2021). Eye movements of proficient L2 readers are similar to L1, but the less proficient the L2 reader is, the less sensitive the eye movements become to a word predictability, and more sensitive to its word frequency (Berzak and Levy, 2023). This raises the question of whether minority language status produces a similar processing signature as either L1 speakers, heritage speakers, or exhibit a unique reading profile. Using data from the MultiplEYE corpus (Jakobi *et al.*, 2025), in which all participants read the same texts translated into their respective languages, we present work-in-progress on the reading profiles of Swiss minority and heritage readers of Romansh Sur-silvan and Albanian. We address two research questions: (1) Do we observe longer mean fixations, lower skipping rates, higher regressive saccade rates, less sensitivity to word predictability and enlarged frequency effects in heritage and minority readers compared to native readers of a majority language? (2) Does domain restriction alone, regardless of whether it stems from minority language or heritage status, produce comparable reading profiles? We first fit a linear model comparing Albanian L1 to Albanian heritage speakers, to investigate the profiles of the heritage readers. We then move on to fit Albanian L1 versus Romansh speakers to investigate the effect of minority language status. Finally, we investigate the differences in reading profiles between Romansh and Albanian heritage speakers, contrasting the two presuppositions for domain restriction. Fixation duration measures, skipping probabilities, and regression rates are modeled as dependent variables, and word frequency, word length, and surprisal as main predictors. We hypothesize that the profile for Albanian heritage readers will show the expected heritage profile, consistent with Parshina *et al.* (2021). Romansh readers will be more similar to heritage readers, rather than the Albanian L1 readers.

References

- Berzak Y., Levy R. (2023). Eye Movement Traces of Linguistic Knowledge in Native and Non-Native Reading. *Open Mind* 7, 179–196. doi:10.1162/opmi.00084.
- Jakobi D. N., Stegenwallner-Schütt M., Hollenstein N., Ding C., Kaspere R., Matic' Š' koric' A., Pavlinusic Vilus E., Frank S., Müller M.-L., Jensen de López K. M., Kharlamov N., B. Søndergaard Knudsen H., Berzak Y., Lion E., Sekerina I. A., Acarturk C., Ansari M. F., Harezlak K., Kasprowski P., Bautista A., Beinborn L., Bondar A., Boznou A., Bradshaw L., Hofmann J. M., Krosness T., Soliva N. B., C'epani A., Cergol K., Dos'en A., Palmovic M., C'erpja A., Chirino D., Chromy' J., Demberg V., Š' krjanec I., Deniz N. D., Fajardo D. I., Gime'nez-Salvador M., M'inguez-Lopez X., Filip M., Freibergs Z., Gomes J., Janeiro A., Luegi P., Ver'issimo J. a., Gramatikov S., Hasena' cker J., Haveriku A., Kote N., Kamal M. M., Kundeneddzierska H., Klimek-Jankowska D., Kosutar S., Krakowczyk D. G., Krejtz I., Łockiewicz M., Lo' o K., Motieju nienundefined J., Nasir J. A., Nedergard J. S. K., Ozkan A., Preininger M., Punga' L., Reich D. R., Tschirner C., Rot v., Sa' uberli A., Sole'-Casals J., Strati E., Svoboda I., Trandafil E., Varlokosta S., Vulchanova M., Ja' ger L. A. (2025). MultiplEYE: Creating a Multilingual Eye-tracking-while-reading Corpus. In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications*, ETRA '25, New York, NY, USA. Association for Computing Machinery. doi:10.1145/3715669.3726843.
- Parshina O., Laurinavichyute A. K., Sekerina I. A. (2021). Eye-movement Benchmarks in Heritage Language Reading. *Bilingualism: Language and Cognition* 24(1), 69–82. doi:10.1017/S136672892000019X.



Brain activity correlations in reading: an ongoing study with fNIRS and eye-tracking coregistration

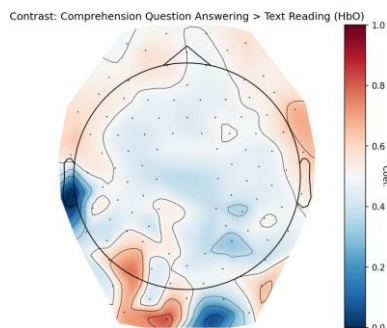
Cengiz Acarturk, Anna Wilkosz, Melike Çağlayan, Jakub Jan Romanowski

¹Centre for Cognitive Science, Jagiellonian University, Poland

*contact: cengiz.acarturk@uj.edu.pl

In the study of eye-movements during reading, a significant gap remains in understanding the relationship between eye-movement patterns and comprehension of text. In the present study, we address this gap by employing coregistration of fNIRS (functional Near-Infrared Spectroscopy) and eye tracking during reading of Polish texts and answering comprehension questions. The central research question has been whether it would be possible to discriminate between the brain activity during reading text and during answering comprehension questions, and to identify eye-movement variables that might correlate with the brain activity, as measured by fNIRS.

Alongside being an indicator of a set of cognitive states in clinical settings, the changes in blood oxygenation have been interpreted as indirect indicators of cognitive load and attention, including studies in reading (Quaresima *et al.*, 2012). The fNIRS is an optical neuroimaging methodology (Jobsis, 1977), designed for measuring brain activity by detecting changes in hemoglobin concentration, thus blood oxygenation, using near-infrared light (Villringer *et al.*, 1993). Although fNIRS provides inferior whole-brain spatial resolution compared to fMRI (functional Magnetic Resonance Imaging), it offers portability, high temporal resolution, and motion tolerance suitable for measuring activity of the cortical surface. For coregistration of fNIRS and eye tracking, we designed an experiment in which text-reading blocks are followed by comprehension-question blocks, and the eye-tracking and fNIRS measures can be separated for the specific task blocks.



