MultiplEYE: Creating a multilingual eye-tracking-while-reading corpus

Deborah Noemie Jakobi
Department of Computational Linguistics
University of Zurich
Zurich, Switzerland
jakobi@cl.uzh.ch

Maja Stegenwallner-Schütz
Humboldt-University of Berlin
Berlin, Germany
University of Koblenz
Koblenz, Germany
maja.stegenwallner-schuetz@hu-berlin.de

Nora Hollenstein
Department of Computational Linguistics
University of Zurich
Zurich, Switzerland
nora.hollenstein@uzh.ch

Cui Ding
Department of Computational Linguistics
University of Zurich
Zurich, Switzerland
cui.ding@uzh.ch

Ramune Kaspere
Kaunas University of Technology
Kaunas, Lithuania
ramune.kasperaviciene@gmail.com

Ana Matić Škorić
Faculty of Education and Rehabilitation Sciences, Dept of Speech and Language pathology, Laboratory for Psycholinguistic Research
University of Zagreb
Zagreb, Croatia
ana.matic@erf.unizg.hr

Eva Pavlinusic Vilus
Faculty of Education and Rehabilitation Sciences, Dept of Speech and Language pathology, Laboratory for Psycholinguistic Research
University of Zagreb
Zagreb, Croatia
eva.pavlinusic.vilus@erf.unizg.hr

Stefan Frank
Centre for Language Studies
Radboud University
Nijmegen, Netherlands
stefan.frank@ru.nl

Marie-Luise Müller
ZPID Leibnitz Institute for Psychology
Trier, Germany
mm@leibniz-psychology.org

Kristine M Jensen de López
Aalborg University
Aalborg, Denmark
kristine@ikp.aau.dk

Nik Kharlamov
Aalborg University
Aalborg, Denmark
nik@ikp.aau.dk

Hanne B. Søndergaard Knudsen
Aalborg University
Aalborg, Denmark
hannebsk@ikp.aau.dk

Yevgeni Berzak
Data and Decision Sciences
Technion
Haifa, Israel
berzak@technion.ac.il

Ella Lion
Data and Decision Sciences
Technion
Haifa, Israel
ellal@technion.ac.il

Irina A. Sekerina
Department of Psychology
College of Staten Island
Staten Island, New York, USA
irina.sekerina@csi.cuny.edu

Cengiz Acarturk
Cognitive Science Department
Jagiellonian University
Krakow, Poland
cengiz.acarturk@uj.edu.pl

Mohd Faizan Ansari
Silesian University of Technology
Gliwice, Poland
mfansari2395@gmail.com

Katarzyna Harezlak
Institute of Informatics
Silesian University of Technology
Gliwice, Poland
katarzyna.harezlak@polsl.pl

Pawel Kasprowski
Institute of Informatics
Silesian University of Technology
Gliwice, Poland
kasprowski@polsl.pl

Ana Bautista
Basque Center on Cognition, Brain & Language
Donostia, Spain
a.bautista@bcbl.eu

Lisa Beinborn
Human-Centered Data Science
University of Göttingen
Göttingen, Germany
l.m.beinborn@uva.nl

Anna Bondar
Department of Computational Linguistics
University of Zurich
Zurich, Switzerland
Digital Society Initiative
University of Zurich
Zurich, Switzerland
anna.bondar@uzh.ch

Antonia Boznou
Lab of Psycholinguistics & Neurolinguistics
National and Kapodistrian University of Athens
Athens, Greece
antoniaboznou@gmail.com

Leah Bradshaw
Department of Computational Linguistics
University of Zurich
University of Zurich, Switzerland
leah.bradshaw@uzh.ch

Jana Mara Hofmann
Department of Computational Linguistics
University of Zurich
Zurich, Switzerland
janamara.hofmann@uzh.ch

Thyra Krosness
Department of Computational Linguistics
University of Zurich
Zurich, Switzerland
thyrasigynhedda.krosness@uzh.ch

Not Battesta Soliva
University of Zurich
Zurich, Switzerland
notbattesta.soliva@uzh.ch

Anila Çepani
University of Tirana
Tirana, Albania
anilacepani@gmail.com

Kristina Cergol
Faculty of Teacher Education
University of Zagreb
Zagreb, Croatia, Hrvatska, Croatia
Laboratory for Psycholinguistic Research
University of Zagreb
Zagreb, Croatia, Hrvatska, Croatia
kristina.cergol@ufzg.hr

Ana Došen
University of Zagreb
Zagreb, Croatia
ana.dosen@erf.unizg.hr

Marijan Palmovic
Dept. of Speech & Lang. Path., Laboratory for Psycholinguistic Research
University of Zagreb
Zagreb, Croatia, Hrvatska, Croatia
marijan.palmovic@erf.unizg.hr

Adelina Çerpja
Academy of Sciences of Albania
Tirana, Albania
adicerpja@gmail.com

Dalí Chirino
Radboud University
Nijmegen, Netherlands
dali.chirino@ru.nl

Jan Chromý
Institute of Czech Language and Theory of Communication
Faculty of Arts, Charles University
Prague, Czech Republic
jan.chromy@ff.cuni.cz

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

Iza Škrjanec
Saarland University
Saarbruecken, Germany
skrjanec@lst.uni-saarland.de

Nazik Dinçtopal Deniz
Bogazici University
Istanbul, Turkiye
nazik.dinctopal@bogazici.edu.tr

Dr. Inmaculada Fajardo
Reading Research Unit/ Dept. Developmental and Educational Psychology
University of Valencia
Valencia, Valencia, Spain
inmaculada.fajardo@uv.es

Mariola Giménez-Salvador
Universitat de València
València, Spain
mariola.gimenez@uv.es

Xavier Mínguez-López
Universitat de València
València, Spain
xavier.minguez@uv.es

Maroš Filip
Institute of Psychology
Czech Academy of Sciences
Prague, Czech Republic
filipmaross@gmail.com

Zigmunds Freibergs
Institute of Public Health
Riga Stradins University
Riga, Latvia
University of Latvia
Riga, Latvia
zigmunds.freibergs1@gmail.com

Jéssica Gomes
Center of Linguistics, School of Arts and Humanities
University of Lisbon
Lisbon, Portugal
jgomes4@campus.ul.pt

Andreia Janeiro
Center of Linguistics, School of Arts and Humanities
University of Lisbon
Lisbon, Portugal
andreia-janeiro@edu.ulisboa.pt

