

Eye Tracking and NLP: A Survey and an Outlook

Yevgeni Berzak[✉], David R. Reich^{🇨🇭}, Omer Shubi[✉] and Lena A. Jäger^{🇨🇭}

[✉]Technion – Israel Institute of Technology ^{🇨🇭}University of Zurich

{berzak@technion, shubi@campus.technion}.ac.il

{davidrobert.reich, lenaann.jaeger}@uzh.ch

Abstract

With the rise of LLMs, NLP has been re-defining itself as a field of study, with an increasing interest shift towards areas such as NLP for science, model analysis, human-centered AI and user-facing applications. As part of this transformation, in recent years, a new area of NLP which combines text processing with eye movements in reading has been emerging. Recent work has demonstrated that this area has a strong potential for driving innovation in NLP and enhancing its scientific and societal impact. In this survey, we characterize the main streams of work on the intersection of eye tracking for reading and NLP, provide critical commentary on promises and challenges, and identify key open directions that can guide future work in this area.

1 Introduction

The unprecedented success of LLMs has led NLP into a new era in which it is redefining itself as a discipline. As part of this transformation, much focus has been shifting away from developing models for narrow textual tasks and towards research areas spanning the characterization of model mechanisms and capabilities, human-computer interaction, and the societal implications of deploying these models at scale. It is now also more feasible than ever to leverage NLP for scientific inquiry. As LLMs are fundamentally models of language, there has specifically been a strong interest in their potential role as modeling tools in the cognitive science of language. At the same time, the flexibility and unprecedented capabilities of current models make them increasingly suitable for user-facing applications. Indeed, many millions of people now use LLMs on a daily basis.

In this survey, we present a nascent area of research in NLP, namely the integration of text-based NLP with behavioral traces of eye move-

ments in reading. This area is part of a larger emerging body of work on integrating NLP with cognitive and neural signals, and is intimately connected with many of the transformational processes that NLP currently undergoes. In particular, it provides a new and thus far underexplored frontier for modeling innovation on a challenging and highly interesting type of multimodal data of eye movements and text. It is also a key domain in which NLP can be used as a scientific tool for cognitive modeling of human language processing and reading. Perhaps most importantly, the combination of eye tracking and NLP provides fertile ground for re-imagining the range and nature of possible NLP applications, with an emphasis on high societal impact domains such as education and information accessibility.

Our survey provides a structured entry point to the area of eye tracking and NLP, which has been gaining traction in recent years through a growing body of research, dedicated venues (Acartürk et al., 2025, 2026b), shared tasks (Hollenstein et al., 2021, 2022b), evaluation benchmarks (Hollenstein et al., 2023; Shubi et al., 2025b) and tutorials (Beinborn et al., 2021; Reich et al., 2025c). The interest in this area is also reflected in several prior surveys on eye movements for augmenting NLP (Mishra and Bhattacharyya, 2018; Hollenstein et al., 2019; Barrett and Hollenstein, 2020; Mathias et al., 2020; López-Cardona et al., 2025a; Acartürk et al., 2026a). While these overviews are highly valuable, the current survey takes a broader perspective by presenting not only uses of eye tracking for augmenting NLP models on textual tasks, but also uses of NLP for cognitive and eye movement modeling. Importantly, we also outline and emphasize emerging human-centered applications that involve real-time predictions about the reader and their interactions with the text.

The current survey is based in part on an ACL 2025 tutorial on eye tracking and NLP (Reich

et al., 2025c). It distills the main lines of work on eye tracking for reading and NLP, and takes stock of the work that has been carried out in this area thus far. We further identify key open challenges and discuss possible directions for addressing them. Importantly, we propose promising open avenues, which we envision to be instrumental for guiding future work in this domain.

While the survey primarily addresses the NLP and machine learning communities, it is intended to be accessible to a wide readership from different disciplines, including linguistics, psychology, cognitive science, and human computer interaction, and is introductory in nature. At the same time, we aim for our synthesis to be informative also to experts already working on the intersection of eye tracking and NLP, and to direct and inspire new research in this and adjacent areas.

The survey is structured as follows. In Section 2 we discuss eye movements in reading and their representation. Section 3 presents the current state of publicly available datasets with eye movements in reading. Following these preliminaries, we present three primary lines of work on eye tracking and NLP. The first is using eye tracking data for improving NLP (Section 4). The second, complementary direction, leverages NLP for scientific research in psycholinguistics, including modeling of eye movements in reading (Section 5). The third, and most forward-looking direction, is human-centered applications which support reading via predictions about readers and reading interactions (Section 6). In Section 7, we conclude by identifying promising avenues for addressing existing research bottlenecks and carrying out high-impact work in this area.

2 Eye Movements in Reading

For decades, eye tracking during reading has been a leading methodology for studying the cognitive processes underlying reading and language processing (see Rayner, 1998; Hyönä and Kaakinen, 2019; Rayner et al., 2006a; Schotter and Dillon, 2025, for overviews). When we read, our eyes move over the text in a sequence of fixations, stops during which the gaze remains relatively still and visual information is obtained, and saccades, which are rapid transitions between fixations. According to the eye-mind assumption (Just and Carpenter, 1980), the word currently fixated is also being currently processed, with little or no tempo-

ral lag between fixation and cognitive processing. This assumption, which was empirically validated to a large extent, posits that eye movements are tightly coupled with the underlying cognitive processes during reading and provides a central foundation for eye-tracking research on reading.

Consistent with the eye-mind assumption, the characteristics of the eye movement trajectory are strongly conditioned on the linguistic characteristics of the text. For example, longer, less frequent, and less predictable words are fixated for longer periods of time, and skipped more (e.g., Kliegl et al., 2004; Rayner et al., 2004; Staub et al., 2010; Staub, 2011; Smith and Levy, 2013). Eye movements during reading are also strongly conditioned on the reader. They vary substantially across readers, reflecting individual differences in factors such as age, working memory capacity, and language proficiency (e.g., Rayner et al., 2006b; Kuperman and Van Dyke, 2011; Staub, 2021; Berzak and Levy, 2023). Finally, eye movements also reflect the nature of the interaction of the reader with the text (e.g., Just et al., 1982b; Kaakinen et al., 2002; Hahn and Keller, 2023; Shubi and Berzak, 2023) and their moment-by-moment cognitive state during reading (e.g., Just and Carpenter, 1980; Reichle et al., 2010). Despite the rich information contained in eye movements about the text, the reader, and the reading interaction, they also contain a substantial amount of unexplained variance (i.e., noise), which introduces practical challenges for predictive modeling and the integration of this signal with NLP.

2.1 Data Collection

Recordings of eye movements in reading are typically obtained from participants in lab settings. In most cases, eye movement data is collected with specialized eye trackers that record the position of the eye as horizontal and vertical on-screen coordinates, at a high frame rate (250–2000 Hz) and with high positional accuracy. Such equipment is needed for reading studies in which the visual interest areas (typically words) are relatively small (but see Section 7.1 for more scalable alternatives). In addition to the eye tracking experiment itself, studies often include additional tasks that do not involve eye tracking, such as demographic and language history questionnaires, as well as language, reading and psychometric tests.

2.2 Eye Movements Representation

The eye movement signal can be used in its raw form, as samples from a continuous signal consisting of positional coordinates over time, typically sampled at intervals of 0.5 to several milliseconds. More commonly, however, it is preprocessed into a *scanpath* - an alternating sequence of fixations and saccades. This preprocessing step is done using algorithms for segmentation of raw gaze positions into oculomotor events (Salvucci and Goldberg, 2000; Engbert, 2006; Andersson et al., 2017). Note that the choice of the segmentation algorithm may influence performance on downstream tasks (Reich et al., 2025b).

Scanpaths are most commonly represented at the fixation level, where each fixation and its preceding or following saccade are characterized by features such as fixation location, fixation duration, and saccade direction. Individual fixations and saccades can be further aggregated into word-level measures, such as the word’s total fixation duration (the sum of all fixation durations on a given word) or whether a regression (backward saccade) was initiated from that word. Unlike fixation-level representations, which directly capture the chronological order of the scanpath, word-level representations are centered around the order of the words.

Both fixation-level and word-level representations can be further aggregated into summary measures. For example, the words’ total fixation duration can be averaged across reading trials (the participant’s scanpath over, e.g., a text screen), texts, participants, or both texts and participants. The different approaches to eye movement data aggregation reflect a tradeoff between expressivity on the one hand and robustness to noise on the other. They further interact with different approaches for aligning eye movements with the underlying text, which we discuss below.

2.3 Alignment of Eye Movements and Text

Using eye tracking data in the context of NLP requires addressing a fundamental challenge: aligning the eye movement trajectory to its underlying textual stimulus. This is a unique and difficult challenge for multimodal modeling, as the text is *static* and linearly ordered, while eye movements are *dynamic*: they unfold over time in a non-linear order over the stimulus text. At the same time, there are systematic and linguistically driven rela-

tions between the eye movement trajectory and the characteristics of the text.

The alignment strategy for eye movements and text depends on the granularity level of the eye movement representation. It may further be constrained by the model architecture and the task to be addressed. We also note that the same alignment strategy can be implemented differently in different model architectures, and even in the same type of architecture. The models of eye movements and text that were developed in the work surveyed in Sections 4 to 6 provide different solutions to the text-eye movements alignment problem, which remains a cardinal open challenge.

3 Datasets of Eye Movements in Reading

A crucial enabling factor for work on eye tracking and NLP is the availability of high-quality broad coverage datasets with eye movements during naturalistic reading. The first such major dataset, the Dundee Corpus (Kennedy et al., 2003), a 20-adult-participant corpus of L1 reading in English and French, was introduced in 2003. For over a decade since its introduction, Dundee was the primary eye tracking in naturalistic reading dataset available to the research community, which limited the possible use cases of eye movement data in NLP, and subsequently the volume and diversity of NLP work that used eye tracking data.

In recent years, however, the data landscape has been changing dramatically, as more and more publicly available eye tracking datasets for naturalistic reading are introduced. Jakobi et al. (2026) provide a living survey of existing resources, which estimates that there are currently 56 publicly available eye tracking datasets with naturalistic reading, with a total of 8,680 participants.¹ Currently available eye tracking corpora span over 30 languages, including multilingual datasets over parallel corpora (Siegelman et al., 2022, 2025), and reader populations such as L2 learners (Cop et al., 2017; Li and Clariana, 2019; Berzak et al., 2022; Kuperman et al., 2023, 2025), individuals with dyslexia (Nilsson Benfatto et al., 2016; Jakovljević et al., 2021; Dostálová et al., 2024; Hollenstein et al., 2022a), and participants with different levels of domain expertise (Jakobi et al., 2025a). They also include reading regimes beyond reading for comprehension,

¹Counts obtained from the living online table at <https://t.uzh.ch/1Yh>, accessed July 29, 2026.

such as information seeking and repeated reading (Berzak et al., 2025), acceptability judgments (Bondar et al., 2025b), and linguistic annotations (Hollenstein et al., 2018, 2020). Additional large-scale data collection projects, such as the multilingual corpus MultiPEYE (Jakobi et al., 2025c; Jäger et al., 2026), are currently underway.