Regarding the expected findings, we anticipate correlations between localized oxygenation and specific oculomotor behaviors, such as pupil dilation. From a methodological point of view, our initial findings show that the coregistration of fNIRS and eye-tracking has the potential to allow researchers to measure and interpret eye-movement patterns in relation to comprehension indicators, especially pupillometry and brain activity. Additionally, the initial results have shown that it is possible to discriminate between text-reading sessions and comprehension-question sessions in terms of fNIRS variables alongside eye-movement variables, as exemplified in the figure.

Consequently, this ongoing study aims to expand global multilingual reading datasets, by bridging the existing gap between linguistics-oriented research and cognitive science, and inform the next generation of psychophysiological grounded computational reading models.

References

- Jobsis F. F. (1977). Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science* 198(4323), 1264–1267.
- Quaresima V., Bisconti S., Ferrari M. (2012). A brief review on the use of functional near-infrared spectroscopy (fnirs) for language imaging studies in human newborns and adults. *Brain and language* 121(2), 79–89.
- Villringer A., Planck J., Hock C., Schleinkofer L., Dirnagl U. (1993). Near infrared spectroscopy (nirs): a new tool to study hemodynamic changes during activation of brain function in human adults. *Neuroscience letters* 154(1-2), 101–104.



Reading comprehension, executive functions and vocabulary in Danish adult L1 speakers

Hanne B. Søndergaard Knudsen, Kristine Jensen de Lopez, Nik Kharlamov

Department of Culture and Communication, research group: Reading Cognition, Aalborg University

*contact: hannebsk@ikk.aau.dk

Background: Executive Functions (EF), especially working memory (WM), are related to reading comprehension (Kim, 2020; Butterfuss & Kendeou, 2018) and eye movements during reading (Traxler et al., 2015). It is, however, unclear whether these relationships depend on the specific language and its linguistic characteristics. In this study we investigate reading comprehension and cognition in Danish-speaking adults. Danish is characterized by deep orthography (Seymour et al., 2003) and tendency for reduction of word boundaries (Basbøll, 2005) and hence may place higher demands on EF for language use generally and for reading comprehension in particular.

Method: We investigated a sample of monolingual Danish-speaking young adults for their reading comprehension skills, multiple EF abilities (verbal and non-verbal WM, inhibition, shifting) and vocabulary to investigate the relation between cognitive abilities and reading comprehension related to different eye tracking measures. The design, procedure and stimuli align with what has been developed in the MultiplEYE COST Action CA21131 (see Müller et al., 2024).

Results & Discussion: As we are currently in the process of collecting data, preliminary results will be presented at the MultiplEYE Final Conference. Results from reading comprehension by young Danish-speaking children suggest that cognitive flexibility predicts their sentence reading comprehension (see Knudsen et al., 2018), and therefore we expect to see significant associations between cognitive abilities and reading comprehension of Danish adults. Specifically, for participants with lower scores on executive function measures, we expect longer dwell times and larger number of regressions especially for complex words.

References

- Basbøll, H. (2005). *The phonology of Danish*. Oxford University Press.
- Butterfuss, R., & Kendeou, P. (2018). The role of executive functions in reading comprehension. *Educational Psychology Review*, 30(3), 801–826. <https://doi.org/10.1007/s10648-017-9422-6>
- Kim, Y.-S. G. (2020). Toward integrative reading science: The direct and indirect effects model of reading. *Journal of Learning Disabilities*, 53(6), 469–491. <https://doi.org/10.1177/0022219420908239>
- Müller, M.-L., Jakobi, D., Stegwallner-Schuetz, M., Kasperè, R., Hollenstein, N., & Jäger, L. (2024). *Data Management Plan (DMP) for COST Action "MultiplEYE" (CA21131)*. PsychArchives. <https://doi.org/10.23668/psycharchives.15018>
- Seymour, P.H.K., Aro, M., Erskine, J.M. and collaboration with COST Action A8 network (2003), Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94: 143–174. <https://doi.org/10.1348/000712603321661859>
- Søndergaard Knudsen, H. B., Jensen de López, K., & Archibald, L. M. D. (2018). The contribution of cognitive flexibility to children's reading comprehension—The case for Danish. *Journal of Research in Reading*, 41(Suppl 1), S130–S148. <https://doi.org/10.1111/1467-9817.12251>
- Traxler, M. J., Long, D. L., Tooley, K. M., Johns, C. L., Zirnstein, M., & Jonathan, E. (2012). Individual differences in eye-movements during reading: Working memory and speed-of-processing effects. *Journal of Eye Movement Research*, 5(1), 1–16. <https://doi.org/10.16910/jemr.5.1.5>



Working memory capacity and eye movement patterns in reading: A correlational analysis of regressions in the Slovenian MultiplEYE dataset

Lamija Zahirović, Lea Popović, Anastasija Žarevac, Eva Viktorija Zajc, Katarina Lasković, Kristian Elersič, Anja Podlesek

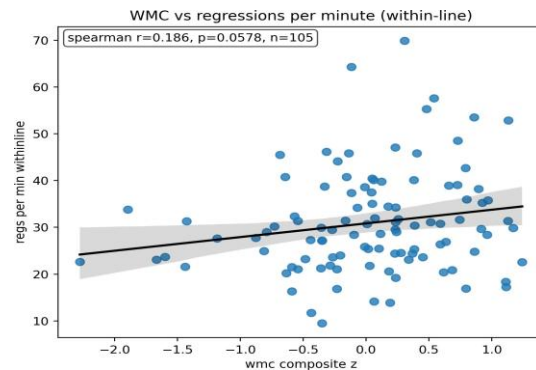
Department of Psychology, Faculty of Arts, University of Ljubljana

*contact: lz67616@student.uni-lj.si, lp58329@student.uni-lj.si

Working Memory Capacity (WMC) reflects an individual's ability to maintain and process information simultaneously and is considered a key factor in reading comprehension efficiency (Daneman & Carpenter, 1980; Just & Carpenter, 1992). Prior research suggests that limited WMC leads to processing difficulties, which may manifest in eye movement behavior such as increased regressions (backward eye movements) (Just & Carpenter, 1992; Rayner, 1998). The present preliminary analysis aims to investigate whether lower WMC scores are associated with a higher frequency of regressions in the Slovenian language in native speakers during reading tasks, focusing on trials where readers demonstrated at least basic text comprehension.