Paula Luegi
Center of Linguistics, School of Arts and Humanities
University of Lisbon
Lisbon, Portugal
paularibeiro@edu.ulisboa.pt

João Veríssimo
Center of Linguistics, School of Arts and Humanities
University of Lisbon
Lisbon, Portugal
jlverissimo@edu.ulisboa.pt

Sasho Gramatikov
Ss. Cyril and Methodius University in Skopje
Skopje, Macedonia
sasho.gramatikov@finki.ukim.mk

Jana Hasenäcker
Universität Erfurt
Erfurt, Germany
jana.hasenaecker@uni-erfurt.de

Alba Haveriku
Faculty of Information Technology, Polytechnic University of Tirana
Tirana, Albania
alba.haveriku@fti.edu.al

Nelda Kote
Polytechnic University of Tirana
Tirana, Albania
nkote@fti.edu.al

Muhammad M. Kamal
NLP Research Centre, Knowledge Platform
Islamabad, Pakistan
mkamal@knowledgeplatform.com

Hanna Kędzierska
University of Wroc?aw
Wroclaw, Poland
hanna.kedzierska2@uwr.edu.pl

Dorota Klimek-Jankowska
University of Wroc?aw
Wroclaw, Lower Silesia, Poland
Dorota.klimek-jankowska@uwr.edu.pl

Sara Kosutar
Department of Language and Culture
UiT The Arctic University of Norway
Tromso, Norway
sara.kosutar@uit.no

Daniel G. Krakowczyk
University of Potsdam
Potsdam, Germany
daniel.krakowczyk@uni-potsdam.de

Izabela Krejtz
Psychology Department
SWPS University of Social Sciences and Humanities
Warsaw, Poland
ikrejtz@swps.edu.pl

Marta Łockiewicz
University of Gdansk
Gdanks, Poland
marta.lockiewicz@ug.edu.pl

Kaidi Lõo
University of Tartu
Tartu, Estonia
kaidi.loo@ut.ee

Jurgita Motiejūnienė
Kaunas University of Technology
Kaunas, Lithuania
jurgita.motiejuniene@ktu.lt

Jamal A. Nasir
University of Galway
Galway, Ireland
jamal.nasir@universityofgalway.ie

Johanne Sofie Krog Nedergård
University of Copenhagen
Copenhagen, Denmark
jskn@hum.ku.dk

Ayşegül Özkan
Independent researcher
Ankara, Turkiye
aysglozkn@gmail.com

Mikuláš Preininger
Faculty of Arts
Charles University
Prague, Czech Republic
Institute of Psychology
Czech Academy of Sciences
Prague, Czech Republic
mikulas.preininger@ff.cuni.cz

Loredana Pungă
West University of Timi?oara
Timisoara, Romania
loredana.punga@e-uvt.ro

David Robert Reich
Universität Potsdam
Potsdam, Germany
Department of Computational Linguistics
University of Zurich
Zurich, Switzerland
david.reich@uni-potsdam.de

Chiara Tschirner
University of Potsdam
Potsdam, Germany
Department of Computational Linguistics
University of Zürich
Zürich, Switzerland
tschirner@uni-potsdam.de

Špela Rot
University of Ljubljana
Ljubljana, Slovenia
spelca.rot@gmail.com

Andreas Säuberli
LMU Munich & MCML
Munich, Germany
andreas.saeuberli@lmu.de

Jordi Solé-Casals
University of Vic - Central University of Catalonia
Vic, Spain
jordi.sole@uvic.cat

Ekaterina Strati
Aleksandër Moisiu University of Durrës
Durrës, Albania
ekaterinastrati@uamd.edu.al

Igor Svoboda
National Technical University of Ukraine
Kyiv, Ukraine
i.svoboda@kpi.ua

Evis Trandafili
University of New York in Tirana
Tirana, Albania
evis.trandafili@gmail.com

Spyridoula Varlokosta
Linguistics/Pgilology/Psycholinguistics and Neurolinguistics Lab
National and Kapodistrian University of Athens
Athens, Greece
Archimedes
Athena RC
Athens, Greece
svarlokosta@phil.uoa.gr

Mila Vulchanova
Language Acquisition and Language Processing Lab
Norwegian University of Science & Technology
Trondheim, Norway
mila.vulchanova@ntnu.no

Lena A. Jäger
Department of Computational Linguistics
University of Zurich
Zurich, Switzerland
jaeger@cl.uzh.ch

Abstract

Eye-tracking-while-reading data provide valuable insights across multiple disciplines, including psychology, linguistics, natural language processing, education, and human-computer interaction. Despite its potential, the availability of large, high-quality, multilingual datasets remains limited, hindering both foundational reading research and advancements in applications. The MultiplEYE project

ETRA '25, Tokyo, Japan

ACM ISBN 979-8-4007-1487-0/25/05
https://doi.org/10.1145/3715669.3726843

addresses this gap by establishing a large-scale, international eye-tracking data collection initiative. It aims to create a multilingual dataset of eye movements recorded during natural reading, balancing linguistic diversity, while ensuring methodological consistency for reliable cross-linguistic comparisons. The dataset spans numerous languages and follows strict procedural, documentation, and data pre-processing standards to enhance eye-tracking data transparency and reproducibility. A novel data-sharing framework, integrated with data quality reports, allows for selective data filtering based on research needs. Researchers and labs worldwide are invited to join the initiative. By establishing and promoting standardized practices and open data sharing, MultiplEYE facilitates interdisciplinary research and advances reading research and gaze-augmented applications.

CCS Concepts

• **Applied computing → Psychology**.