The recent proliferation of eye tracking for reading datasets creates the necessary basis for meaningful integration of eye movement signals into modern-day NLP and machine learning models, whose effectiveness is known to crucially depend on the size of the training data, as well as on its diversity. Despite this very positive trend, due to the reliance on in-lab data collection using costly high-precision eye tracking equipment, the amount of available data relative to many other domains in which machine learning and NLP are applied remains relatively limited. This data collection bottleneck is especially severe for populations such as children and the elderly, as well as for languages that are spoken in countries where there is no appropriate infrastructure for eye tracking data collection. In Section 7.1 we discuss a number of avenues for scaling eye tracking data collection, which is critical for the future success of the field.

4 Eye Tracking for Improving Text-Based NLP

One of the most longstanding and common combinations of eye tracking and NLP to date has been the integration of eye movement signals from reading into NLP models, with the aim of improving performance on standard textual NLP tasks. Such tasks include part-of-speech tagging (Barrett and Søgaard, 2015a; Barrett et al., 2016), dependency parsing (Barrett and Søgaard, 2015b; Strzyz et al., 2019), sentiment analysis (Mishra et al., 2016; Barrett et al., 2018; Ren and Xiong, 2021; McGuire and Tomuro, 2021; Yang and Hollenstein, 2023; Kiegeland et al., 2024), named entity recognition (Hollenstein and Zhang, 2019; Ren and Xiong, 2021; Yu et al., 2024), relation extraction (Ren and Xiong, 2021), readability assessment (González-Garduño and Søgaard, 2017), question answering (Malmaud et al., 2020; Zhang and Hollenstein, 2024), hallucination detection (Maharaj et al., 2023), grammaticality judgments (Bondar et al., 2025a), sentence compression (Klerke et al., 2016; Sood et al., 2020b), paraphrasing (Sood et al., 2020b),

and general-purpose language modeling and open-ended generation (Kiegeland et al., 2024; Bondar et al., 2025a; López-Cardona et al., 2025c; Galiyamov et al., 2025; Säuberli et al., 2026; Chang et al., 2026). Several studies addressed multiple tasks using standard NLP evaluation benchmark datasets (Deng et al., 2023a, 2024).

Early approaches to the integration of eye movements in NLP models relied primarily on the concatenation of eye movement and textual features at the text- (e.g., Mishra et al., 2016), sentence- (e.g., Barrett and Søgaard, 2015b) or word-level (e.g., Barrett et al., 2016), using traditional machine learning models such as logistic regression, SVMs and HMMs. Newer modeling approaches marked two departures relative to this work. The first is relying on modern neural language model backbones. The second is an increased focus on word-level and fixation-level alignment approaches. Many of the recent modeling approaches are based on standard language model architectures such as RNNs and transformers, often already pretrained on text. These models were then modified to integrate word-level eye movement information, for example, by adding gaze features in the input (Hollenstein and Zhang, 2019; López-Cardona et al., 2025c), modifying model internals (e.g., attention, Sood et al., 2020b), or predicting gaze (see below). A notable example of a fixation-level approach applied to NLP tasks is the reordering of word embeddings according to the fixation sequence (Yang and Hollenstein, 2023).

A major challenge for the integration of eye movements in textual NLP tasks is the need for predictions on arbitrary textual data, where the text can be either human or model-generated. Since human gaze data are generally unavailable at inference time, several strategies have been proposed to address this challenge. These include using gaze as a *prediction target* in an auxiliary training objective (e.g., Klerke et al., 2016; Barrett et al., 2018; Malmaud et al., 2020) or preference optimization (Kiegeland et al., 2024; López-Cardona et al., 2025c; Bondar et al., 2025a; Galiyamov et al., 2025; Chang et al., 2026) that is discarded after training, using gaze as an *auxiliary input* during training while requiring only text at inference time (Deng et al., 2024), or replacing human gaze with synthetic gaze generated by computational models (Sood et al., 2020b; Khu-

rana et al., 2023; Deng et al., 2023a, 2024) during training, inference or both.

Broadly, this line of work reflects a shared hypothesis, according to which injecting data on how humans read texts may provide the right inductive bias for improving text processing in NLP models. While this is a plausible hypothesis, work in this area rarely specifies which precise aspects of the eye movement signal are relevant for the task at hand. This question is especially pertinent to many papers where eye movement data is from participants reading for comprehension rather than performing the task provided to the model (but see e.g., Mishra et al., 2016, Malmaud et al., 2020 and Bondar et al., 2025a for exceptions).

Despite the variety of tasks and architectures in this line of work, the performance gains from adding eye movement data to NLP models have typically been modest. Further, work in this space has been scarce recently. This, we believe, is no accident, as current LLMs trained at scale are already highly capable on a large variety of tasks, and it is challenging to improve the performance of such models via behavioral data augmentation, especially when such data is not task-specific, is limited in size, and is noisy, as is the case with eye movement data. While low-resource scenarios are in principle promising for the integration of eye tracking signals (Deng et al., 2023a), practically, limited textual resources are also often accompanied by limited eye tracking resources.

Overall, we conclude that the goal of improving state-of-the-art model performance on standard textual tasks via eye movement signals may currently be less feasible compared to other combinations of eye tracking and NLP described below. This, however, may change in the future as more eye tracking data becomes available, including synthetic eye movements which can be generated instantaneously at scale (Section 7.1), and as more effective models of eye movements and text are introduced (Section 7.2).

5 NLP for Cognitive and Eye Movement Modeling

A complementary research direction to integrating eye movement signals into NLP is *using* NLP for modeling reading and human language processing. While human language processing during reading can be studied using a range of behavioral methodologies (Section 7.1.2), we focus on

eye tracking, the dominant behavioral methodology in reading research (e.g., Rayner, 1998).

5.1 Broad Coverage Annotations for Psycholinguistic Modeling

NLP models can be instrumental for testing psycholinguistic theories of human language processing. Traditionally, testing such theories relied on manual linguistic annotations or responses to linguistic stimuli elicited from humans (e.g., next word predictions). Modern NLP enables automation and broad coverage extraction of cognitively relevant quantities from text. These quantities are then compared to reading times as measures of cognitive effort. A key example is using language models for extracting word surprisal as a proxy for contextual word predictability. Language model based surprisal has often been used in conjunction with eye movements in reading data to test predictions of surprisal theory (Hale, 2001; Levy, 2008) in English (Smith and Levy, 2013; Shain et al., 2024) and other languages (Wilcox et al., 2023). Ryu and Lewis (2021) used language model surprisal and eye movements data for testing surprisal and interference-based explanations of syntactic agreement phenomena. Berzak and Levy (2023) used surprisal to study lexical and contextual processing in L1 and L2 speakers, and Haller et al. (2026) to assess the measurement reliability of individual-level surprisal effects.

Another type of automated annotations obtainable with NLP models is syntactic trees. Demberg and Keller (2008) leveraged such annotations along with eye movement data to test the predictions of surprisal theory and dependency locality theory (Gibson, 1998, 2000). Similarly, Haller et al. (2026) used dependency trees to investigate processing difficulty associated with syntactic integration. In addition to surprisal and syntactic information, psycholinguistic studies have also used contextual word vectors from language models. For example, Smith and Vasishth (2020) used word embeddings as cognitive representations of lexical items in memory. Similarly, Meghdadi et al. (2026) replaced hand-crafted feature representations with word embeddings in the ACT-R model (Anderson and Lebiere, 1998; Anderson et al., 2004; Lewis and Vasishth, 2005).

We note that scalable LLM-based annotation can come at the cost of accuracy, interpretability, and theoretical grounding. Methodological chal-

lenges include subword-to-word surprisal aggregation, lower annotation accuracy, particularly beyond English, and the limited interpretability of neural representations. Moreover, LLM-based estimates may obscure theoretically relevant distinctions, such as that between syntactic and lexicalized surprisal.

5.2 Evaluating Human Alignment of LLMs using Eye Movements

A closely related line of work to the usage of NLP models for extracting cognitively relevant quantities discussed above has been testing whether language models are good models of human language processing, in many cases via the same quantities (e.g., surprisal). Eye tracking data has been playing an important role in addressing this question by providing a fine-grained behavioral benchmark for language models.

A popular line of work around this question has been testing the predictive power of language model surprisal for reading data obtained with eye tracking across different models (Goodkind and Bicknell, 2018; Oh and Schuler, 2023; Shain et al., 2024; Gruteke Klein et al., 2024), model layers (Kuribayashi et al., 2025) and decoding strategies (Bolliger and Jäger, 2025; Bolliger et al., 2024). Language model predictive power for eye movement measures was also examined for different reader groups (Haller et al., 2024; Škrjanec and Demberg, 2026), and for texts generated by LLMs (Bolliger et al., 2024; Bolliger and Jäger, 2025).

More broadly, a large body of work in psycholinguistics has been examining the linguistic capabilities of language models, with a special focus on models trained on data comparable in size to the linguistic exposure of humans (e.g., Linzen et al., 2016; Marvin and Linzen, 2018; Gulordava et al., 2018; Futrell et al., 2019; Hu et al., 2020; Warstadt et al., 2020). Such work has been instrumental for progress on long-standing debates about syntactic learnability hypotheses in linguistics and cognitive science (Chomsky, 1965; Lewis and Elman, 2001; Wilcox et al., 2024b). While most commonly, these studies rely only on textual stimuli (both naturalistic and controlled), some have also used behavioral evaluations using SPR and Maze (van Schijndel and Linzen 2018; Wilcox et al. 2021, see Section 7.1.2 on these methods).

A related line of work seeks not only to measure language model alignment with behavioral traces

of human language processing, but also to improve it by developing more cognitively plausible model architectures (Beinborn and Hollenstein, 2023; Oh and Linzen, 2026). Within this framework, several studies introduced a recency bias (de Varda and Marelli, 2024; Clark et al., 2025) and memory constraints (McCurdy and Hahn, 2024; Li et al., 2026; Thamma and Heilbron, 2026; Xu et al., 2026), including scanpath-based constraints (Mooney and Ho, 2026), for transformer language model architectures, evaluated against eye tracking measures.

At the moment, eye tracking does not provide a special advantage over methods such as SPR and Maze for much of this work, and the work presented in Section 5.1. This is because studies tend to use a single reading measure, which can be obtained more easily and at larger scale with these methods. To better leverage eye tracking for such work, future studies can focus on more fine-grained analyses of multiple reading measures, as well as other scanpath representations, which can better reveal processing dynamics over time. Across all methodologies, more work is called for with behavioral data collected from additional languages, populations and different profiles of linguistic knowledge (e.g., L1 vs L2).

5.3 NLP for Generative Models of Eye Movements in Reading

A longstanding challenge in the psychology of reading is the development of computational models of eye movements in reading. Such models generate one or more trajectories of eye movements for a given text. Beyond scientific inquiry, these models can also be highly useful for many applications, especially those in which relevant data from human participants is limited or infeasible to collect. Computational models of eye movements in reading can be broadly divided into two groups, theory driven *cognitive* models and theory agnostic *machine learning* models.