The preliminary analysis is based on eye-tracking data collected in Slovenia as part of the MultiplEYE project. Regressions were extracted from EyeLink saccade events and defined as leftward saccades exceeding 50 pixels within the same text line, with vertical deviation < 30 pixels, then normalized per minute to control for reading time. WMC was operationalized as a composite z-score derived from memory updating, operation span, sentence span, and spatial short-term memory tasks using partial-credit unit scoring. Text understanding was assessed with six comprehension questions after each text and used as a filtering criterion. The dataset included 105 participants and 10 texts from different categories (literary, instructional, popular science, and argumentative), yielding 1050 participant–text observations with comprehension above 60%.

Preliminary results showed a small positive but non-significant association between WMC and regression rate (Spearman's $r = 0.186$, $p = 0.0578$; Figure 1). Additionally, we tested whether comprehension was predicted by regressions per minute, WMC, and their interaction, while controlling for text category and accounting for repeated measurements within participants. The results showed no significant effects of regressions per minute, WMC, or their interaction, while text category emerged as the strongest predictor.



The observed trend between WMC and regression rate does not align with the initial assumption that regressions primarily reflect processing difficulty. Some eye movement research suggests that regressions are not exclusively indicators of failure of text understanding but can also reflect adaptive rereading strategies that support integration and comprehension (Staub & Benatar, 2013). In this context, the present trend may indicate that higher-capacity readers may use rereading strategies more frequently. To better understand how WMC relates to eye-movement behavior, the analysis should also test whether the effect is moderated by individual differences in reading skill, examine each WMC component separately, assess whether the effect varies with text difficulty and distinguish between different types of regressions, such as short corrective regressions and long-distance rereading.

References

- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450–466.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122–149.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422.
- Staub, A., & Benatar, A. (2013). Individual differences in fixation duration distributions in reading. *Attention, Perception, & Psychophysics*, 75(6), 1304–1315.



Effects of context length and text genre on Turkish language models' prediction of human reading: A lossy-context surprisal analysis

Onur Keleş¹, Sezen Bektaş Yüksel², Servet Arda Kayhan¹, Feyza Budan¹, İris Beyza Deveci¹, Nazik Dinctopal Deniz^{2,3}

¹Department of Linguistics, Boğaziçi University, ²Department of Foreign Language Education, Boğaziçi University, ³M.A. Program in Cognitive Science, Boğaziçi University

*contact: onur.keles1@bogazici.edu.tr

Surprisal theory predicts slower reading for less predictable words (Levy, 2008). Psychometric predictive power (PPP; Goodkind & Bicknell, 2018) tests this by examining the log-likelihood gain in predicting human reading behavior through a language model's (LM) surprisal value. Kuribayashi et al. (2022) showed that LMs can become more human-like when prior context is restricted, suggesting that full-context Transformers use information that readers do not always maintain online. We compare full and restricted context LM surprisal in Turkish, an agglutinative SOV language, and ask whether their fit to human eye-movement data depends on model size, text genre, and comprehension accuracy. We used eye-tracking data from 57 native Turkish readers who read 10 MultiplEye passages (Jakobi et al., 2025) from Literary, Institutional, Argumentative, and Popular Science genres. Seven Turkish causal LMs were screened by tokenizer fertility (mean subwords/word; Rust et al., 2021). Kanarya (Safaya et al., 2022) had the lowest fertility (1.35); so, we analysed Kanarya-750m and Kanarya-2b. For each model we computed word surprisal under full context and under lossy n -gram contexts retaining the last 20, 10, 7, 5, 3, or 2 words (where 20 words correspond to broad local context and 2 words correspond to extremely limited context), and examined whether surprisal predicted human eye-tracking measures: log first fixation duration (FFD), gaze duration (GD), total duration (TD), re-reading duration (RRD), total incoming regressions (TRC_{in}), and probability of regression out (PRO). Frequency and word length were fixed effects; subjects and stimulus texts were random effects. Separate PPP models were fit for each context length. We report the results for full-context vs. 2-gram and full-context vs. 7-gram contrasts because they test extreme and intermediate context limitation, respectively. These contrasts were tested by paired permutation tests on by-token squared residuals. Comprehension accuracy was overall 65.3%. It was the highest for Institutional (69.3%) and Argumentative (69.7%) texts and the lowest for Literary texts (62.8%); accuracy for Popular Science texts (66.7%) was in the middle; $\beta = -.069$, $p = .001$. Per-word GD was slower for Institutional and Literary than Argumentative texts, both $p < .001$. Figures 1 and 2 summarize the GD PPP results (reported as $\Delta\text{LogLik} \times 10^{-3}$) by context length and genre. The context-length effect differed by model size. For Kanarya-750m, predictive power was highest, for GD, when context was not limited (PPP = 5.77) and lost its predictive power decreased in 2-gram contexts ($\Delta = -0.95$; permutation test: $p < .002$). The full context was favored in Kanarya-750m for TD ($p < .001$), RRD ($p = .012$), and PRO ($p < .001$), and in Kanarya-2b for PRO ($p < .001$), as well. Kanarya-750m also showed no 7-gram advantage on any measure ($ps \geq .09$). However, for Kanarya-2b, 7-gram context predicted GD ($\Delta = +0.12$; $p = .002$), TD ($\Delta = +0.86$; $p < .001$), and RRD ($\Delta = +0.69$; $p = .004$) better than full context. Context length effects also interacted with text genre: both full-context vs. 2-gram and full-context vs. 7-gram contrasts showed significant genre effects ($p's < .001$). Institutional was the only genre where 2-gram surprisal improved prediction for both models. ($\Delta \approx +1.5$ to $+1.8$; $p < .001$). In contrast, Literary and Popular Science texts showed the strongest full-context advantage ($\Delta \approx -2.5$ to -5.3), suggesting greater dependence on longer prior context. We also examined whether readers whose TD was more strongly related to surprisal showed higher comprehension accuracy. The results showed that readers whose total duration was more surprisal-sensitive comprehended better: the surprisal–TD relation correlated positively with comprehension accuracy in Popular Science ($r = .442$, $p < .001$) and Literary texts ($r = .267$, $p = .045$). In sum, lossy-context in Turkish was not uniformly beneficial: it improved prediction for the larger Kanarya-2b model, and for Institutional texts but full context was favored for the smaller Kanarya-750m model and for later eye-movement measures. Our results, thus, show that the restricted context effect reported for English and Japanese extends to Turkish (Kuribayashi et al., 2022), but that the useful context window varies with model size, measure, and genre.