Keywords

Eye-tracking, reading, psycholinguistics, multilingual, open science

ACM Reference Format:

Deborah Noemie Jakobi, Maja Stegenwallner-Schütz, Nora Hollenstein, Cui Ding, Ramune Kaspere, Ana Matić Škorić, Eva Pavlinusic Vilus, Stefan Frank, Marie-Luise Müller, Kristine M Jensen de López, Nik Kharlamov, Hanne B. Søndergaard Knudsen, Yevgeni Berzak, Ella Lion, Irina A. Sekerina, Cengiz Acarturk, Mohd Faizan Ansari, Katarzyna Harezlak, Pawel Kasprowski, Ana Bautista, Lisa Beinborn, Anna Bondar, Antonia Boznou, Leah Bradshaw, Jana Mara Hofmann, Thyra Krosness, Not Battesta Soliva, Anila Çepani, Kristina Cergol, Ana Došen, Marijan Palmovic, Adelina Çerpja, Dalí Chirino, Jan Chromý, Vera Demberg, Iza Škrjanec, Nazik Dinçtopal Deniz, Dr. Inmaculada Fajardo, Mariola Giménez-Salvador, Xavier Mínguez-López, Maroš Filip, Zigmunds Freibergs, Jéssica Gomes, Andreia Janeiro, Paula Luegi, João Veríssimo, Sasho Gramatikov, Jana Hasenäcker, Alba Haveriku, Nelda Kote, Muhammad M. Kamal, Hanna Kędzierska, Dorota Klimek-Jankowska, Sara Kosutar, Daniel G. Krakowczyk, Izabela Krejtz, Marta Łockiewicz, Kaidi Lõo, Jurgita Motiejūnienė, Jamal A. Nasir, Johanne Sofie Krog Nedergård, Ayşegül Özkan, Mikuláš Preininger, Loredana Pungă, David Robert Reich, Chiara Tschirner, Špela Rot, Andreas Säuberli, Jordi Solé-Casals, Ekaterina Strati, Igor Svoboda, Evis Trandafili, Spyridoula Varlokosta, Mila Vulchanova, and Lena A. Jäger. 2025. MultiplEYE: Creating a multilingual eye-tracking-while-reading corpus. In *2025 Symposium on Eye Tracking Research and Applications (ETRA '25), May 26–29, 2025, Tokyo, Japan.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3715669.3726843

1 Introduction

Eye-tracking-while-reading data are a valuable resource for various disciplines, including psychology, linguistics, natural language processing (NLP), education, and human-computer interaction. Its versatility and potential have made eye-tracking an invaluable tool in numerous fields. In cognitive psychology [Hessels and Hooge 2019] and linguistics [Conklin and Pellicer-Sánchez 2016], it is considered the gold-standard dependent variable for studying reading processes [Rayner 1998]. In computational linguistics, eye-tracking data provide human feedback to better align large language models with human linguistic preferences [Deng et al. 2024]. In education, researchers are exploring the potential of eye-tracking data to enhance and automate reading competence assessments [Berzak et al. 2018; Halszka et al. 2017; Reich et al. 2022].

While researchers often publish the eye-tracking data collected for a particular study for secondary scientific use, the need for additional data continues to grow. First, across disciplines, from foundational psycholinguistic reading research to machine learning-driven applications, a scarcity of eye-tracking data remains a major bottleneck. In psycholinguistics, more data is needed to increase statistical power, while for innovative machine learning methods, complex and powerful models require large datasets to be trained. Currently, there are hardly any large *multilingual* datasets available. An exception is the Multilingual Eye-Movement Corpus (MECO-L1, [Siegelman et al. 2022]. The MultiplEYE corpus presented in this work differs from MECO-L1 in terms of the genres of text, the assessment of reading comprehension, and the protocols that are applied to ensure data consistency, among other aspects. Therefore, the MultiplEYE corpus complements existing datasets well and adds more diversity to available eye-tracking datasets. Multilingual parallel corpora are necessary for cross-linguistic comparisons and for evaluating theories of language processing, oculomotor control, and other cognitive processes across speakers of different languages. Along the same lines, naturalistic reading data, especially from diverse text types, allow for evaluating the ecological validity of existing theories and models and their ability to generalize across a wide range of diverse reading materials.

Second, many of the datasets from previous eye-tracking studies offer only limited re-usability and do not allow for types of analyses different from the original ones. The main problem is the heterogeneity of the data because they are generated from discipline-specific perspectives. Typically, eye-tracking data are created in formats designed for proprietary analyses, which can vary substantially between subfields. Data pre-processing is often intransparent, as many researchers rely on proprietary, black-box software or focus only on the specific analyses relevant to their study. Consequently, many existing data sets provide only final aggregated measures, such as reading times or fixation counts, rather than raw eye-tracking data. This not only makes it nearly impossible to build upon previous work because analyses and results are difficult to reproduce but also hinders the re-use of the data for different types of analyses that require different data formats. For example, a typical psycholinguistic dataset providing word-level reading measures, such as first-pass reading times and regression counts, cannot be re-used to analyze the readers' scanpaths [von der Malsburg and Vasishth 2011] or for developing neural sequence architectures for eye-tracking-while-reading data [Bolliger et al. 2023; Deng et al. 2023; Sood et al. 2020]. The situation is exacerbated by a lack of commonly agreed-upon data sharing and publication practices within the eye-tracking-while-reading community. Data that were deemed irrelevant for the primary purpose of the study are often excluded, which heavily limits their re-usability for new research questions or meta-analyses. For example, data with suboptimal calibration are often discarded, yet are necessary to develop algorithms for post-hoc drift correction [Jakobi et al. 2024]. Likewise, noisy raw data are crucial for developing and testing robust algorithms for gaze event detection (e.g., detecting fixations). Moreover, initiatives that combine data from multiple labs often face challenges due to differences in lab or experiment setups — such as variations in eye-tracking hardware, calibration procedures, display

settings, and participant positioning. These inconsistencies can confound cross-linguistic comparisons and limit the re-usability of the data. To enhance the re-usability of eye-tracking-while-reading data, it is crucial for new datasets to adopt standardized procedures and protocols that ensure consistency across all data sources.

In this work, we present the MultiplEYE initiative which 1) involves the collection of a multilingual eye-tracking-while-reading dataset, and 2) aims to establish new standards for the collection, documentation, pre-processing, and sharing of multilingual eye-tracking-while-reading data.

2 The MultiplEYE data collection

The MultiplEYE project is an international, multi-lab eye-tracking data collection initiative that aims to create a large, multilingual dataset of eye movements recorded while reading natural texts from a parallel corpus.