Cognitive models of eye movements implement different cognitive theories of reading and language processing. They are developed from first principles, based on known characteristics of oculomotor control in reading and the relation between eye movements and linguistic characteristics of the text (e.g., more frequent words are skipped more often). They use a relatively small number of parameters and are highly inter-

pretable. Prominent examples include E-Z Reader (Reichle et al. 1999, 2003, 2009, serial attention), SWIFT (Engbert et al. 2005; Seelig et al. 2020, parallel attention), SEAM (Rabe et al. 2024, activation coupling) and OB1-Reader (Snell et al. 2018, Bayesian inference). The linguistic representations used by these models are hand-crafted features rather than representations derived from embedding based language models. In principle, however, cognitive models could instead employ neural word embeddings to represent lexical knowledge, as demonstrated by recent work (Smith and Vasishth, 2020; Meghdadi et al., 2026) with the cognitive architecture ACT-R (Anderson et al., 2004; Lewis and Vasishth, 2005).

An alternative approach for modeling eye movements in reading, which has been gaining momentum in recent years, is data-driven, theory-agnostic modeling based on supervised machine learning from human eye movement data. In this approach, models are trained to generate synthetic eye movements or derivatives thereof, such as word reading times or fixation locations, based on existing training data from human participants. Early methods in this line of work predicted the next fixation location using logistic regression, representing the prior scanpath using word-level features of the text and of eye movements (Nilsson and Nivre, 2009, 2011). Newer models build on language models, using their contextual word representations as the textual component of the architecture. Among these, Eyettention (Deng et al., 2023b, 2026) predicts fixation duration and fixation location, derived from saccade distance, by aligning text and eye movements via cross-attention between a transformer text encoder and an LSTM fixation encoder. ScanDL (Bolliger et al., 2023, 2025) predicts the sequence of fixations using a diffusion model, where eye movements are aligned with the text using shared word index embeddings.

An additional body of work generates fixation sequences (Re et al., 2025), completions of fixation sequences (Sood et al., 2025) or raw gaze samples (Fuhl and Kasneci, 2018; Fuhl et al., 2018, 2021; Fuhl and Kasneci, 2022; Prasse et al., 2023; Jiao et al., 2026) without relying on the textual stimulus. Beyond scanpath prediction, another line of work predicts word-level eye movement measures (see Hollenstein et al., 2021, 2022b, for overviews). A notable example is

NEAT (Hahn and Keller, 2023), which predicts the probability of fixating words using an LSTM-based model.

Neural models are less interpretable than cognitive models. At the same time, their strength lies in closer fits to human eye movements. Moreover, such models can be more easily adjusted to model specific languages, textual domains, reader populations and specific readers, enabling bespoke large-scale generation of synthetic eye movement data. We envision that this capability will be a key enabler for future work on eye movements and NLP (Sections 7.1.3, 7.2, 7.3 and 7.6).

6 Human-Centered Applications

An exceptionally promising research area on the intersection of eye tracking and NLP is human-centered applications that rely on real-time predictions about the characteristics of the reader, or their cognitive state while interacting with the text, from eye movements in reading. Such applications expand the range of uses for NLP, and substantially broaden its impact potential in domains such as education and information accessibility, which we further discuss in Section 7.3.

In recent years, several proof-of-concept studies on various prediction tasks from eye movements in reading have laid the foundations for such applications. Multiple studies have obtained promising preliminary results on inferring different *characteristics of the reader*. These include aspects of the linguistic knowledge of the reader, such as native language (Berzak et al., 2017; Ahn et al., 2020; Reich et al., 2022; Skerath et al., 2023) and language proficiency (Berzak et al., 2018; Demareva et al., 2022; Steichen et al., 2022; Shalileh et al., 2024; Frenkel et al., 2026), as well as extralinguistic knowledge, such as domain expertise (Jakobi et al., 2025b). Within this line of work, a growing body of research has been specifically focused on the detection of dyslexia (Rello and Ballesteros, 2015; Nilsson Benfatto et al., 2016; Jothi Prabha and Bhargavi, 2020; Raatikainen et al., 2021; Nerušil et al., 2021; Haller et al., 2022; Björnsdóttir et al., 2023; Shalileh et al., 2023; Laurinavichyute et al., 2025).

Prediction tasks have been further explored for *reader cognitive states* that reflect different interactions of the reader with the text. Such tasks include predicting different manners of reading: information seeking versus reading for compre-

hension (Shubi et al., 2025a), linguistic annotation versus reading for comprehension (Hollenstein et al., 2023), skimming versus proofreading (Biedert et al., 2012; Chen et al., 2023), first versus repeated reading (Meiri et al., 2025), detecting mind wandering during reading (Hutt et al., 2019; Mills et al., 2021), and textual decoding of information seeking goals (Hadar et al., 2026).

A longstanding problem in the area of cognitive state prediction tasks that remains extremely challenging is predicting reading comprehension from eye movements. This task is rooted in psycholinguistic literature which suggests that eye movements reflect online comprehension processes (e.g., Just and Carpenter, 1980; Reichle et al., 2010; Schotter et al., 2014; Rayner et al., 2016), and that eye movement features differ on average across readers and texts as a function of text comprehension level (Mézière et al., 2023, 2024). However, to date, predictive models of comprehension from eye movements have yielded small improvements over text-only baselines, especially at the granularity of a single participant reading a single text (Copeland and Gedeon, 2013; Makowski et al., 2019; Ahn et al., 2020; Reich et al., 2022; Shubi et al., 2024; Zhang et al., 2025). There are two key hindering factors for this line of work. First, there are to date only very few publicly available eyetracking-while-reading datasets that provide standardized or piloted reading comprehension questions that go beyond attention checks which typically show ceiling effects. The second hindering factor has been the limitation to adult L1 data, often from highly proficient readers, which is less central for this task compared to, e.g., children, L2 speakers, less skilled or populations with cognitive impairments. In Section 7.3 we stress the importance of collecting data from other populations, which are key for comprehension prediction tasks.

Similarly to the domains of text-based NLP tasks and predictive modeling of eye movements in reading, human-centric prediction tasks typically involve joint modeling of eye movements and text. As such, many of the models and alignment approaches developed for these tasks are similar to those presented in Sections 4 and 5.3, only that instead of predicting a property of the text or the eye movement trajectory, the models are trained to predict properties of the reader or their cognitive state. A notable difference from

the models in Sections 4 and 5.3 is that models for human-centered tasks typically do require eye movement data from human participants at test time. Recent models for human-centered tasks commonly rely on word- or fixation-based alignment of eye movements and text via modifications to standard pretrained language models (e.g., transformer encoders), or adaptations of language-vision models (Shubi et al., 2024). Models that were originally developed for next fixation prediction (Deng et al., 2023b) were also adapted for such tasks (Shubi et al., 2024). An alternative to the language modeling approach used vision models instead, representing the eye movements, the text, or both as images (Bhattacharya et al., 2020; Shubi et al., 2025a; Zhang et al., 2025).

EyeBench (Shubi et al., 2025b) is a recent benchmark that brings human-centered prediction tasks and the models developed for them under a common evaluation framework. Current EyeBench evaluations suggest that many of these tasks are very challenging, and none of the existing modeling approaches for eye movements and text generalize across prediction tasks.

7 Outlook

The work presented in this survey reflects the state of the art of a young field. In this concluding section, we outline directions for addressing existing challenges in research on eye tracking and NLP, and identify research directions that are likely to drive scientific and engineering innovation as well as societal impact in this area.

7.1 Scaling Eye Tracking Data Collection

Despite the growing volume and diversity of eye tracking for reading data in recent years (Section 3), nearly all data collection projects to date rely on in-lab experiments with costly eye tracking equipment. While high-quality data collected in controlled in-lab settings has been highly valuable for jump-starting fruitful interactions between the psychology of reading and NLP, long-term progress in these areas will require solving the current logistical and financial challenges of eye tracking data collection. Below, we outline three main directions for addressing this core issue.

7.1.1 Eye Tracking with Ordinary Cameras

So far, very few attempts were made to collect eye movements in reading data with standard cameras such as webcams, and built-in cameras on laptops,

tablets and mobile phones (Ribeiro et al., 2023; Hutt et al., 2024). The primary reason for this is the low sampling rate and, even more importantly, the low accuracy of open source software for eye movement data collection with standard cameras. For example, WebGazer (Papoutsaki et al., 2016), perhaps the most popular open source software for webcam eye tracking, was estimated to have an average error of 4.17° of visual angle, which makes reliable mapping of fixations to words infeasible.

However, recent studies suggest not only that fixations and saccades may be identified at low-frequency sampling rates (Angele et al., 2025), but also that commercial software for webcam eye tracking can reach a rather low calibration error (1.45° , Kaduk et al. 2024). Preliminary results suggest that webcams may suffice for relatively reliable mapping of fixations to words when the text is sufficiently large, and yield reading times and frequency effects similar to those obtained with high-end eye trackers (Serrano-Carot et al., 2025).

Leveraging these developments, there is a need for further validation studies of current webcam eye tracking software, and a potential for scaling reading data collection with such software in the near future, including naturalistic everyday reading environments using tablets and phones. In the longer term, we identify an open opportunity for the community: the development of high-accuracy open-source software solutions for webcam eye tracking and their validation in reading studies.

7.1.2 Proxy Behavioral Methods

Additional alternatives to in-lab eye tracking data collection are simpler and cheaper behavioral methods that are commonly used in psycholinguistics and can be deployed online at scale. One such method is Self-Paced Reading (SPR) (Just et al., 1982a; Aaronson and Ferres, 1984), where participants read the text by revealing one word at a time using key presses. In Maze (Forster et al., 2009; Boyce et al., 2020), which similarly uses key presses, participants sequentially choose between each word and an implausible distractor. In both methods, reading times are recorded between key presses. In a more recent method, Mouse Tracking (MoTR) (Wilcox et al., 2024a), reading is performed by unblurring a small area around the mouse cursor of an otherwise blurred text, mimicking foveal and parafoveal vision.

All three methods support the collection of word reading times. However, differently from

eye tracking, they are not based on standard natural reading, and involve noise from motor planning and execution of key presses or mouse movement. SPR and Maze are limited to a single reading measure (per-word response time), while MoTR provides a richer signal that more closely resembles eye tracking data in natural reading (Wilcox et al., 2024a). This is consistent with preliminary experiments demonstrating that unlike SPR, MoTR is beneficial for pre-training eye tracking inference models (Reich et al., 2025a). Crucially, all three methods can support large-scale online data collection without specialized equipment.

7.1.3 Synthetic Eye Movements

The third solution to the eye movement data collection bottleneck is the generation of synthetic eye movements. In Section 5.3 we presented neural models for generating eye movements in reading. These models are currently underutilized. Furthermore, we anticipate that in the coming years, these models will continue improving in precision and efficiency. The convergence of the increasing amount and diversity of high-quality eye tracking data from humans, with modeling innovation in this space, opens the door for models that will generate synthetic eye movements for extremely large amounts of text, and can be customized to different languages, text genres, specific populations and individual readers. To support the development of such models, new evaluation measures for scanpaths (von der Malsburg and Vasishth, 2011; Fahimi and Bruce, 2021) that generalize across different texts are needed.