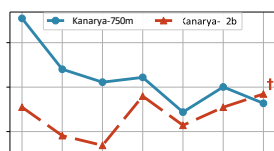


Figure 1. GD PPP by context length and model size.

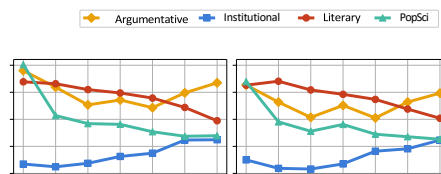


Figure 2. GD PPP by context length, model size, and genre.

References

Goodkind & Bicknell (2018); Jakobi et al. (2025); Kuribayashi et al. (2022); Levy (2008); Rust et al. (2021); Safaya et al. (2022).



Investigating the effect of text simplification on reading comprehension in advanced L2 English speakers

Rares-Alexandru Roscan, Oleksandra Kuvshynova, Mircea-Mihai Marin, Sergiu Nisioi

*contact: sergiu.nisioi@unibuc.ro

Recent insights from the Text Simplification domain (Alva-Manchego *et al.*, 2025) show that text simplification can be adapted for different levels of language proficiency. However, few studies Berzak *et al.* (2025) investigate the actual reading improvements gained by simplified texts.

This study investigates whether algorithmically reducing English text complexity (from C1-C2 to B1-B2) genuinely improves reading comprehension for advanced C-level readers. The pool of participants consists of approximately 100 native Romanian speakers, with reading levels of C1 (27%) - C2 (63%), age 23-35, students enrolled in either Computer Science or Cognitive Science programs. We utilize a subset of the English stimulus texts from the MultiPLEYE dataset (Kaspere *et al.*, 2026) and we augment the data with three additional C2-level texts: one argumentative text on Marxist feminist critical theory from a book by the sociologist Lise Vogel, an institutional text from the WHO on maternal, infant and young child nutrition, and an argumentative text from UNESCO about the ancestral practices in Yasun'í Biosphere Reserve. This includes additional comprehension questions created using the MultiPLEYE guidelines.

Several simplification strategies have been manually compared, with the best approach being based on Gemma 3 12B-IT using a multi-stage pipeline. After simplification, we verify that the comprehension questions are still answerable from the modified text. Each participant receives a random subsample of 50% original texts and 50% simplified texts, is asked to read and respond to the comprehension questions. Additionally, for each participant we conduct an English reading CEFR-level assessment using EF-SET, a vocabulary size estimation using Lextale, and a reverse digit span task. From a total of 113 participants, 16 have been removed due to accidental or text or question skips.

Additional post-hoc data analysis is ongoing, and preliminary empirical results from fitting GLMM models re-reveals a surprising trend: comprehension accuracy remains stable across both simplified and original texts, and simplified texts show no significant improvement in reading comprehension. However, reading times for the simplified texts exhibit a statistically significant reduction of approximately 20%. This suggests that for advanced learners, text simplification acts primarily as a processing optimizer rather than a necessary comprehension aid. Furthermore, there are significant interactions between the participants' grammar score and reading times, where participants with lower grammar scores show significant improvements in reading times ($p = 0.003$).

Finally, we report a critical methodological limitation regarding cohort generalization. An attempt to replicate the study with a substantial secondary cohort of nearly 30 high school students (aged 15-16) yielded compromised data. The inherent length and cognitive demand of the texts induced rapid fatigue and satisficing behavior (e.g., frequent skipping), rendering the experimental design suboptimal for this age group. This highlights a future need to calibrate the experiment construction adequately to different age groups.

References

- Alva-Manchego F., Stodden R., Imperial J. M., Barayan A., North K., Tayyar Madabushi H. (2025, November). Findings of the TSAR 2025 shared task on readability-controlled text simplification. In M. Shardlow, F. Alva-Manchego, K. North, R. Stodden, H. Saggion, N. Khallaf, et al. Hayakawa (Eds.), *Proceedings of the Fourth Workshop on Text Simplification, Accessibility and Readability (TSAR 2025)*, Suzhou, China, pp. 116–130. Association for Computational Linguistics. URL: <https://aclanthology.org/2025.tsar-1.8/>, doi:10.18653/v1/2025.tsar-1.8.
- Berzak Y., Malmaud J., Shubi O., Meiri Y., Lion E., Levy R. (2025). Onestop: A 360-participant English eye tracking dataset with different reading regimes. *Scientific Data*.
- Kaspere, R., Bondar, A., Nisioi, S., Stegenwallner-Schütz, M., Søndergaard Knudsen, H. B., Matić, A., ... & Jäger, L. A. (2026). The MultiPLEYE text corpus: Towards a diverse and ever-expanding multilingual text corpus. In 15th International Conference on Language Resources and Evaluation. European Language Resources Association (ELRA).



Eye-movement patterns in Romanian text reading: A linguistic perspective

Daria Horga, Antonia-Maria Popeanga

Faculty of Mathematics and Computer Science, University of Bucharest

*contact: daria.horga@s.unibuc.ro, antonia-maria.popeanga@s.unibuc.ro

Background: Eye-tracking research has consistently shown that reading behavior is shaped by linguistic factors such as word length and grammatical category (Luke and Christianson, 2017). However, most studies have focused on languages with relatively simple morphology, leaving morphologically rich languages comparatively underexplored. This study investigates reading behavior in Romanian, a language characterized by complex inflectional morphology and clitic constructions, using eye-tracking data from the MultiplEYE corpus (Jakobi, 2025).