The MultiplEYE project covers different high- and low-resource languages. Currently, the following languages are included in MultiplEYE: Albanian, Arabic, Basque, Catalan, Chinese, Croatian, Czech, Danish, Dutch, English, Estonian, French, Farsi, German, Greek, West Greenlandic (Kalaallisut), Hebrew, Hindi, Italian, Latvian, Lithuanian, Macedonian, Norwegian, Polish, Portuguese, Romanian, Romansh, Russian, Serbian, Slovenian, Spanish, Swedish, Turkish, Ukrainian, and Urdu. A second major contribution is that the entire data collection follows strict standards with regard to the procedure, metadata documentation, data pre-processing, data sharing, and data publication. By integrating these standards into MultiplEYE, we not only ensure consistency within our dataset but also advance standardized data-sharing processes and Open Research Data (ORD) practices to enhance the re-usability of eye-tracking datasets more broadly. To facilitate widespread adoption of these standards, we provide comprehensive guidelines, templates, and forms, that are openly available for re-use. More details are described in the project's Data Management Plan [Müller et al. 2024]. In addition, we have developed a new pre-processing pipeline that includes the development of data quality reports, allowing us to share all data without exclusions. In many studies, data of insufficient quality are excluded and therefore not published, with the data exclusion criteria remaining unspecified. Our use of data quality reports accompanying the data enables the publication of all data, while still allowing users to exclude selected data for a particular use case. Users can choose subsets based on the level of quality they require, as indicated by the data quality reports. The MultiplEYE data are stored in the newly established database and data sharing platform EyeStore (https://rdc-psychology.org/en/eyestore), which enables users to filter the data based on, for example, the quality reports, language, or other meta-data. New eye-tracking data beyond the MultiplEYE samples can be added to EyeStore in the future.

The MultiplEYE project strives to balance diversity and consistency, two key principles that are challenging to achieve simultaneously. It embraces diversity in order to enable greater generalizability and ecological validity by accommodating a wide range of languages, scripts, text types, participant backgrounds, and eye-tracking equipment. Besides generalizability, allowing diversity in eye-tracking equipment is essential to enable as many labs as possible to contribute. Ensuring consistency allows for reliable cross-linguistic comparisons of reading behavior by minimizing confounding variables, such as differences in stimulus layout or lab setup.

2.1 Procedure

The MultiplEYE data collection consists of two sessions: (1) an eye-tracking-while-reading experiment followed by a participant questionnaire, and (2) a psychometric assessment. The first session is mandatory and includes the collection of the eye-tracking data followed by a short participant questionnaire with questions about the participant's demographics (e.g., age), (socio-)linguistic background (e.g., native language(s)), and current physical state (e.g., tiredness). The second session is optional but highly encouraged for the participating labs. It includes nine psychometric assessments of cognitive measures that have been found to correlate with individual reading behavior [Cunnings and Felser 2013; Fedorenko 2014; Haller et al. 2024; Kuperman et al. 2018; Nicenboim et al. 2016; Novick et al. 2014] and are therefore a valuable addition to the eye-tracking data. Figure 1 visualizes the procedure.

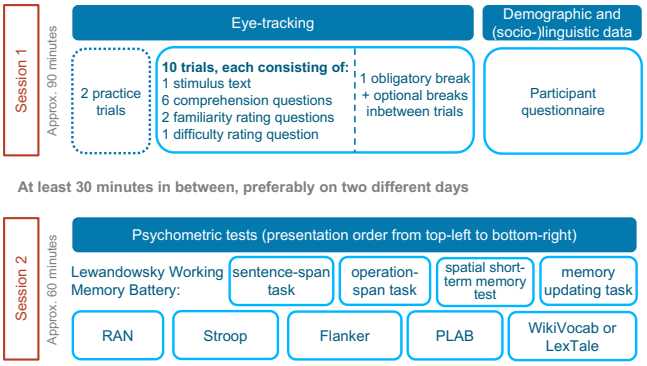

Figure 1: MultiplEYE experimental procedure with two sessions: In Session 1, eye movements are recorded during reading, accompanied by text comprehension and rating questions, and followed by a participant questionnaire. Session 2, which is optional but recommended, consists of a series of psychometric assessments.

2.1.1 Session 1: Eye-tracking and participant questionnaire. In the eye-tracking session, participants are presented with ten natural texts, each followed by six multiple-choice comprehension questions and three rating questions (two familiarity ratings and one perceived difficulty rating). Two practice trials precede the session. The session requires one obligatory break and as many optional breaks as necessary. The last part is the participant questionnaire.

The detailed session procedure is described in full in the experimenter script (https://tinyurl.com/tehp59w2). In order to achieve consistency in data quality across labs, there are recommendations for calibration and validation, as well as practical tips on how to achieve good calibration accuracy. Each session is documented in a session documentation sheet (https://tinyurl.com/uv6x8cea), which includes notes and relevant session-specific metadata.

2.1.2 Session 2: Psychometric assessment. In the second, optional, session, a psychometric assessment is performed consisting of nine computerized tests. The order of tests is fixed to enhance comparability between participants. The participant can automatically proceed from one test to the next without constant supervision. In the order of presentation, the tests are: The Lewandowsky Working Memory Capacity (WMC) Battery [Lewandowsky et al. 2010] comprising four tests (a sentence-span task, an operation-span task, a spatial short-term memory task, and a memory updating task), the Rapid Automatized Naming (RAN) Task (processing speed and efficiency in retrieving information from memory) [Denckla and Rudel 1974], the Stroop Task (cognitive control and inhibition) [Stroop 1935], the Flanker Task (attentional control) [Eriksen and Eriksen 1974], the Pimsleur Language Aptitude Battery (PLAB) Test (language learning aptitude) [Pimsleur 1966], the WikiVocab Test [van Rijn et al. 2023] or LexTALE, where available [Lemhöfer and Broersma 2012] (lexical knowledge).

As for the first session, experimenters consult an experimenter script (https://tinyurl.com/tehp59w2) describing the session procedure to ensure consistent administration across all labs.

2.2 Materials of the MultiplEYE natural reading experiment

The stimulus corpus consists of **ten natural texts** from the 20th and 21st centuries, along with two practice texts, covering a range of text types: fiction, popular science, institutional texts, and argumentative multi-document items. These texts were considered suitable for the MultiplEYE collection because they originated in various languages and had existing translations available in multiple languages. All materials will be linguistically annotated with regard to relevant lexical, syntactic, and morphological features. Following Kintsch [1998], a set of **questions was designed to assess the reader's text comprehension** at various levels. These questions are organized into three conditions that target different levels of text comprehension, ranging from literal text-based comprehension to the integration of prior knowledge in constructing a situation model. All questions are presented in a multiple-choice format with four answer options. Crucially, we adopted the annotation scheme by Berzak et al. [2020] for systematically designing distractors (i.e., plausible yet incorrect answer options) for each of the three question types. Additionally, following this scheme, we annotated the text spans that correspond to the information referenced by specific answer options. We also designed three **rating questions** for each text, in order to gather information about participants' responses, specifically in relation to their eye movements. Those include two types of text familiarity questions and one question assessing the reader's perceived text difficulty.