Large scale generation of synthetic eye movements will enable pretraining multimodal text-eye movement foundation models (Section 7.2) that can be used for enhancing text processing, as well as developing models for various eye movement centered prediction tasks. Crucially, eye movement generations will enable *instantaneous* eye movement based evaluations of model generated text, which are not possible with human eye movements (Section 7.6). Furthermore, while many user-facing capabilities require human eye movements, some can be supported by the ability of such models to generate eye movements on demand, instantly, and in real time, for example by selecting or generating future texts for a given reader using predicted eye movements based on their past reading patterns (Section 7.3).

7.2 Models of Eye Movements and Text

Joint modeling of eye movements and text which supports inferences about readers is an underexplored multimodal modeling challenge for NLP and machine learning. Despite substantial recent progress in this area, many existing models are ad-hoc modifications of standard language model architectures and training objectives. Importantly, most developed models are *text-centric*: their backbones are models that are well suited for text and are often pretrained on large amounts of text, and then modified using small amounts of eye movement data. While such approaches can be effective, they are not necessarily optimal for effective processing of eye movement data, leaving much room for further development of first-principles models of eye movements and text in which eye movements are first-class citizens.

In particular, there is currently no “foundation” model of eye movements and text which is pretrained on eye tracking-while-reading data. The recent dramatic improvements in the eye tracking data landscape, as well as current and future possibilities for large-scale generation of eye movement training data discussed in Section 7.1.3, make the development of such models a feasible goal for the coming years. EyeBench (Shubi et al., 2025b) provides implementations of existing architectures for modeling eye movements and text, and a unified evaluation framework on a variety of prediction tasks, supporting future progress in this area.

An additional cardinal challenge for joint modeling of eye movements and text is accounting for the substantial variability in eye movement patterns across readers. Addressing this variability requires not only collecting more data, but also leveraging information that is readily available in existing datasets. In particular, the large majority of existing work on modeling eye movements and text does not use information collected in questionnaires and tests administered in addition to the eye tracking recording session, such as demographic information and linguistic background and proficiency. Such information can improve the performance of human-centered applications, as exemplified by Laurinavichyute et al. (2025) on the task of dyslexia prediction, and can further support customization of eye movement generation models to specific populations and readers discussed in Section 7.1.3.

7.3 Human-Centered Applications

The studies surveyed in Section 6 provide proofs-of-concept for capabilities that can support new types of human-centered applications. However, these studies only scratch the surface of the opportunities that arise in this space. As NLP is expanding towards user-facing applications, eye tracking data can provide the basis for a variety of new applications that support human users in areas such as information accessibility, education and digital learning, and interactive online systems.

In broad strokes, in the future, eye movement data could be collected and analyzed in real time as users read. This data could be used for assessing the reader, as well as for real-time inferences about the reading process and the interaction of the reader with the text (e.g., monitoring attention, text comprehension or reading goal). One can imagine future applications in classrooms where the linguistic and reading abilities of children, and their performance on specific reading tasks are assessed directly from eye movements, without the need for developing and deploying costly standard reading evaluations. Pre-screening for dyslexia can also be deployed at scale, potentially improving diagnosis and monitoring the development of reading abilities over time.

Future human-centered applications are not limited to assessments. Real time monitoring of reading can also enhance the reading experience and support reading interventions. For example, if an eye movement-based system detects that a reader is struggling to comprehend a text, another system component can select the next text to be easier, or simplify the current text in real time. For students and language learners, such systems can enhance language learning and reading development. For other populations, such as the elderly, special assistance can be triggered when the user is accessing and not comprehending critical information.

Crucially, these applications will require far more data and development focus on their main audiences, often not represented in convenience samples, including children, adults with reading difficulties, the elderly and L2 learners. Additional research will also be needed on fast real-time inference, which is required for such applications.

7.4 Structured Evaluations

Enabling the applications discussed above will require rigorous model evaluations. As in other ar-

eas of NLP, evaluations should include appropriate statistical tests. Eye tracking datasets pose additional challenges for statistical testing because their samples are often non-i.i.d., but contain multiple measurements from both participants and items: multiple participants read the same texts, and each participant reads multiple texts. Many evaluations require accounting for these dependencies using mixed-effects models (Barr et al., 2013). Furthermore, train–test splits should respect the crossed participant–item structure to enable fine-grained assessments of model generalization to unseen participants, unseen texts, and their combination.

These evaluation regimes directly correspond to real-world deployment scenarios. The *seen-text, new-participant* regime is relevant for assessment-like cases, where the set of textual materials is fixed. The *seen-participant, new-text* regime is suitable for systems that adapt to user behavior over time, such as digital learning platforms. The *new-participant, new-text* regime is the most challenging and general setting, with no prior eye movement data for the participant nor the text. Specifying the evaluation regimes is crucial for future work, both for understanding model generalization and for gauging relevance for applications. Finally, demonstrating relevance to real-time applications may also require evaluations on varying amounts of training and testing data, and different degrees of data quality.

7.5 Eye Tracking in Current Day NLP

In Section 4 we noted that improving NLP model performance using eye movements is highly challenging due to the drastic recent improvements in the performance of current LLMs. These improvements have accelerated a shift in academic NLP from model development to model analysis, and interpretability of model-internal mechanisms (Belinkov and Glass, 2019; Madsen et al., 2022; Mueller et al., 2026). Eye movement data could potentially play a role in this transition, and a number of studies have already examined the alignment between transformer attention mechanisms and human attention as operationalized with eye tracking measures (Sood et al., 2020a; Eberle et al., 2022; Wang et al., 2024). Further work in this area could widen the set of existing tools and benchmarks for analyzing language models.

Furthermore, as NLP models are improving,

there is an ever growing need to ensure their safe and fair deployment. Concerns about these issues are highly relevant to multimodal models of eye movements and text, especially for user-facing applications (Sections 6 and 7.3). Recent work has demonstrated that such models can indeed be biased. For example, models that predict language proficiency can favor speakers of certain native languages over others (Frenkel et al., 2026). Additional biases may be present for different reading strategies, educational backgrounds, ages, gender and others. Privacy and the possibility of identifying readers from their eye movements (Kasprowski and Ober, 2004; Jäger et al., 2020; Makowski et al., 2019), including associated biases (Prasse et al., 2022), are also becoming prominent concerns. We anticipate that addressing these questions will become front and center in the coming years.

Finally, many current LLMs are multimodal models that process, among others, both text and visual information. Eye movements during inspection of visual scenes, and specifically visual saliency (e.g., Fosco et al., 2020), are longstanding research areas in psychology, computer vision and HCI. In NLP, eye movements over images were used for image captioning (Takmaz et al., 2020), visual question answering (Sood et al., 2023) and prediction of human preferences for images (Papadopoulos et al., 2026). While the focus of the current survey is on eye movements for reading, we anticipate that work on eye movements over images and videos will continue gaining momentum in NLP, and potentially be integrated in the future with eye movements in reading.

7.6 Evaluation of NLP Model Outputs with Eye Movements

Many NLP evaluation methodologies rely on “gold standard” behavioral human responses, or LLMs which approximate them. This data can come in the form of annotations, human generations such as summaries and translations, user preference responses, and other forms of feedback from human users. In most cases, these are *offline* signals reflecting the end products of human linguistic processing and decision making processes.

Eye tracking offers an alternative, and thus far underexplored, approach to NLP system evaluation by capturing *online* information on cognitive processes during reading. Such information can

enable more fine-grained evaluation of NLP system outputs and provide direct evidence for real-time processing difficulty, which is more challenging to obtain with offline methods. Further, unlike other applications which require large amounts of eye tracking data, many evaluations are feasible even with limited eye movement data, as well as with proxy methods discussed in Section 7.1.

A few studies have already introduced eye movement-based NLP system evaluations in the domains of text generation (Zarrieß et al., 2015), readability (Gruteke Klein et al., 2025), simplification (Säuberli et al., 2026), machine translation (Doherty et al., 2010; Sajjad et al., 2016), and summarization (Ikhwantri et al., 2024), demonstrating the value of this approach. Much room remains for further work on eye movement-based measures as standalone evaluations or as complementary methods for standard offline evaluations. Beyond evaluation, as discussed in Section 4, the ability of eye movements to discriminate between higher- and lower-quality model outputs also makes them attractive as supervision signals for model optimization (Bondar et al., 2025b; López-Cardona et al., 2025b; He et al., 2026).

8 Conclusion

We presented the main streams of work on the intersection of eye tracking for reading and NLP. The current momentum in research that combines eye tracking with NLP reflects the potential of this area in providing opportunities for exciting work on modeling, evaluation, basic science and user-facing applications. These research directions will substantially broaden the scope of NLP, and enhance its scientific and societal value. The current survey provides a roadmap towards this vision.

References

Doris Aaronson and Steven Ferres. 1984. [The word-by-word reading paradigm: An experimental and theoretical approach 1](#). In David E. Kieras and Marcel A. Just, editors, *New Methods in Reading Comprehension Research*, pages 31–68. Routledge, New York, United States.

Cengiz Acartürk, Burcu Can, Melike Caglayan, Jamal A. Nasir, and Çağrı Çöltekin. 2026a. [A survey of incorporating gaze data into natural language processing models and applica-](#)

[tions](#). In *Proceedings of the Second International Workshop on Eye-Tracking Resources and Evaluation for Human-Aligned NLP*, pages 64–76.

Cengiz Acartürk, Burcu Can, Jamal A. Nasir, and Çağrı Çöltekin, editors. 2026b. *Proceedings for the Second International Workshop on Eye-Tracking Resources and Evaluation for Human-Aligned NLP*. ELDA.

Cengiz Acartürk, Jamal A. Nasir, Burcu Can, and Çağrı Çöltekin, editors. 2025. *Proceedings of the First International Workshop on Gaze Data and Natural Language Processing*. INCOMA Ltd.

Seoyoung Ahn, Conor Kelton, Aruna Balasubramanian, and Greg Zelinsky. 2020. [Towards predicting reading comprehension from gaze behavior](#). In *Proceedings of the 2020 Symposium on Eye Tracking Research and Applications*, pages 1–5.

John R. Anderson, Daniel Bothell, Michael D. Byrne, Scott Douglass, Christian J. Lebiere, and Yulin Qin. 2004. An integrated theory of the mind. *Psychological review*, 111(4):1036–1060.

John R. Anderson and Christian J. Lebiere. 1998. *The Atomic Components of Thought*. Psychology Press, New York, United States.

Richard Andersson, Linnea Larsson, Kenneth Holmqvist, Martin Stridh, and Marcus Nyström. 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.

Bernhard Angele, Zeynep Gunes Ozkan, Marina Serrano-Carot, and Jon Andoni Duñabeitia. 2025. [How low can you go? Tracking eye movements during reading at different sampling rates](#). *Behavior Research Methods*, 57(7):195.

Dale J. Barr, Roger P. Levy, Christoph Scheepers, and Harry J. Tily. 2013. [Random effects structure for confirmatory hypothesis testing: Keep it maximal](#). *Journal of Memory and Language*, 68(3):255–278.