Method: To support this analysis, we first detected fixations using the pymovements library (Krakowczyk et al., 2023), and subsequently implemented a velocity-based procedure for saccade detection, together with a set of metrics capturing the spatial and temporal characteristics of fixations. Building on this preprocessing pipeline, we then developed predictive models that integrate standard eye-tracking metrics with variables tailored to Romanian morphology, including enclitic definite articles (-ul, -ului, -lor), clitic pronoun constructions (s-au, s-i-a), noun case and number marking, diacritics (ă, â, î, ș, ț), hyphenated forms (după -amiază). In this pilot study, we evaluated the predictability of eye movements using classifiers for word skipping, fixation duration, and regression events. To ensure generalizability across reading patterns, all models were trained using a subject-independent leave-one-subject-out validation framework.

Results & Discussion: Our results show that word skipping in Romanian is primarily driven by word length and frequency, with additional linguistic features—including function word status and clitic constructions—providing further predictive signal beyond these factors (ROC-AUC $\approx$ 0.78).

Fixation duration, modeled as a classification task (short, medium, long; ROC-AUC 0.75), is primarily influenced by word length and word position. Romanian-specific features, such as enclitic articles, clitic constructions, and case marking, provide additional predictive value, while diacritics also contribute, likely reflecting increased visual and orthographic processing demands.

For regression events, the models achieved a mean ROC-AUC $\approx$ 0.68. These regressions are influenced by both general and language-specific features, including fixation duration and surprisal (Jakobi et al., 2023). Shorter regressions tend to land on oblique cases and function words, while longer ones target content words, suggesting a greater loss of context.

In conclusion, this study takes a first step toward addressing the scarcity of Romanian eye-tracking research, emphasizing the value of incorporating language-specific cues for more linguistically informed reading models.

References

- Jakobi D. N., Kern T., Reich D. R., Haller P., Jaeger L. A. (2023). Potec: A german naturalistic eye-tracking-while-reading corpus. *Scientific Data* 10(1), 640. doi:10.1038/s41597-023-02534-1.
- Jakobi D. N. e. a. (2025). MultiplEYE: Creating a multilingual eye-tracking-while-reading corpus. In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications (ETRA '25)*, New York, NY, USA. Association for Computing Machinery.
- Krakowczyk D. G., Reich D. R., Chwastek J., Jakobi D. N., Prasse P., Suess A., Turuta O., Kasprowski P., Jaeger L. A. (2023). pymovements: A python package for processing eye movement data. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*, New York, NY, USA. Association for Computing Machinery. URL: <https://link.springer.com/article/10.3758/s13428-017-0908-4>.
- Luke S. G., Christianson K. (2017). Predictors of eye movements during reading: A replication of rayner, slattery, drieghe, and liv-ersedge (2011) with english readers. *Behavior Research Methods* 49(2), 623–631. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC12208991/#CR55>.



A work-in-progress towards guidelines for designing Easy Language multiple choice reading comprehension items

Thyra Krosness^{1,2}, Annette Lichtenauer², Claire Reymond³, Lena Ann Jäger¹, Simone Girard-Groeber²

¹Department of Computational Linguistics, University of Zurich, ²FHNW School of Social Work, University of Applied Sciences and Arts Northwestern Switzerland, ³Basel Academy of Art and Design FHNW

*contact: thyrasignhedda.krosness@uzh.ch

Guidelines for creating multiple-choice reading comprehension questions (MCRC) do not cover the creation of Easy Language MCRC items for psycholinguistic studies. However, psycholinguists widely use MCRC items in reading research to assess participants' understanding of texts, and increasingly study populations that rely on Easy Language (EL), such as second language readers or readers with intellectual disabilities. Easy Language "refers to modified forms of standard languages, which aim to facilitate reading and language comprehension (...)" (Lindholm and Vanhatalo, 2021). Its core linguistic features include short, syntactically simple sentences, high-frequency and concrete vocabulary, consistent use of the same words for the same concepts, and the avoidance of negations and passive constructions. Constructing MCRC items for Easy Language materials is particularly challenging due to the reduced linguistic complexity, with short sentences and unambiguous wording. In this paper, we therefore propose a set of guidelines that will aid researchers in creating valid MCRC items for studies involving target groups of EL using EL texts, and enable them to claim that their RC data measures true reading comprehension.

This work is separated into two parts; first a systematic literature review on guidelines for constructing MCRC materials from 2016-2026 to identify gaps relevant to EL, with particular attention to considerations taken in relation to the target groups and implications for item design. Second, we present a work-in-progress toward a structured framework for designing MCRC questions for EL contexts, covering the following: all items must adhere to EL rules; the question item can be formulated as a question or a statement; and we adapt the structured annotation scheme of Berzak *et al.* (2020) to EL materials, motivating specific adaptations, most notably the omission of an incorrect-interpretation distractor, which is not feasible when texts leave little room for misinterpretation.

As next steps, we plan to adapt the set of item-writing flaws for the EL context, distinguishing those that make assessments unintentionally easier, such as implausible distractors, grammatical cues, or answer options identifiable without reading the text, from those that make them unintentionally harder, such as language that does not comply with EL norms or negatively framed stems. The next steps also include a structured reporting check-list to standardize how MCRC materials and design decisions are documented, improving transparency and reproducibility. We further plan to pilot and test the resulting MCRC items with EL target readers and a control group.

References

- Berzak Y., Malmaud J., Levy R. (2020). STARC: Structured Annotations for Reading Comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, pp. 5726–5735. Association for Computational Linguistics. [doi:10.18653/v1/2020.acl-main.507](https://doi.org/10.18653/v1/2020.acl-main.507).
- Lindholm C., Vanhatalo U. (Eds.) (2021). *Handbook of Easy Languages in Europe*. Number 8 in Easy – Plain – Accessible. Berlin: Frank & Timme. [doi:10.26530/20.500.12657/52628](https://doi.org/10.26530/20.500.12657/52628).