All the materials were prepared in English in a multistage collaborative effort including an alternation of item evaluation experiments (in-lab and online) and review rounds by professional text editors. The English version serves as the foundation for translations into other languages created by MultiplEYE.

2.3 Coordination and Communication

To effectively organize and coordinate this large-scale data collection effort, we have implemented several practices to help navigate the data collection and to make sure that all problems can be solved in a timely fashion, and that expertise and experiences are shared across the labs:

(1) Working group meetings: Regular online meetings are held for all those involved in the data collection, where we discuss current issues, provide updates on changes, and share best practices. Labs already collecting data are encouraged to share their experiences. Additionally, several times a year, we organize on-site meetings in varying locations, including training schools, workshops, and hackathons, to foster the exchange of knowledge and collaborative work on the project.

(2) Office hours for stimulus translation and technical problems: Each week, a dedicated time slot is available via an online video conference platform with project members coordinating the translation efforts, during which any MultiplEYE member can reach out with questions regarding the translation of stimulus materials. This ensures that any translation-related issues can be addressed promptly, fostering consistency across the project's multilingual stimulus materials. Similarly, office hours for technical problems encountered during the local installation of the experiment allow to address such problems in a timely fashion and support participating labs to be able run the experiment.

(3) pre-processing meetings in collaboration with *pymovements*: The pre-processing pipeline is developed by members of MultiplEYE in collaboration with contributors to the *pymovements* Python package [Krakowczyk et al. 2023]. Regular meetings are held twice a month (see the *pymovements* website), and workshops, hackathons, and additional meetings are organized as needed to discuss and refine the pre-processing pipeline and procedure.

(4) Collaborative Software Development: To facilitate efficient software development for the experiment implementation and ensure transparent version control, we use GitHub. This platform allows MultiplEYE members to share resources, collaborate on code, and document procedures.

(5) Pre-registration: We utilize pre-registration to define responsibilities, maintain an overview of the progress of data collection, and specify lab setups. Additionally, pre-registration clarifies whether collected data will be integrated into the MultiplEYE corpus or considered an add-on dataset.

(6) Image Generation: A dedicated MultiplEYE team is responsible for generating stimulus images for each language while adapting them to the specific setup of each participating lab, using the information from pre-registration. This approach ensures that the pre-processing software can run accurately across datasets from different labs, even when they use different setups.

(7) Experiment implementation and support team: A significant part of MultiplEYE involves the development of the presentation software, which includes various modules. While initial meetings were held to discuss implementation details, the focus has now shifted to providing support for labs collecting data. A support team has been formed, that holds weekly meetings to bundle support, ensure consistency, and distribute tasks to the responsible team members. Given that many support topics are lab-specific, there are no regular open meetings, but support meetings are scheduled when needed with the respective labs. As the data collection

progresses, labs that are further along in the process are now actively helping those in the initial stages, creating a collaborative flow of support.

(8) Data collection newsletter: Given that many labs are collecting data at the same time, it is important to share experiences and successful solutions to any issues that arise. The newsletter serves this purpose by keeping everyone informed about updates, highlighting shared challenges, and providing insights into effective solutions across the labs.

(9) Short-term scientific missions (STSMs): STSMs offer the possibility for researchers who are members of the MultiplEYE COST Action (https://www.cost.eu/actions/CA21131/) to travel to another participating lab for knowledge transfer. The visiting researcher may either teach at the lab, assist with a relevant task such as setting up the experiment or running the data collection, or learn from the host lab to implement new methods or practices in their own lab. Numerous STSMs have been organized throughout the project, helping labs to successfully collect data and fostering collaboration across the network.

(10) Sharing eye-tracking equipment: To facilitate collaboration and ensure participating labs have access to necessary equipment, MultiplEYE members have successfully navigated the complexities of insurance and customs for traveling with eye-tracking equipment.

(11) FAQs in MultiplEYE data collection guidelines: Many questions that have occurred throughout the process are collected in FAQ sections in the MultiplEYE data collection guidelines (https://tinyurl.com/3rpe6m4h). Each section contains FAQs related to the section's topic.

(12) Document hub: On our website, we present an overview of all the documents and resources that are necessary for the data collection (https://multipleye.eu/forms-and-documentation-hub/).

(13) Communication via email and Slack: While important or urgent news are communicated via email, there exists a Slack workspace that can be used for more informal communication and quickly organizing subgroups.

(14) Mandatory submission of pilot data: We have established a standardized piloting procedure to ensure that all data collection practices are aligned across the participating labs. During this phase, pilot data is submitted for review, allowing us to conduct thorough sanity checks and data quality assessments.

(15) Data collection support: For lab-specific questions or problems, the MultiplEYE support team can be reached via email. We aim to distribute tasks quickly to experts in our team in the field of the problem or question, which includes labs that have already concluded parts of the data collection and can provide assistance for other labs. This process ensures a collaborative effort to advance the data collection and exchange of knowledge and expertise.

2.4 Standardization and implementing ORD principles

In order to achieve MultiplEYE's second goal—setting new ORD standards for collecting, documenting, pre-processing, and sharing a large multilingual eye-tracking-while-reading dataset—the project implements key measures. First, the stimulus layout, the presentation size of stimuli, and the eye-to-screen distance are standardized. To further strengthen ORD principles, the MultiplEYE team provides open-source software for stimulus presentation (see https://tinyurl.com/3yhfa7w2), which is compatible with the different eye trackers used in participating labs. Pre-registering each data collection is required, and stimuli images are generated according to each lab's hardware setup. This ensures that stimuli are presented with uniform font and presentation size, resulting in consistent areas-of-interest dimensions across labs. Precise guidelines are provided for the lab setup to ensure a standardized lab environment.