Maria Barrett, Joachim Bingel, Nora Hollenstein, Marek Rei, and Anders Sjøgaard. 2018. [Sequence classification with human attention](#). In

- Proceedings of the 22nd Conference on Computational Natural Language Learning*, pages 302–312.
- Maria Barrett, Joachim Bingel, Frank Keller, and Anders Søgaard. 2016. [Weakly supervised part-of-speech tagging using eye-tracking data](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 579–584.
- Maria Barrett and Nora Hollenstein. 2020. [Sequence labelling and sequence classification with gaze: Novel uses of eye-tracking data for natural language processing](#). *Language and Linguistics Compass*, 14(11):1–16.
- Maria Barrett and Anders Søgaard. 2015a. [Reading behavior predicts syntactic categories](#). In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, pages 345–349.
- Maria Barrett and Anders Søgaard. 2015b. [Using reading behavior to predict grammatical functions](#). In *Proceedings of the Sixth Workshop on Cognitive Aspects of Computational Language Learning*, pages 1–5.
- Lisa Beinborn and Nora Hollenstein. 2023. *Cognitive Plausibility in Natural Language Processing*. Springer, Cham, Switzerland.
- Lisa Beinborn, Nora Hollenstein, and Willem Zuidema. 2021. [Analyzing the cognitive plausibility of deep language models](#). Advanced course at the 32nd European Summer School in Logic, Language and Information (ESSLLI 2021), online. July 26–30, 2021. Last accessed July 30, 2026.
- Yonatan Belinkov and James Glass. 2019. [Analysis methods in neural language processing: A survey](#). *Transactions of the Association for Computational Linguistics*, 7:49–72.
- Yevgeni Berzak, Boris Katz, and Roger P. 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)*, pages 1986–1996.
- Yevgeni Berzak and Roger P. Levy. 2023. [Eye movement traces of linguistic knowledge in native and non-native reading](#). *Open Mind*, 7:179–196.
- Yevgeni Berzak, Jonathan Malmaud, Omer Shubi, Yoav Meiri, Ella Lion, and Roger P. Levy. 2025. [OneStop: A 360-participant English eye-tracking dataset with different reading regimes](#). *Scientific Data*, 12(1):1995.
- Yevgeni Berzak, Chie Nakamura, Suzanne Flynn, and Boris Katz. 2017. [Predicting native language from gaze](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pages 541–551.
- Yevgeni Berzak, Chie Nakamura, Amelia Smith, Emily Weng, Boris Katz, Suzanne Flynn, and Roger P. Levy. 2022. [CELER: A 365-participant corpus of eye movements in L1 and L2 English reading](#). *Open Mind*, 6:41–50.
- Nilavra Bhattacharya, Somnath Rakshit, Jacek Gwizdka, and Paul Kogut. 2020. [Relevance prediction from eye-movements using semi-interpretable convolutional neural networks](#). In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, pages 223–233.
- Ralf Biedert, Jörn Hees, Andreas Dengel, and Georg Buscher. 2012. [A robust realtime reading-skimming classifier](#). In *Proceedings of the 2012 Symposium on Eye Tracking Research and Applications*, pages 123–130.
- Marina Björnsdóttir, Nora Hollenstein, and Maria Barrett. 2023. [Dyslexia prediction from natural reading of Danish texts](#). In *Proceedings of the 24th Nordic Conference on Computational Linguistics*, pages 60–70.
- Lena S. Bolliger, Patrick Haller, and Lena A. Jäger. 2024. [On the alignment of LM language generation and human language comprehension](#). In *Proceedings of the 7th BlackboxNLP Workshop: Analyzing and Interpreting Neural Networks for NLP*, pages 217–231.
- 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*, pages 15513–15538.
- Lena S. Bolliger, David R. Reich, and Lena A. Jäger. 2025. ScanDL 2.0: A generative model of eye movements in reading synthesizing scanpaths and fixation durations. *Proceedings of the ACM on Human-Computer Interaction*, 9(3):1–29.
- Lena Sophia Bolliger and Lena Ann Jäger. 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*, pages 7459–7476, Suzhou, China. Association for Computational Linguistics.
- Anna Bondar, David R. Reich, and Lena A. Jäger. 2025a. AIEYEgument: Leveraging eye-tracking-while-reading to align language models with human preferences. In *Proceedings of the First International Workshop on Gaze Data and Natural Language Processing*, pages 58–70.
- Anna Bondar, David R. Reich, and Lena A. Jäger. 2025b. CoLAGaze: A corpus of eye movements for linguistic acceptability. In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications*.
- Veronica Boyce, Richard Futrell, and Roger P. Levy. 2020. Maze made easy: Better and easier measurement of incremental processing difficulty. *Journal of Memory and Language*, 111:104082.
- Haw-Shiuan Chang, Jeffrey Gomez, Mehul Patwari, Aryan Sajith, and Hamed Zamani. 2026. Your mouse and eyes secretly leak your preference: LLM alignment using implicit feedback from users. arXiv preprint arXiv:2606.20482.
- Xiuge Chen, Namrata Srivastava, Rajiv Jain, Jennifer Healey, and Tilman Dingler. 2023. Characteristics of deep and skim reading on smartphones vs. desktop: A comparative study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*.
- Noam Chomsky. 1965. *Aspects of the Theory of Syntax*. MIT Press, Cambridge, Massachusetts, United States.
- Christian Clark, Byung-Doh Oh, and William Schuler. 2025. Linear recency bias during training improves transformers’ fit to reading times. In *Proceedings of the 31st International Conference on Computational Linguistics (COLING 2025)*, pages 7735–7747.
- Uschi Cop, Nicolas Dirix, Denis Drieghe, and Wouter Duyck. 2017. Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior Research Methods*, 49(2):602–615.
- Leana Copeland and Tom Gedeon. 2013. Measuring reading comprehension using eye movements. In *Proceedings of the 4th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 791–796.
- V. A. Demareva, A. V. Golubinskaya, Yu. A. Edelova, and R. V. Golubin. 2022. Evaluation of level of foreign language proficiency based on eye movement data. *Journal of Optical Technology*, 89(8):484–489.
- Vera Demberg and Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2):193–210.
- Shuwen Deng, Cui Ding, David R. Reich, Paul Prasse, and Lena A. Jäger. 2026. Eyettention II: A dual-sequence architecture for modeling fixation location, within-word landing position, and fixation duration in reading. *Behavior Research Methods*. In press.
- Shuwen Deng, Paul Prasse, David R. Reich, Tobias Scheffer, and Lena A. Jäger. 2023a. Pre-trained language models augmented with synthetic scanpaths for natural language understanding. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6500–6507.
- Shuwen Deng, Paul Prasse, David R. Reich, Tobias Scheffer, and Lena A. 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)*, pages 217–224.
- Shuwen Deng, David R. Reich, Paul Prasse, Patrick Haller, Tobias Scheffer, and Lena A. Jäger. 2023b. [Eyettention: An attention-based dual-sequence model for predicting human scanpaths during reading](#). *Proceedings of the ACM on Human-Computer Interaction*, 7(ETRA):1–24.
- Stephen Doherty, Sharon O’Brien, and Michael Carl. 2010. [Eye tracking as an MT evaluation technique](#). *Machine Translation*, 24(1):1–13.
- Nicol Dostálová, Roman Švaříček, Jan Sedmidubský, Wolf Culemann, Čeněk Šašinka, Pavel Zezula, and Jiří Čeněk. 2024. [ETDD70: Eye-tracking dyslexia dataset](#). Dataset, Zenodo.
- Oliver Eberle, Stephanie Brandl, Jonas Pilot, and Anders Sjøgaard. 2022. [Do transformer models show similar attention patterns to task-specific human gaze?](#) In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4295–4309.
- Ralf Engbert. 2006. [Microsaccades: A microcosm for research on oculomotor control, attention, and visual perception](#). In Susana Martinez-Conde, Stephen L. Macknik, Luis M. Martínez, Jose-Manuel Alonso, and Peter Ulric Tse, editors, *Visual Perception*, volume 154 of *Progress in Brain Research*, pages 177–192. Elsevier.
- Ralf Engbert, Antje Nuthmann, Eike M. Richter, and Reinhold Kliegl. 2005. [SWIFT: A dynamical model of saccade generation during reading](#). *Psychological Review*, 112(4):777.
- Ramin Fahimi and Neil D. B. Bruce. 2021. [On metrics for measuring scanpath similarity](#). *Behavior Research Methods*, 53(2):609–628.
- Kenneth I. Forster, Christine Guerrero, and Lisa Elliot. 2009. [The maze task: Measuring forced incremental sentence processing time](#). *Behavior Research Methods*, 41(1):163–171.
- Camilo Fosco, Anelise Newman, Pat Sukhum, Yun Bin Zhang, Nanxuan Zhao, Aude Oliva, and Zoya Bylinskii. 2020. [How much time do you have? Modeling multi-duration saliency](#). In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4473–4482.
- Shachar Frenkel, Ido Falah, Omer Shubi, and Yevgeni Berzak. 2026. [Language proficiency assessment from eye movements in naturalistic passage reading](#). arXiv preprint arXiv:2608.30583.
- Wolfgang Fuhl and Enkelejda Kasneci. 2018. [Eye movement velocity and gaze data generator for evaluation, robustness testing and assess of eye tracking software and visualization tools](#). arXiv preprint arXiv:1808.09296.
- Wolfgang Fuhl and Enkelejda Kasneci. 2022. [HPCGen: Hierarchical k-means clustering and level based principal components for scan path generation](#). In *Proceedings of the 2022 Symposium on Eye Tracking Research and Applications*, pages 1–7.
- Wolfgang Fuhl, Yao Rong, and Enkelejda Kasneci. 2021. [Fully convolutional neural networks for raw eye tracking data segmentation, generation, and reconstruction](#). In *Proceedings of the 25th International Conference on Pattern Recognition (ICPR)*, pages 142–149.
- Wolfgang Fuhl, Thiago Santini, Thomas Kuebler, Nora Castner, Wolfgang Rosenstiel, and Enkelejda Kasneci. 2018. [Eye movement simulation and detector creation to reduce laborious parameter adjustments](#). arXiv preprint arXiv:1804.00970.
- Richard Futrell, Ethan G. Wilcox, Takashi Morita, Peng Qian, Miguel Ballesteros, and Roger P. Levy. 2019. [Neural language models as psycholinguistic subjects: Representations of syntactic state](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 32–42.
- Karim Galliamov, Ivan Titov, and Ilya Pershin. 2025. [Enhancing RLHF with human gaze modeling](#). In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 30637–30643.