Irony processing in bilingual adults with autism: Exploring the pragmatic-cognitive interface through eye-tracking

Qingyang Tang, Bonnie Wing-Yin Chow, Alex Perovic

Linguistics, University College London

*contact: qingyang.tang.24@ucl.ac.uk

Pragmatic difficulties are widely reported in autism, particularly in the comprehension of non-literal language such as irony, which requires integrating contextual information and inferring speaker intent. However, findings on irony comprehension in autistic adults are mixed, and research has largely focused on monolingual populations. This study investigates irony processing in bilingual and monolingual autistic adults and examines whether executive functions (EF) and theory of mind (ToM) predict pragmatic performance.

Using a quantitative cross-sectional design, approximately 90 high-functioning autistic adults (aged 18-50) will be recruited and stratified by language status (bilingual vs. monolingual). EF and ToM will be assessed through established behavioural tasks administered online. Participants will then complete a laboratory-based eye-tracking task in which they read short scenarios containing ironic or literal utterances. Eye-movement measures across predefined regions of interest, including first-pass reading time, regression-path time, total reading time, and comprehension accuracy, will capture real-time processing dynamics. Data will be analysed using mixed-effects modelling in R to examine the relationships between EF, ToM, bilingualism, and irony processing.



Reading goals in eye-tracking research: A comparative analysis of purpose-driven processing in L1 and L2 readers

Hatice Ozata¹, Filiz Rızaoğlu²

¹Foreign Language Education, Giresun University, ²Foreign Language Education, Pamukkale University

*contact: hatice.ozata@giresun.edu.tr

This study explores how reading purpose influences cognitive processing during reading, with a specific focus on eye movement patterns in first (L1) and second language (L2) contexts. It also investigates possible interactions between working memory capacity (WMC) and reading purpose. Drawing on experimental data from 75 university students (L1 English and L1 Turkish-L2 English speakers), the study examines how different reading goals (reading for information vs. reading for study) influence eye movement patterns and comprehension outcomes. Eye-tracking measures, including fixation behavior, regressions, and rereading patterns, were analyzed along-side comprehension performance. Comprehension was assessed through recall and recognition measures, while WMC was measured using a standardized reading span task. Results indicate that reading purpose plays a critical role in comprehension strategies, particularly in L2 readers, who demonstrated enhanced recall performance under study-oriented conditions. Eye movement data suggest increased engagement in rereading and regression behavior during purposeful reading, reflecting deeper processing. Notably, L2 readers showed greater adaptability to task demands compared to L1 readers. While eye movement measures showed limited sensitivity to WMC differences, comprehension outcomes revealed important cognitive interactions. Reading purpose enhanced recall performance, particularly among L2 readers with lower working memory capacity, suggesting a compensatory role of goal-directed processing. These findings underscore the role of cognitive re-sources in reading and suggest that task demands can differentially support readers depending on their working memory profiles. The findings highlight the importance of goal-directed reading strategies in L2 contexts and suggest pedagogical implications for designing reading tasks that support deeper comprehension and retention.



Eye movements and reading comprehension in neurodivergent readers: A study on children with formal ADHD diagnosis

Cengiz Acarturk¹, ESMEHAN ÖZER², Belgin Üstün Güllü³, Halime Betül Kırmızı⁴, Kardelen Çiftci⁵

¹Centre for Cognitive Science, Jagiellonian University, Poland, ²Division of Mental Disabilities Education, Department of Special Education, Gazi University, Türkiye, ³Department of Psychology, Faculty of Arts and Sciences, Düzce University, Türkiye, ⁴Linguistics Lab (DiLab), Graduate School Social Sciences, Ankara University, Türkiye, ⁵Psychology Department, Faculty of Language, History and Geography, Ankara University, Türkiye

*contact: cengiz.acarturk@uj.edu.pl

Attention-Deficit/Hyperactivity Disorder (ADHD) is a condition frequently associated with reading difficulties, which can significantly impact academic achievement and daily functioning. Its prevalence has been reported as 7.6% in children and 5.6% in adolescents (Salari *et al.*, 2023). As a neurodevelopmental disorder, ADHD characterized by symptoms of inattention, hyperactivity, and impulsivity (APA, 2013). The available research has shown that individual readers, diagnosed with ADHD, also exhibit challenges in reading comprehension. Eye-tracking methodology has played a methodological role in the study of these challenges, by providing data that facilitate the investigation of oculomotor processes underlying reading. However, the use of eye tracking for studying reading difficulties in children has led to divergent findings, indicating the need for methodological standardization and expansion towards multiple languages (Papadopoulos *et al.*, 2026).

The present study aims to contribute to the research on ADHD by providing data from Turkish, a language characterized by regular and consistent grapheme-to-phoneme mappings. We present findings from ongoing research, particularly focusing on eye-movement metrics, such as fixation durations, saccadic amplitudes, and regression rates, commonly conceived as common indicators of attention and executive function deficits that may occur during reading. The eye-tracking data have been analyzed in relation to a set of cognitive tests, such as WISC-IV Wechsler Intelligence Scale for Children–Fourth Edition (Wechsler, 2003) and behavioral games designed for diagnosing verbal and visual working memory in primary school children (Van de Weijer-Bergsma *et al.*, 2015, 2016), to allow researchers controlling for further metrics, and to study the correlates between visual attention and executive functioning. The data recorded from 52 children diagnosed with ADHD (N = 21) and their typically developing peers (N = 31) reveal significant differences in eye-movement patterns in reading, especially in mean fixation durations on single words.