Additionally, the experimental procedure is standardized across labs to ensure data collection follows the same protocol. This is achieved through the MultiplEYE presentation software and experimenter scripts (https://tinyurl.com/tehp59w2 and https://tinyurl.com/22w7d6v7), which guide experimenters. Further standardization is achieved by publishing various tutorials, such as calibrating (https://tinyurl.com/492fpdvr), setting up the MultiplEYE experiment (https://tinyurl.com/js7ed4rh), and performing the dominant eye test (https://tinyurl.com/4v43587j). All software and accompanying materials, including experimenter scripts and tutorials, are openly accessible, ensuring transparency. Moreover, labs receive feedback on the quality of their pilot data. Only when the pilot data meets MultiplEYE's quality standards can the lab proceed with data collection. To enhance transparency and reusability, MultiplEYE has developed comprehensive metadata schemes to ensure data is properly labeled and organized, making it easier for researchers to use in future studies.

Furthermore, the data pre-processing pipeline is standardized and implemented by MultiplEYE members using the open-source Python package *pymovements* [Jakobi et al. 2024; Krakowczyk et al. 2023]. The data is published at all pre-processing stages, including raw sample data, along with the pipeline and quality reports. This approach ensures transparency, reproducibility, and the generalization of the pre-processing pipeline to eye-tracking data beyond MultiplEYE.

3 Contributing to MultiplEYE

The standardization efforts outlined in the previous section lay the foundation for ensuring consistency across participating labs, enabling the collection of eye-tracking-while-reading data across different languages. However, fully addressing the challenges of data scarcity and enhancing the multilingual scope of the dataset requires the participation of new contributors. By expanding the range of languages and datasets included in MultiplEYE data collections, as well as further refining the practices of standardized data preprocessing and archiving, researchers can collectively strengthen cross-linguistic eye-tracking research.

MultiplEYE allows new contributors to join anytime. All stimuli and text materials have been prepared in English and can be translated into any new language. Figure 2 summarizes the necessary steps for contributing a new language or an additional dataset for a language already included in MultiplEYE. Importantly, it is also possible to contribute the stimulus materials for a new language without committing to the collection of eye-tracking data. Once the stimuli exist in the respective language, it is easier to later find a lab

Name and link of the document	Description
MultiplEYE data collection guidelines (https://tinyurl.com/3rpe6m4h)	These guidelines form the backbone of the data collection and pre-processing procedure and implementation. They should be read in full to start and again be consulted and followed for each new step from the beginning of the data collection until the very end.
pre-registration form (https://multipleye.eu/multipleye-pre-registration-form/)	With this form, a data collection can be registered as a MultiplEYE dataset. Researchers are required to submit their lab specifications and provide information about the setup, timeline, and hardware. This information is crucial to be able to properly set up the experiment in the respective lab.
Experimenter script (1^{st} session) (https://tinyurl.com/tehp59w2)	The experimenter script for the first (eye-tracking) session specifies the exact procedure and should be consulted by everyone collecting data. This includes research assistants as well as lab managers and ideally a print-out copy should be available during each session to make sure that the procedure is consistent.
Experimenter script (2^{nd} session) (https://tinyurl.com/22w7d6v7)	Analogous to the first session, the experimenter script for the second session specifies the exact procedure of this session and should be consulted by everyone collecting data before and during each session.
Experimental presentation (https://tinyurl.com/3yhfa7w2)	Our presentation software is made available through a GitHub repository. It contains detailed information about how to set up the experiment locally. Data collectors should consult all READMEs and documents in order to start the data collection.

Table 1: Summary of the most relevant documents for new contributors and what they include.