- Edward Gibson. 1998. [Linguistic complexity: Locality of syntactic dependencies](#). *Cognition*, 68(1):1–76.
- Edward Gibson. 2000. [The dependency locality theory: A distance-based theory of linguistic complexity](#). In Alec Marantz, Yasushi Miyashita, and Wayne O’Neil, editors, *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium*, pages 95–126. MIT Press, Cambridge, Massachusetts, United States.
- Ana Valeria González-Garduño and Anders Søgaard. 2017. [Using gaze to predict text readability](#). In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 438–443.
- Adam Goodkind and Klinton Bicknell. 2018. [Predictive power of word surprisal for reading times is a linear function of language model quality](#). In *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2018)*, pages 10–18.
- Keren Gruteke Klein, Shachar Frenkel, Omer Shubi, and Yevgeni Berzak. 2025. [Eye tracking based cognitive evaluation of automatic readability assessment measures](#). arXiv preprint arXiv:2502.11150.
- Keren Gruteke Klein, Yoav Meiri, Omer Shubi, and Yevgeni Berzak. 2024. [The effect of surprisal on reading times in information seeking and repeated reading](#). In *Proceedings of the 28th Conference on Computational Natural Language Learning*, pages 219–230.
- Kristina Gulordava, Piotr Bojanowski, Edouard Grave, Tal Linzen, and Marco Baroni. 2018. [Colorless green recurrent networks dream hierarchically](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1195–1205.
- Cfir Avraham Hadar, Omer Shubi, Yoav Meiri, Amit Heshes, and Yevgeni Berzak. 2026. [Decoding open-ended information seeking goals from eye movements in reading](#). In *Proceedings of the Fourteenth International Conference on Learning Representations*.
- Michael Hahn and Frank Keller. 2023. [Modeling task effects in human reading with neural network-based attention](#). *Cognition*, 230:105289.
- John Hale. 2001. [A probabilistic Earley parser as a psycholinguistic model](#). In *Second Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL 2001)*, pages 1–8.
- Patrick Haller, Lena S. Bolliger, and Lena A. 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*, pages 7878–7892.
- Patrick Haller, Cui Ding, Maja Stegenwallner-Schütz, David R. Reich, Iva Koncic, Silvia Makowski, and Lena A. Jäger. 2026. [Replicate me if you can: Assessing measurement reliability of individual differences in reading across measurement occasions and methods](#). *Cognitive Science*, 50(1):e70121.
- Patrick Haller, Andreas Säuberli, Sarah Kiener, Jinger Pan, Ming Yan, and Lena A. Jäger. 2022. [Eye-tracking based classification of Mandarin Chinese readers with and without dyslexia using neural sequence models](#). In *Proceedings of the Workshop on Text Simplification, Accessibility, and Readability*, pages 111–118.
- Xiaoshan He, Xiaoqun Liu, Haodong He, Yu Wang, and Yang Xu. 2026. [Eye movement features can predict human preferences on machine-generated texts](#). In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, pages 470–486.
- Nora Hollenstein, Maria Barrett, and Marina Björnsdóttir. 2022a. [The Copenhagen corpus of eye tracking recordings from natural reading of Danish texts](#). In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1712–1720.
- Nora Hollenstein, Maria Barrett, Marius Troendle, Francesco Bigioli, Nicolas Langer, and Ce Zhang. 2019. [Advancing NLP with cognitive language processing signals](#). arXiv preprint arXiv:1904.02682.

- Nora Hollenstein, Emmanuele Chersoni, Cassandra L. Jacobs, Yohei Oseki, Laurent Prévot, and Enrico Santus. 2021. [CMCL 2021 shared task on eye-tracking prediction](#). In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2021)*, pages 72–78.
- Nora Hollenstein, Emmanuele Chersoni, Cassandra L. Jacobs, Yohei Oseki, Laurent Prévot, and Enrico Santus. 2022b. [CMCL 2022 shared task on multilingual and crosslingual prediction of human reading behavior](#). In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2022)*, pages 121–129.
- Nora Hollenstein, Jonathan Rotsztein, Marius Troendle, Andreas Pedroni, Ce Zhang, and Nicolas Langer. 2018. [ZuCo, a simultaneous EEG and eye-tracking resource for natural sentence reading](#). *Scientific Data*, 5(1):180291.
- Nora Hollenstein, Marius Troendle, Ce Zhang, and Nicolas Langer. 2020. [ZuCo 2.0: A dataset of physiological recordings during natural reading and annotation](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 138–146.
- Nora Hollenstein, Marius Tröndle, Martyna Plomecka, Samuel Kiegeland, Yilmazcan Özyurt, Lena A. Jäger, and Nicolas Langer. 2023. [The ZuCo benchmark on cross-subject reading task classification with EEG and eye-tracking data](#). *Frontiers in Psychology*, 13:1028824.
- Nora Hollenstein and Ce Zhang. 2019. [Entity recognition at first sight: Improving NER with eye movement information](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1–10.
- Jennifer Hu, Jon Gauthier, Peng Qian, Ethan G. Wilcox, and Roger P. Levy. 2020. [A systematic assessment of syntactic generalization in neural language models](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1725–1744.
- Stephen Hutt, Kristina Krasich, Caitlin Mills, Nigel Bosch, Shelby White, James R. Brockmole, and Sidney K. D’Mello. 2019. [Automated gaze-based mind wandering detection during computerized learning in classrooms](#). *User Modeling and User-Adapted Interaction*, 29(4):821–867.
- Stephen Hutt, Aaron Wong, Alexandra Papoutsaki, Ryan S. Baker, Joshua I. Gold, and Caitlin Mills. 2024. [Webcam-based eye tracking to detect mind wandering and comprehension errors](#). *Behavior Research Methods*, 56(1):1–17.
- Jukka Hyönä and Johanna K. Kaakinen. 2019. [Eye movements during reading](#). In Christoph Klein and Ulrich Ettinger, editors, *Eye Movement Research: An Introduction to Its Scientific Foundations and Applications*, pages 239–274. Springer, Cham, Switzerland.
- Fariz Ikhwantri, Hiroaki Yamada, and Takenobu Tokunaga. 2024. [Analyzing interpretability of summarization model with eye-gaze information](#). In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 939–950.
- Lena A. Jäger, Nora Hollenstein, Ana Matić Škorić, Deborah N. Jakobi, Maja Stegenwallner-Schütz, Cui Ding, Eva Pavlinušić Vilus, Ramunė Kasperė, and Marie-Luise Müller. 2026. [MultiPEYE: Enabling multilingual eye-tracking data collection for human and machine language processing research](#). Preregistration, PsychArchives.
- Lena A. Jäger, Silvia Makowski, Paul Prasse, Sascha Liehr, Maximilian Seidler, and Tobias Scheffer. 2020. [Deep eyedentification: Biometric identification using micro-movements of the eye](#). In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 299–314.
- Deborah N. Jakobi, Thomas Kern, David R. Reich, Patrick Haller, and Lena A. Jäger. 2025a. [PoTeC: A German naturalistic eye-tracking-while-reading corpus](#). *Behavior Research Methods*, 57(8):211.
- Deborah N. Jakobi, David R. Reich, Paul Prasse, Jana M. Hofmann, Lena S. Bolliger, and

- Lena A. Jäger. 2026. Eye-tracking-while-reading: A living survey of datasets with open library support. *Behavior Research Methods*, 58:263.
- Deborah N. Jakobi, David R. Reich, Paul Prasse, and Lena A. Jäger. 2025b. [Neural additive models uncover predictive gaze features in reading](#). In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications*.
- Deborah N. Jakobi, Maja Stegenwallner-Schütz, Nora Hollenstein, Cui Ding, Ramune Kaspere, Ana Matić Škorić, Eva Pavlinusic Vilus, Stefan L. Frank, Marie-Luise Müller, Kristine M. Jensen de López, Nik Kharlamov, Hanne B. Søndergaard Knudsen, Yevgeni Berzak, Ella Lion, Irina A. Sekerina, Cengiz Acartürk, Mohd Faizan Ansari, Katarzyna Harezlak, Paweł Kasprowski, Ana Bautista, Lisa Beinhorn, Anna Bondar, Antonia Boznou, Leah Bradshaw, Jana M. 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, Inmaculada Fajardo, Mariola Giménez-Salvador, Xavier Mínguez-López, Maroš Filip, Zigmunds Freibergs, Jéssica Gomes, Andreia Janeiro, Paula Luegi, João Veríssimo, Sasho Gramatkov, 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 Motiejūniene, Jamal A. Nasir, Johanne Sofie Krog Nedergård, Ayşegül Özkan, Mikuláš Preininger, Loredana Pungă, David R. Reich, Chiara Tschirner, Špela Rot, Andreas Säuberli, Jordi Solé-Casals, Ekaterina Strati, Igor Svoboda, Evis Trandafili, Spyridoula Varlokosta, Mila Vulchanova, and Lena A. Jäger. 2025c. [MultiPLEYE: Creating a multilingual eye-tracking-while-reading corpus](#). In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications*.
- Tamara Jakovljević, Milica M. Janković, Andrej M. Savić, Ivan Soldatović, Gordana Čolić, Tadeja Jere Jakulin, Gregor Papa, and Vanja Ković. 2021. [The relation between physiological parameters and colour modifications in text background and overlay during reading in children with and without dyslexia](#). *Brain Sciences*, 11(5):539.
- Chuhan Jiao, Guanhua Zhang, Yeonjoo Cho, Zhiming Hu, and Andreas Bulling. 2026. [DiffEyeSyn: User-specific subtle eye movement synthesis using diffusion models](#). In *Proceedings of the 2026 IEEE 20th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–10.
- A. Jothi Prabha and R. Bhargavi. 2020. [Predictive model for dyslexia from fixations and saccadic eye movement events](#). *Computer Methods and Programs in Biomedicine*, 195:105538.
- Marcel A. Just and Patricia A. Carpenter. 1980. [A theory of reading: From eye fixations to comprehension](#). *Psychological Review*, 87(4):329–354.
- Marcel A. Just, Patricia A. Carpenter, and Jacqueline D. Woolley. 1982a. [Paradigms and processes in reading comprehension](#). *Journal of Experimental Psychology: General*, 111(2):228.
- Marcel Adam Just, Patricia A Carpenter, and MEJ Masson. 1982b. What eye fixations tell us about speed reading and skimming. *Eye-lab Technical Report Carnegie-Mellon University, Pittsburgh*.
- Johanna K. Kaakinen, Jukka Hyönä, and Janice M. Keenan. 2002. [Perspective effects on online text processing](#). *Discourse Processes*, 33(2):159–173.
- Tobiasz Kaduk, Caspar Goeke, Holger Finger, and Peter König. 2024. [Webcam eye tracking close to laboratory standards: Comparing a new webcam-based system and the EyeLink 1000](#). *Behavior Research Methods*, 56(5):5002–5022.
- Paweł Kasprowski and Józef Ober. 2004. Eye movements in biometrics. In *Biometric Authentication*, pages 248–258, Berlin, Heidelberg, Germany. Springer.
- Alan Kennedy, Robin Hill, and Joël Pynte. 2003. The Dundee Corpus. In *12th European Conference on Eye Movements*. Poster presentation.