References

- APA (2013). *American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders* (5 ed.).
- Papadopoulos T. C., Fella A., Acarturk C., Bahçekapılı E., Canaj K., Cergol K., Dinc, topal Deniz N., Hollenstein N., Hrzica G., Jozipovic M., Kaspere R., Klimek-Jankowska D., ..., & Erol Barkana D. (2026). Eye-tracking in reading research: A systematic review of studies with children of varying reading abilityeye-tracking in reading research: A systematic review of studies with children of varying reading ability. *European Psychologist* 31(2), 134–168.
- Salari N., Ghasemi H., Abdoli N., Rahmani A., Shiri M. H., Hashemian A. H., ..., Mohammadi M. (2023). The global prevalence of adhd in children and adolescents: A systematic review and meta-analysis. *Italian Journal of Pediatrics* 49(1), 48.
- Van de Weijer-Bergsma E., Kroesbergen E. H., Jolani S., Van Luit J. E. (2016). The monkey game: A computerized verbal working memory task for self-reliant administration in primary school children. *Behavior Research Methods* 48(2), 756–771.
- Van de Weijer-Bergsma E., Kroesbergen E. H., Prast E. J., Van Luit J. E. (2015). Validity and reliability of an online visual-spatial working memory task for self-reliant administration in school-aged children. *Behavior Research Methods* 47(3), 708–719.
- Wechsler D. (2003). *Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV)*. The Psychological Corporation.



EvaTrack: a flexible framework for standardised eye tracking research in reading research

Akshay Mendhakar¹, Helmut Leder²

¹Laboratoire de Psychologie Sociale et Cognitive, Université Clermont Auvergne, ²EVA labs, University of Vienna

*contact: akshay.mendhakar@uca.fr

Eye-tracking research in reading relies on fine-grained measures of visual attention, such as fixations and regressions to index cognitive processing during text comprehension (Rayner, 2009; Reichle et al., 2003). However, the field remains fragmented due to differences in experimental design, preprocessing pipelines and reporting practices, which limit cross-study comparability and reproducibility across laboratories and experimental contexts.

To address these challenges, we introduce *EvaTrack*, a unified all-purpose framework designed to support homogeneous eye-tracking for all eye tracking tasks, including while-reading research. *EvaTrack* integrates experimental presentation, automated preprocessing and feature extraction within a single pipeline implementing standardised procedures for gaze filtering, event detection and alignment with a variety of stimulus types (Duchowski, 2007; Holmqvist et al., 2011). The tool provides a consistent data harmonisation and metadata annotation to ensure interoperability across datasets. Emphasis is placed on reproducibility, with transparent parameter settings and documented workflows (Andersson et al., 2017; Hessels et al., 2018; Holmqvist et al., 2012).

Preliminary evaluations of *EvaTrack* on webcam-based eye-tracking datasets (N = 74 participants) were collected under controlled reading conditions. Across datasets, the system achieved a mean spatial accuracy of 2.3°–2.6° and precision of 1.9°–2.2°, consistent with recent webcam-based systems approaching laboratory standards (Kaduk et al., 2023; Papoutsaki et al., 2016; Semmelmann & Weigelt, 2017; Yang & Krajbich, 2021). Raw gaze samples processed with *EvaTrack* showed correlations of 0.80–0.91 with reference pipelines. Importantly, *EvaTrack* reduced preprocessing-induced variability in fixation duration estimates by 18–27% and improved cross-lab consistency of regression measures by 15–22%, with fixation detection agreement exceeding 85%. These results demonstrate that *EvaTrack* provides a robust and reproducible pipeline for eye-tracking research in reading. By standardizing preprocessing and feature extraction, the framework reduces methodological variability and supports reliable cross-lab and ecological valid analyses. It also facilitates faster and more reliable extraction of standard reading measures with standard features of classical eye tracking methods. Further case studies using multilingual reading datasets show that the tool can successfully harmonise data across languages and experimental paradigms, enabling cross-linguistic analyses.

EvaTrack contributes to ongoing efforts in the eye-tracking-while-reading community to standardise methodologies and promote open, reproducible research. By providing a unified pipeline, the tool lowers technical barriers and enables researcher to focus on theoretical and empirical questions rather than implementation details.

References

- Andersson R., Larsson L., Holmqvist K., Stridh M., Nyström M. (2017). One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior Research Methods* 49(2), 616–637.
- Duchowski A. T. (2007). *Eye tracking methodology: Theory and practice*. Springer.
- Hessels R. S., Niehorster D. C., Nyström M., Andersson R., Hooge I. T. C. (2018). Is the eye-movement field confused about fixations and saccades? A survey among 124 researchers. *Royal Society Open Science* 5(8), 180502.
- Holmqvist K., Nyström M., Andersson R., Jarodzka H., van de Weijer J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Holmqvist K., Nyström M., Mulvey F. (2012). Eye tracker data quality: What it is and how to measure it. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 45–52.
- Kaduk T., Goeke C., Finger H., König P. (2024). Webcam eye tracking close to laboratory standards: Comparing a new webcam-based system and the EyeLink 1000. *Behavior Research Methods* 56, 5002–5022.
- Papoutsaki A., Sangkloy P., Laskey J., Daskalova N., Huang J., Hays J. (2016). WebGazer: Scalable webcam eye tracking using user interactions. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 3839–3845.



- Rayner K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology* 62(8), 1457–1506.
- Reichle E. D., Rayner K., Pollatsek A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences* 26(4), 445–476.
- Semmelmann K., Weigelt S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods* 50(2), 451–465.
- Yang X., Krajbich I. (2021). Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making* 16(6), 1485–1505.



Surprisal by design: A principled pipeline for human-LLM stimulus pairs in eye-tracking reading experiments

Isabelle Caroline Rose Cretton¹, David Robert Reich^{1,2}, Lena Ann Jäger¹

¹Department of Computational Linguistics, University of Zurich, ²Department of Computer Science, University of Potsdam

*contact: isabellecarolinerose.cretton@uzh.ch

Behavioural studies of reading often rely on surprisal as a linking hypothesis between language model predictions and fixation durations (Smith and Levy, 2013; Levy, 2008). This framework was originally operationalised using n-gram and probabilistic phrase-structure models of human language (Hale, 2001; Boston et al., 2008; Demberg and Keller, 2008). Yet no principled methodology exists for constructing eye-tracking stimuli that isolate generation source as an experimental variable. Existing human vs LLM text corpora are observational: uncontrolled for topic, length, register, and crucially, not guaranteed to exhibit a positive predictability gap. We present a pipeline that treats stimulus construction as an engineering problem with explicit quality constraints, and solves it.