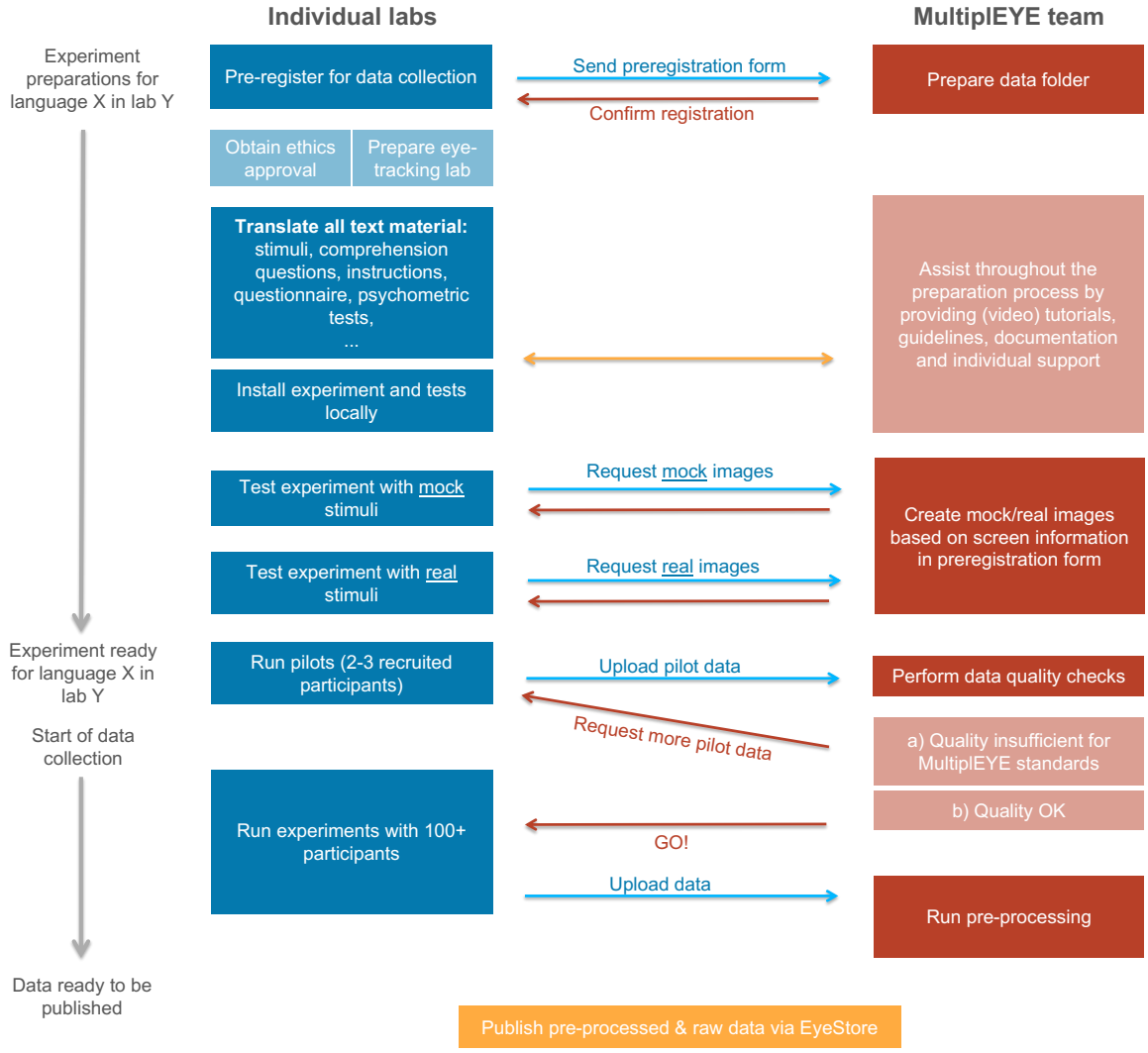

Figure 2: MultiplEYE data collection. The flowchart outlines the steps from experiment preparation to data sharing, including ethics approval, stimulus preparation, pilot testing, data collection, and final data review.

that can collect the data, as a large part of the work has already been contributed. In addition to contributing a new language, it is possible to help develop the tools such as the pre-processing pipeline, or join the technical support team. More detailed information on how to contribute can be found in the MultiplEYE data collection guidelines (https://tinyurl.com/3rpe6m4h). The documents in Table 1 should be consulted as a first step by new or interested contributors.

3.1 Requirements

To ensure consistency and comparability across data collection sites, participating labs must adhere to the following constraints: All data collections must follow the MultiplEYE data collection guidelines (https://tinyurl.com/3rpe6m4h). As a first step, contributors are required to **pre-register the dataset** via our website (https://multipleye.eu/multipleye-pre-registration-form/). While we aim at allowing for as much diversity as possible in terms of eye-tracking device and hardware, the guidelines list a **few hardware and setup requirements** such as the minimal screen size or the eye-to-screen distance which must be controlled to ensure consistent visual alignment of each character on screens across labs. Detailed setup guidelines are being provided to ensure compliance. Further, participating labs must **use the standardized MultiplEYE reading experiment**. Data collected using variants of the MultiplEYE stimuli or participant populations can still be submitted; however, these datasets will be considered add-on datasets, which may limit their direct comparability with the main corpus. Adding a new language to the project requires researchers to **translate the provided stimulus materials** following the MultiplEYE guidelines (https://tinyurl.com/3rpe6m4h). All other aspects of the experiment, including the study design, data collection workflow, and pre-processing pipeline, are already prepared and standardized, ensuring a seamless integration into the dataset. Finally, all participating labs are required to **fill in a session documentation form** for each individual session and complete an additional **data collection metadata form** once the data collection for this lab is completed which summarizes the entire data collection. These documents are crucial as they allow us to handle special cases appropriately and include as much data as possible without discarding any. For example: if one participant could not complete one of the trials, this session can still be included if appropriately labeled and documented.

3.2 Where to start

The MultiplEYE initiative welcomes collaboration from researchers and institutions interested in contributing to the collection of a multilingual eye-tracking-while-reading dataset. Visit our MultiplEYE website (https://multipleye.eu/) for an overview of the project and opportunities for participating. Specifically, there is a page that helps new contributors find all the necessary information: https://multipleye.eu/contribute/. Then, join the MultiplEYE Cost Action (https://www.cost.eu/actions/CA21131/). Once the COST Action is completed but also already now, researchers interested in participating can contact us via email (for more details see https://multipleye.eu/contact/) .

4 Outlook and future work

While the MultiplEYE Cost Action grant period has a fixed duration (until September 2026), the MultiplEYE data collection can still continue. All materials are made available beyond the project's lifetime and crucially, the database and data sharing platform EyeStore (https://rdc-psychology.org/en/eyestore) allows researchers to add new datasets anytime. It is in MultiplEYE's biggest interest that the efforts for standardization, preparing guidelines and templates as well as publishing tools, specifically an eye-tracking pre-processing pipeline, are openly accessible to be used for future research.

Acknowledgments

This work was partially funded by COST Action MultiplEYE (Action number: CA21131), supported by COST (European Cooperation in Science and Technology), the Swiss National Science Foundation under grants 212276 (MeRID, PI: Lena Jäger) and IZCOZ0_220330 (EyeNLG, PI: Lena Jäger), Swissuniversities under grants EyeStore and EyeStore+ (PI: Lena Jäger), the Estonian Research Council under grant PSG743 (PI: Kaidi Lõo), the German Federal Ministry for Education and Research under grant 01|S20043 (AEye, PI: Lena Jäger), the Croatian Science Foundation under grant IPCH-2022-04-3316 (MeRID, PI: Marijan Palmovic), the Foundation for Research in Science and the Humanities at the University of Zurich under grant STWF-22-020 (PI: Lena Jäger), the Carlsberg Foundation under grants CF23-1627 (PI: Hanne B. Søndergaard Knudsen, Danish data collection by Reading Cognition Lab, Aalborg University) and CF24-2005 (West Greenlandic data collection led by Johanne Sofie Krog Nedergård, University of Copenhagen with support (eye-tracker) from Reading Cognition Lab, Aalborg University, financed by The Obel Family Foundation grant to Aalborg University), the Foundation for Science and Technology, Portugal under grant UID/00214 (Center of Linguistics of the University of Lisbon), the Digital Society Initiative at the University of Zurich through a PhD scholarship (Anna Bondar), the German Federal Ministry of Education and Research through the DAAD programme Konrad Zuse Schools of Excellence in Artificial Intelligence (Iza Škrjanec).

References

Yevgeni Berzak, Boris Katz, and Roger Levy. 2018. Assessing Language Proficiency from Eye Movements in Reading. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, Marilyn Walker, Heng Ji, and Amanda Stent (Eds.). Association for Computational Linguistics, New Orleans, Louisiana, 1986–1996. https://doi.org/10.18653/v1/N18-1180

Yevgeni Berzak, Jonathan Malmaud, and Roger Levy. 2020. STARC: Structured annotations for reading comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 5726–5735.

Lena S. Bolliger, David R. Reich, Patrick Haller, Deborah N. Jakobi, Paul Prasse, and Lena A. Jäger. 2023. ScanDL: A Diffusion Model for Generating Synthetic Scanpaths on Texts. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 15513–15538. https://aclanthology.org/2023.emnlp-main.960

Kathy Conklin and Ana Pellicer-Sánchez. 2016. Using eye-tracking in applied linguistics and second language research. *Second Language Research* 32, 3 (2016), 453–467.