- Varun Khurana, Yaman Kumar, Nora Hollenstein, Rajesh Kumar, and Balaji Krishnamurthy. 2023. [Synthesizing human gaze feedback for improved NLP performance](#). In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1895–1908.
- Samuel Kiegele, David R. Reich, Ryan Cotterell, Lena A. Jäger, and Ethan G. Wilcox. 2024. [The pupil becomes the master: Eye-tracking feedback for tuning LLMs](#). In *Proceedings of the ICML 2024 Workshop on LLMs and Cognition*.
- Sigrid Klerke, Yoav Goldberg, and Anders Søgaard. 2016. [Improving sentence compression by learning to predict gaze](#). In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1528–1533.
- Reinhold Kliegl, Ellen Grabner, Martin Rolfs, and Ralf Engbert. 2004. [Length, frequency, and predictability effects of words on eye movements in reading](#). *European Journal of Cognitive Psychology*, 16(1-2):262–284.
- Victor Kuperman, Sascha Schroeder, Cengiz Acartürk, Niket Agrawal, Dominick M. Alexandre, Lena S. Bolliger, Jan Brasser, César Campos-Rojas, Denis Drieghe, Dušica Filipović Đurđević, Luiz Vinicius Gadelha de Freitas, Sofya Goldina, Romualdo Ibáñez Orellana, Lena A. Jäger, Ómar I. Jóhannesson, Anurag Khare, Nik Kharlamov, Hanne B. S. Knudsen, Árni Kristjánsson, Charlotte E. Lee, Jun Ren Lee, Marina P. T. Leite, Simona Mancini, Nataša Mihajlović, Ksenija Mišić, Miloslava Orekhova, Olga Parshina, Milica Popović Stijačić, Athanassios Protopapas, David R. Reich, Anurag Rimzhim, Rui Rothe-Neves, Thais M. M. Sá, Andrea Santana Covarrubias, Irina A. Sekerina, Heida M. Sigurdardottir, Anna Smirnova, Priyanka Srivastava, Elisangela N. Teixeira, Ivana Ugrinic, Kerem Alp Usal, Karolina Vakulya, João M. M. Vieira, Ark Verma, Denise H. Wu, Jin Xue, Sunčica Zdravković, Junjing Zhuo, Laoura Zizaka, and Noam Siegelman. 2025. [New data on text reading in English as a second language: The wave 2 expansion of the Multilingual Eye-Movement Corpus \(MECO\)](#). *Studies in Second Language Acquisition*, 47(2):677–695.
- Victor Kuperman, Noam Siegelman, Sascha Schroeder, Cengiz Acartürk, Svetlana Alexeeva, Simona Amenta, Raymond Bertram, Rolando Bonandrini, Marc Brysbaert, Daria Chernova, Sara Maria Da Fonseca, Nicolas Dirix, Wouter Duyck, Argyro Fella, Ram Frost, Carolina A. Gattei, Areti Kalaitzi, Kaidi Lõo, Marco Marelli, Kelly Nisbet, Timothy C. Papadopoulos, Athanassios Protopapas, Satu Savo, Diego E. Shalom, Natalia Slioussar, Roni Stein, Longjiao Sui, Analí Taboh, Veronica Tønnesen, and Kerem Alp Usal. 2023. [Text reading in English as a second language: Evidence from the Multilingual Eye-Movements Corpus](#). *Studies in Second Language Acquisition*, 45(1):3–37.
- Victor Kuperman and Julie A Van Dyke. 2011. [Effects of individual differences in verbal skills on eye-movement patterns during sentence reading](#). *Journal of Memory and Language*, 65(1):42–73.
- Tatsuki Kuribayashi, Yohei Oseki, Souhaib Ben Taieb, Kentaro Inui, and Timothy Baldwin. 2025. [Large language models are human-like internally](#). *Transactions of the Association for Computational Linguistics*, 13:1743–1766.
- Anna Laurinavichyute, Anastasiya Lopukhina, and David R. Reich. 2025. [Automatic detection of dyslexia based on eye movements during reading in Russian](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 59–66.
- Roger P. Levy. 2008. [Expectation-based syntactic comprehension](#). *Cognition*, 106(3):1126–1177.
- John D. Lewis and Jeffrey L. Elman. 2001. Learnability and the statistical structure of language: Poverty of stimulus arguments revisited. In *BUCLD 26: Proceedings of the 26th Annual Boston University Conference on Language Development*, pages 359–370.
- Richard L. Lewis and Shravan Vasishth. 2005. [An activation-based model of sentence processing as skilled memory retrieval](#). *Cognitive Science*, 29(3):375–419.

- Jiaxin Li, Michael Hahn, and Vera Demberg. 2026. [Modeling working memory capacity in sentence processing with lossy-context surprisal](#). PsyArXiv preprint.
- Ping Li and Roy B. Clariana. 2019. [Reading comprehension in L1 and L2: An integrative approach](#). *Journal of Neurolinguistics*, 50:94–105.
- Tal Linzen, Emmanuel Dupoux, and Yoav Goldberg. 2016. [Assessing the ability of LSTMs to learn syntax-sensitive dependencies](#). *Transactions of the Association for Computational Linguistics*, 4:521–535.
- Ángela López-Cardona, Sebastian Idesis, and Ioannis Arapakis. 2025a. [Integrating cognitive processing signals into language models: A review of advances, applications and future directions](#). In *Proceedings of the 2025 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.
- Ángela López-Cardona, Sebastian Idesis, Miguel Barreda-Ángeles, Sergi Abadal, and Ioannis Arapakis. 2025b. [OASST-ETC dataset: Alignment signals from eye-tracking analysis of LLM responses](#). *Proceedings of the ACM on Human-Computer Interaction*, 9(3).
- Ángela López-Cardona, Carlos Segura, Alexandros Karatzoglou, Sergi Abadal, and Ioannis Arapakis. 2025c. [Seeing eye to AI: Human alignment via gaze-based response rewards for large language models](#). In *Proceedings of the Thirteenth International Conference on Learning Representations*. International Conference on Learning Representations (ICLR 2025).
- Andreas Madsen, Siva Reddy, and Sarath Chandar. 2022. [Post-hoc interpretability for neural NLP: A survey](#). *ACM Computing Surveys*, 55(8):1–42.
- Kishan Maharaj, Ashita Saxena, Raja Kumar, Abhijit Mishra, and Pushpak Bhattacharyya. 2023. [Eyes show the way: Modelling gaze behaviour for hallucination detection](#). In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 11424–11438.
- Silvia Makowski, Lena A. Jäger, Ahmed Abdelwahab, Niels Landwehr, and Tobias Schef-fer. 2019. [A discriminative model for identifying readers and assessing text comprehension from eye movements](#). In *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2018), Part I*, pages 209–225.
- Jonathan Malmaud, Roger P. Levy, and Yevgeni Berzak. 2020. [Bridging information-seeking human gaze and machine reading comprehension](#). In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 142–152.
- Titus von der Malsburg and Shravan Vasishth. 2011. [What is the scanpath signature of syntactic reanalysis?](#) *Journal of Memory and Language*, 65(2):109–127.
- Rebecca Marvin and Tal Linzen. 2018. [Targeted syntactic evaluation of language models](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1192–1202.
- Sandeep Mathias, Diptesh Kanojia, Abhijit Mishra, and Pushpak Bhattacharya. 2020. [A survey on using gaze behaviour for natural language processing](#). In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 4907–4913. Survey track.
- Kate McCurdy and Michael Hahn. 2024. [Lossy context surprisal predicts task-dependent patterns in relative clause processing](#). In *Proceedings of the 28th Conference on Computational Natural Language Learning*, pages 36–45.
- Erik McGuire and Noriko Tomuro. 2021. [Sentiment analysis with cognitive attention supervision](#). In *Proceedings of the 34th Canadian Conference on Artificial Intelligence*.
- Maryam Meghdadi, John Duff, and Vera Demberg. 2026. [Integrating language model embeddings into the ACT-R cognitive modeling framework](#). *Frontiers in Language Sciences*, 5:1721326.
- Yoav Meiri, Omer Shubi, Cfir Avraham Hadar, Ariel Kreisberg Nitzav, and Yevgeni Berzak. 2025. [Déjà vu? Decoding repeated reading from eye movements](#). In *Proceedings of*

- the 63rd Annual Meeting of the Association for Computational Linguistics, pages 19460–19482.
- Diane C. Mézière, Lili Yu, Genevieve McArthur, Erik D. Reichle, and Titus von der Malsburg. 2024. [Scanpath regularity as an index of reading comprehension](#). *Scientific Studies of Reading*, 28(1):79–100.
- Diane C. Mézière, Lili Yu, Erik D. Reichle, Titus von der Malsburg, and Genevieve McArthur. 2023. [Using eye-tracking measures to predict reading comprehension](#). *Reading Research Quarterly*, 58(3):425–449.
- Caitlin Mills, Julie Gregg, Robert Bixler, and Sidney K. D’Mello. 2021. [Eye-Mind reader: An intelligent reading interface that promotes long-term comprehension by detecting and responding to mind wandering](#). *Human–Computer Interaction*, 36(4):306–332.
- Abhijit Mishra and Pushpak Bhattacharyya. 2018. *Cognitively Inspired Natural Language Processing: An Investigation Based on Eye-tracking*. Springer, Singapore, Singapore.
- Abhijit Mishra, Diptesh Kanojia, Seema Nagar, Kuntal Dey, and Pushpak Bhattacharyya. 2016. [Leveraging cognitive features for sentiment analysis](#). In *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*, pages 156–166.
- Michael Mooney and Edmond SL Ho. 2026. Towards a scanpath-conditioned surprisal theory: Modeling reader information states. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 41366–41380.
- Aaron Mueller, Jannik Brinkmann, Millicent Li, Samuel Marks, Koyena Pal, Nikhil Prakash, Can Rager, Aruna Sankaranarayanan, Arnab Sen Sharma, Jiuding Sun, Eric Todd, David Bau, and Yonatan Belinkov. 2026. [The quest for the right mediator: Surveying mechanistic interpretability for NLP through the lens of causal mediation analysis](#). *Computational Linguistics*, 52(1):331–378.
- Boris Nerušil, Jaroslav Polec, Juraj Škunda, and Juraj Kačur. 2021. [Eye tracking based dyslexia detection using a holistic approach](#). *Scientific Reports*, 11(1):15687.
- Mattias Nilsson and Joakim Nivre. 2009. [Learning where to look: Modeling eye movements in reading](#). In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009)*, pages 93–101.
- Mattias Nilsson and Joakim Nivre. 2011. [Entropy-driven evaluation of models of eye movement control in reading](#). In *Proceedings of the 8th International Workshop on Natural Language Processing and Cognitive Science (NLPCS)*, pages 201–212. Last accessed July 30, 2026.
- Mattias Nilsson Benfatto, Gustaf Öqvist Seimyr, Jan Ygge, Tony Pansell, Agneta Rydberg, and Christer Jacobson. 2016. [Screening for dyslexia using eye tracking during reading](#). *PLOS ONE*, 11(12):e0165508.
- Byung-Doh Oh and Tal Linzen. 2026. [To model human linguistic prediction, make LLMs less superhuman](#). *Trends in Cognitive Sciences*.
- Byung-Doh Oh and William Schuler. 2023. [Why does surprisal from larger transformer-based language models provide a poorer fit to human reading times?](#) *Transactions of the Association for Computational Linguistics*, 11:336–350.
- Nikolas Papadopoulos, Shreenithi Navaneethan, Sheng Bai, Ankur Samanta, and Paul Sajda. 2026. [Gaze patterns predict preference and confidence in pairwise AI image evaluation](#). In *Proceedings of the 2026 Symposium on Eye Tracking Research and Applications*.
- Alexandra Papoutsaki, Patsorn Sangkloy, James Laskey, Nediya Daskalova, Jeff Huang, and James Hays. 2016. [WebGazer: Scalable webcam eye tracking using user interactions](#). In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pages 3839–3845.
- Paul Prasse, David R. Reich, Silvia Makowski, Seoyoung Ahn, Tobias Scheffer, and Lena A. Jäger. 2023. [Sp-EyeGAN: Generating synthetic eye movement data with generative adversarial networks](#). In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*, pages 1–9.