The pipeline is built around several interlocking design principles that distinguish it from ad hoc generation approaches. Surprisal delta, the difference in bits/token between human and generated text, is estimated by two independent judges that play no role in generation, any pair of language models can serve these roles; a hard invariant enforces agreement at the pair level, making the primary experimental variable model-agnostic by construction. The pipeline provides a framework for diagnosing and addressing corpus-specific failure modes through targeted pre- and post-generation filters; two instantiations are demonstrated here. In factual, entity-dense prose, formulaic register systematically inverts the surprisal direction under news-trained language models, caught by a surprisal-based pre-filter before generation. In L2 argumentative writing, the predictability gap is structurally guaranteed but generation tends toward register drift and essay-mode artefacts, caught by a pattern-level validation layer after generation, with filter logic modular and extensible to any prose type. In this instantiation, generation across both corpora is pooled across model scales and the best output per source text is selected by a composite score weighting surprisal strength, length fidelity, semantic overlap, syntactic similarity, and information density uniformity. Scientific constraints are enforced as hard blocking assertions in a dedicated test suite rather than as post-hoc recommendations, and full reproducibility is ensured via a seeded generation chain spanning sampling, model inference, and dataset construction.

As a proof of concept applied to two typologically distinct corpora, the pipeline yields fully validated stimulus pairs with a mean surprisal delta of 0.957 bits/token, dual-judge agreement on every pair, BERTScore F1 of 0.821, and length ratios above 0.87, all passing a blocking test suite that enforces scientific invariants as hard assertions rather than recommendations. The pipeline is corpus-agnostic, fully reproducible via seeded generation, and releases both stimuli and tooling for the community.

References

- Boston M. F., Hale J. T., Kliegl R., Patil U., Vasishth S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the potsdam sentence corpus. *Journal of Eye Movement Research* 2. doi:<https://doi.org/10.16910/JEMR.2.1.1>.
- Demberg V., Keller F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109(2), 193–210. doi:<https://doi.org/10.1016/j.cognition.2008.07.008>.
- Hale J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*, North American Chapter of the Association for Computational Linguistics, USA, pp. 1–8. Association for Computational Linguistics. doi:<https://doi.org/10.3115/1073336.1073357>.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition* 106(3), 1126–1177. doi:<https://doi.org/10.1016/j.cognition.2007.05.006>.
- Smith N. J., Levy R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition* 128(3), 302–319. doi:<https://doi.org/10.1016/j.cognition.2013.02.013>.



Methodological and accessibility challenges in eye-tracking research with Deaf and hard of hearing

Jurgita Motiejūnienė, Ramunė Kasperė, Vilma Linkevičiūtė, Valdas Grigaliūnas, Paulius Budreika

Kaunas University of Technology

*contact: ramune.kaspere@ktu.lt

Eye tracking has become an important method for investigating perceptual, cognitive, and linguistic processing, offering objective insights into visual attention, reading behaviour, audiovisual processing, and user interaction. In research involving deaf and hard of hearing (DHH) participants, eye tracking is particularly valuable because vision often plays a central role in communication, language access, and information processing. At the same time, conducting eye-tracking studies with DHH populations raises specific methodological and accessibility challenges that remain insufficiently addressed in current literature.

Media accessibility for DHH is one of the fundamental objectives of the European Union, which is explicitly addressed in the European Accessibility Act ([Directive \(EU\) 2019/882, 2019](#)). Promoting the inclusion of people with disabilities, the Act emphasises the significance of research aimed at these target groups. Research involving DHH participants is influenced by several factors, including the considerable heterogeneity of the DHH community and communication barriers. These challenges may contribute to the underrepresentation of DHH individuals in research studies ([Motiejūnienė, 2025](#)). Therefore, guidelines and recommendations for researchers seeking to investigate these populations might be useful in standardizing eye-tracking data collection protocols and fostering collaborative environments.

Our study examines methodological practices in eye-tracking research involving DHH participants and offers guidelines for future work. We discuss eye-tracking research design, drawing on recent literature, ongoing international initiatives and our own experience conducting eye-tracking studies with DHH participants. We draw particularly on the MultiplEYE text corpus and comprehension methodology ([Jakobi et al., 2025](#); [Kasperė et al., 2026](#)). The proposed methodological guidelines cover participant profiling, inclusive recruitment strategies, informed consent procedures, communication during testing, adaptation of instructions into plain language, stimulus design, subtitles, sign language video integration, multimodal layouts, etc.

This work is part of the ongoing EyeSign project (LMTLT, no. S-MIP-24-32) which examines the accessibility and acceptability of machine and AI-supported translation for DHH in Lithuania.

References

- Directive (EU) 2019/882 (2019). *Directive of the European Parliament and of the Council of 17 April 2019 on the accessibility requirements for products and services*. Official Journal of the European Union, L 151, 70–115. URL: <http://data.europa.eu/eli/dir/2019/882/oj>.
- Jakobi D. N., Stegenwallner-Schütz M., Hollenstein N., Ding C., Kasperė R., Matic' Š' koric' A., others, Ja'ger L. A. (2025, May). Multipl-eye: Creating a multilingual eye-tracking-while-reading corpus. In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications*, pp. 1–11. URL: <https://doi.org/10.1145/3715669.3726843>, doi:10.1145/3715669.3726843.
- Kasperė R., Bondar A., Nisioi S., Stegenwallner-Schütz M., Søndergaard Knudsen H. B., Matic' Š' koric' A., Pavlinus'ic' Vilus E., others, Ja'ger L. A. (2026, March). The multipleye text corpus: Towards a diverse and ever-expanding multilingual text corpus. In *Proceedings of the 15th Language Resources and Evaluation Conference (LREC)*. URL: <https://lrec2026.info>.
- Motiejūnienė J. (2025). Reaching deaf and hard-of-hearing people in audiovisual research. *Specijalna edukacija i rehabilitacija* 24(2), 181–194. URL: <https://doi.org/10.5937/specedreh24-52733>, doi:10.5937/specedreh24-52733.