Ian Cunnings and Claudia Felser. 2013. The role of working memory in the processing of reflexives. *Language and Cognitive Processes* 28, 1-2 (2013), 188–219. https://doi.org/10.1080/01690965.2010.548391

Martha Bridge Denckla and Rita Rudel. 1974. Rapid automatized naming of pictured objects, colors, letters and numbers by normal children. *Cortex* 10, 2 (1974), 186–202.

Shuwen Deng, Paul Prasse, David Reich, Tobias Scheffer, and Lena Jäger. 2024. Fine-Tuning Pre-Trained Language Models with Gaze Supervision. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2:*

Short Papers), Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, Bangkok, Thailand, 217–224. https://doi.org/10.18653/v1/2024.acl-short.21

Shuwen Deng, David R Reich, Paul Prasse, Patrick Haller, Tobias Scheffer, and Lena A Jäger. 2023. Eyettention: An attention-based dual-sequence model for predicting human scanpaths during reading. *Proceedings of the ACM on Human-Computer Interaction* 7, ETRA (2023), 1–24.

Barbara A Eriksen and Charles W Eriksen. 1974. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics* 16, 1 (1974), 143–149.

Evelina Fedorenko. 2014. The role of domain-general cognitive control in language comprehension. *Frontiers in Psychology* 5 (2014), 335. https://doi.org/10.3389/fpsyg.2014.00335

Patrick Haller, Lena Bolliger, and Lena Jäger. 2024. Language models emulate certain cognitive profiles: An investigation of how predictability measures interact with individual differences. In *Findings of the Association for Computational Linguistics ACL 2024*, Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, Bangkok, Thailand and virtual meeting, 7878–7892. https://aclanthology.org/2024.findings-acl.469

Jarodzka Halszka, Kenneth Holmqvist, and Hans Gruber. 2017. Eye tracking in educational science: Theoretical frameworks and research agendas. *Journal of eye movement research* 10, 1 (2017), 10–16910.

Roy S. Hessels and Ignace T. C. Hooge. 2019. Eye tracking in developmental cognitive neuroscience—the good, the bad and the ugly. *Developmental Cognitive Neuroscience* 40 (2019), 100710.

Deborah N Jakobi, Daniel G Krakowczyk, and Lena A Jäger. 2024. Reporting eye-tracking data quality: Towards a new standard. In *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications*. 1–3.

Walter Kintsch. 1998. *Comprehension. A paradigm for cognition.* Cambridge University Press.

Daniel G Krakowczyk, David R Reich, Jakob Chwastek, Deborah N. Jakobi, Paul Prasse, Assunta Süss, Oleksii Turuta, Paweł Kasprowski, and Lena A. Jäger. 2023. Pymovements: A python package for eye movement data processing. In *Proceedings of the 2023 ACM Symposium on Eye Tracking Research and Applications*. 1–2.

Victor Kuperman, Kazunaga Matsuki, and Julie A. Van Dyke. 2018. Contributions of reader- and text-level characteristics to eye-movement patterns during passage reading. *Journal of Experimental Psychology: Learning Memory and Cognition* 44 (2018), 1687–1713. Issue 11. https://doi.org/10.1037/xlm0000547

Kristin Lemhöfer and Mirjam Broersma. 2012. Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods* 44 (2012), 325–343.

Stephan Lewandowsky, Klaus Oberauer, Lee-Xieng Yang, and Ullrich KH. Ecker. 2010. A working memory test battery for MATLAB. *Behavior Research Methods* 42, 2 (2010), 571–585.

Marie-Luise Müller, Deborah Jakobi, Maja Stegenwallner-Schuetz, Ramunė Kasperė, Nora Hollenstein, and Lena Jäger. 2024. *Data Management Plan (DMP) for COST Action "MultiplEYE" (CA21131).* https://doi.org/10.23668/psycharchives.15018

Bruno Nicenboim, Pavel Logacev, Carolina Gattei, and Shravan Vasishth. 2016. When high-capacity readers slow down and low-capacity readers speed up: Working memory and locality effects. *Frontiers in Psychology* 7 (2016), 280. https://doi.org/10.3389/fpsyg.2016.00280

Jared M Novick, Erika Hussey, Susan Teubner-Rhodes, J Isaiah Harbison, and Michael F Bunting. 2014. Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience* 29 (2014), 186–217. Issue 2. https://doi.org/10.1080/01690965.2012.758297

Paul Pimsleur. 1966. *Pimsleur language aptitude battery (Form S).* Harcourt, Brace and World.

Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124, 3 (1998), 372.

David Robert Reich, Paul Prasse, Chiara Tschirner, Patrick Haller, Frank Goldhammer, and Lena A Jäger. 2022. Inferring native and non-native human reading comprehension and subjective text difficulty from scanpaths in reading. In *2022 Symposium on eye tracking research and applications*. 1–8.

Noam Siegelman, Sascha Schroeder, Cengiz Acartürk, Hee-Don Ahn, Svetlana Alexeeva, Simona Amenta, Raymond Bertram, Rolando Bonandrini, Marc Brysbaert, Daria Chernova, et al. 2022. Expanding horizons of cross-linguistic research on reading: The Multilingual Eye-movement Corpus (MECO). *Behavior Research Methods* 54, 6 (2022), 2843–2863.

Ekta Sood, Simon Tannert, Philipp Mueller, and Andreas Bulling. 2020. Improving Natural Language Processing Tasks with Human Gaze-Guided Neural Attention. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 6327–6341. https://proceedings.neurips.cc/paper_files/paper/2020/file/460191c72f67e90150a093b4585e7eb4-Paper.pdf

J Ridley Stroop. 1935. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 18, 6 (1935), 643.

Pol van Rijn, Yue Sun, Harin Lee, Raja Marjieh, Ilia Sucholutsky, Francesca Lanzarini, Elisabeth André, and Nori Jacoby. 2023. Around the world in 60 words: A generative vocabulary test for online research. *arXiv preprint arXiv:2302.01614* (2023).

Titus von der Malsburg and Shravan Vasishth. 2011. What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language* 65, 2 (2011), 109–127. https://doi.org/10.1016/j.jml.2011.02.004