- Paul Prasse, David R. Reich, Silvia Makowski, Lena A. Jäger, and Tobias Scheffer. 2022. [Fairness in oculomotoric biometric identification](#). In *Proceedings of the 2022 Symposium on Eye Tracking Research and Applications*.
- Peter Raatikainen, Jarkko Hautala, Otto Loberg, Tommi Kärkkäinen, Paavo H. T. Leppänen, and Paavo Nieminen. 2021. [Detection of developmental dyslexia with machine learning using eye movement data](#). *Array*, 12:100087.
- Maximilian M. Rabe, Dario Paape, Daniela Mertzen, Shravan Vasishth, and Ralf Engbert. 2024. [SEAM: An integrated activation-coupled model of sentence processing and eye movements in reading](#). *Journal of Memory and Language*, 135:104496.
- Keith Rayner. 1998. [Eye movements in reading and information processing: 20 years of research](#). *Psychological Bulletin*, 124(3):372–422.
- Keith Rayner, Jane Ashby, Alexander Pollatsek, and Erik D Reichle. 2004. [The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model](#). *Journal of Experimental Psychology: Human Perception and Performance*, 30(4):720–732.
- Keith Rayner, Kathryn H. Chace, Timothy J. Slatery, and Jane Ashby. 2006a. [Eye movements as reflections of comprehension processes in reading](#). *Scientific Studies of Reading*, 10(3):241–255.
- Keith Rayner, Erik D. Reichle, Michael J. Stroud, Carrick C. Williams, and Alexander Pollatsek. 2006b. [The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers](#). *Psychology and Aging*, 21(3):448–465.
- Keith Rayner, Elizabeth R. Schotter, Michael E. J. Masson, Mary C. Potter, and Rebecca Treiman. 2016. [So much to read, so little time: How do we read, and can speed reading help?](#) *Psychological Science in the Public Interest*, 17(1):4–34.
- Francesco Ignazio Re, Andreas Opedal, Glib Manaiiev, Mario Giulianelli, and Ryan Cotterell. 2025. [A spatio-temporal point process for fine-grained modeling of reading behavior](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 30518–30538.
- David R. Reich, Cui Ding, Lena S. Bolliger, Patrick Haller, Paul Prasse, and Lena A. Jäger. 2025a. [Proxy-based pre-training for eye-tracking applications](#). In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications*.
- David R. Reich, Paul Prasse, and Lena A. Jäger. 2025b. [Evaluating gaze event detection algorithms: Impacts on machine learning-based classification and psycholinguistic statistical modeling](#). *Proceedings of the ACM on Human-Computer Interaction*, 9(3).
- David R. 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 *Proceedings of the 2022 Symposium on Eye Tracking Research and Applications*.
- David R. Reich, Omer Shubi, Lena A. Jäger, and Yevgeni Berzak. 2025c. [Eye tracking and NLP](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 5: Tutorial Abstracts)*, pages 2–2.
- Erik D. Reichle, Keith Rayner, and Alexander Pollatsek. 1999. [Eye movement control in reading: Accounting for initial fixation locations and re-fixations within the E-Z Reader model](#). *Vision Research*, 39(26):4403–4411.
- Erik D. Reichle, Keith Rayner, and Alexander Pollatsek. 2003. [The E-Z Reader model of eye-movement control in reading: Comparisons to other models](#). *Behavioral and Brain Sciences*, 26(4):445–526.
- Erik D. Reichle, Andrew E. Reineberg, and Jonathan W. Schooler. 2010. [Eye movements during mindless reading](#). *Psychological Science*, 21(9):1300–1310.
- Erik D. Reichle, Tessa Warren, and Kerry McConnell. 2009. [Using E-Z Reader to model the effects of higher level language processing on](#)

- eye movements during reading. *Psychonomic Bulletin & Review*, 16(1):1–21.
- Luz Rello and Miguel Ballesteros. 2015. [Detecting readers with dyslexia using machine learning with eye tracking measures](#). In *Proceedings of the 12th International Web for All Conference*, pages 1–8.
- Yuqi Ren and Deyi Xiong. 2021. [CogAlign: Learning to align textual neural representations to cognitive language processing signals](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3758–3769.
- Tiago Ribeiro, Stephanie Brandl, Anders Søgaard, and Nora Hollenstein. 2023. [WebQAmGaze: A multilingual webcam eye-tracking-while-reading dataset](#). arXiv preprint arXiv:2303.17876.
- Soo Hyun Ryu and Richard L. Lewis. 2021. [Accounting for agreement phenomena in sentence comprehension with transformer language models: Effects of similarity-based interference on surprisal and attention](#). In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2021)*, pages 61–71.
- Hassan Sajjad, Francisco Guzmán, Nadir Durrani, Ahmed Abdelali, Houda Bouamor, Irina Temnikova, and Stephan Vogel. 2016. [Eyes don’t lie: Predicting machine translation quality using eye movement](#). In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1082–1088.
- Dario D. Salvucci and Joseph H. Goldberg. 2000. [Identifying fixations and saccades in eye-tracking protocols](#). In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, pages 71–78.
- Andreas Säuberli, Darja Jepifanova, Diego Frassinelli, and Barbara Plank. 2026. [Controlling reading ease with gaze-guided text generation](#). In *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2383–2397.
- Marten van Schijndel and Tal Linzen. 2018. [A neural model of adaptation in reading](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4704–4710.
- Elizabeth R. Schotter and Brian Dillon. 2025. [A beginner’s guide to eye tracking for psycholinguistic studies of reading](#). *Behavior Research Methods*, 57(2):68.
- Elizabeth R. Schotter, Randy Tran, and Keith Rayner. 2014. [Don’t believe what you read \(only once\): comprehension is supported by regressions during reading](#). *Psychological Science*, 25(6):1218–1226.
- Stefan A. Seelig, Maximilian M. Rabe, Noa Malem-Shinitski, Sarah Risse, Sebastian Reich, and Ralf Engbert. 2020. [Bayesian parameter estimation for the SWIFT model of eye-movement control during reading](#). *Journal of Mathematical Psychology*, 95:102313.
- Marina Serrano-Carot, Bernhard Angele, Hemu Xu, and Martin R. Vasilev. 2025. [Webcams can be used to study eye movements during reading](#). PsyArXiv preprint.
- Cory Shain, Clara Meister, Tiago Pimentel, Ryan Cotterell, and Roger P. Levy. 2024. [Large-scale evidence for logarithmic effects of word predictability on reading time](#). *Proceedings of the National Academy of Sciences*, 121(10):e2307876121.
- Soroosh Shalileh, Dmitry Ignatov, Anastasiya Lopukhina, and Olga Dragoy. 2023. [Identifying dyslexia in school pupils from eye movement and demographic data using artificial intelligence](#). *PLOS ONE*, 18(11):e0292047.
- Soroosh Shalileh, Matvey Kairov, Ranga Baminiwatte, Olga Parshina, and Olga Dragoy. 2024. [Predicting first-language and second-language proficiency using eye fixation data and demographic information: Assumptions, data representations, and methods](#). *IEEE Access*, 12:145832–145844.
- Omer Shubi and Yevgeni Berzak. 2023. [Eye movements in information-seeking reading](#). In

- Proceedings of the 45th Annual Meeting of the Cognitive Science Society*, pages 936–943.
- Omer Shubi, Cfir Avraham Hadar, and Yevgeni Berzak. 2025a. [Decoding reading goals from eye movements](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics*, pages 5616–5637.
- Omer Shubi, Yoav Meiri, Cfir Avraham Hadar, and Yevgeni Berzak. 2024. [Fine-grained prediction of reading comprehension from eye movements](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 3372–3391.
- Omer Shubi, David R. Reich, Keren Gruteke Klein, Yuval Angel, Paul Prasse, Lena A. Jäger, and Yevgeni Berzak. 2025b. [EyeBench: Predictive modeling from eye movements in reading](#). In *Advances in Neural Information Processing Systems 38 (NeurIPS 2025) Datasets and Benchmarks Track*.
- Noam Siegelman, Sascha Schroeder, Cengiz Acartürk, Hee-Don Ahn, Svetlana Alexeeva, Simona Amenta, Raymond Bertram, Rolando Bonandrini, Marc Brysbaert, Daria Chernova, Sara Maria Da Fonseca, Nicolas Dirix, Wouter Duyck, Argyro Fella, Ram Frost, Carolina A. Gattei, Areti Kalaitzi, Nayoung Kwon, Kaidi Lõo, Marco Marelli, Timothy C. Papadopoulos, Athanassios Protopapas, Satu Savo, Diego E. Shalom, Natalia Slioussar, Roni Stein, Longjiao Sui, Analí Taboh, Veronica Tønnesen, Kerem Alp Usal, and Victor Kuperman. 2022. [Expanding horizons of cross-linguistic research on reading: The Multilingual Eye-Movement Corpus \(MECO\)](#). *Behavior Research Methods*, 54(6):2843–2863.
- Noam Siegelman, Sascha Schroeder, Yaqian B. Bao, Cengiz Acartürk, Niket Agrawal, Lena S. Bolliger, Jan Brasser, César Campos-Rojas, Denis Drieghe, Dušica Filipović Đurđević, Sofya Goldina, Romualdo Ibáñez Orellana, Lena A. Jäger, Ómar I. Jóhannesson, Anurag Khare, Nik Kharlamov, Hanne B. S. Knudsen, Árni Kristjánsson, Charlotte E. Lee, Jun Ren Lee, Marina P. T. Leite, Simona Mancini, Nataša Mihajlović, Ksenija Mišić, Miloslava Orekhova, Olga Parshina, Milica Popović Stijačić, Athanassios Protopapas, David R. Reich, Anurag Rimzhim, Rui Rothe-Neves, Thais M. M. Sá, Andrea Santana-Covarrubias, Irina A. Sekerina, Heida M. Sigurdardottir, Anna Smirnova, Priyanka Srivastava, Elisangela N. Teixeira, Ivana Ugrinic, Kerem Alp Usal, Karolina Vakulya, Ark Verma, João M. M. Vieira, Denise H. Wu, Jin Xue, Sunčica Zdravković, Junjing Zhuo, Laoura Zikaka, and Victor Kuperman. 2025. [Wave 2 of the Multilingual Eye-Movement Corpus \(MECO\): New text reading data across languages](#). *Scientific Data*, 12(1):1183.
- Lina Skerath, Paulina Toborek, Anita Zielińska, Maria Barrett, and Rob van der Goot. 2023. [Native language prediction from gaze: A reproducibility study](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pages 152–159.
- Iza Škrjanec and Vera Demberg. 2026. [Language models that match reader experience are better predictors of reading times](#). *Journal of Memory and Language*, 146:104677.
- Garrett Smith and Shravan Vasishth. 2020. [A principled approach to feature selection in models of sentence processing](#). *Cognitive Science*, 44(12):e12918.
- Nathaniel J. Smith and Roger P. Levy. 2013. [The effect of word predictability on reading time is logarithmic](#). *Cognition*, 128(3):302–319.
- Joshua Snell, Sam van Leipsig, Jonathan Grainger, and Martijn Meeter. 2018. [OB1-reader: A model of word recognition and eye movements in text reading](#). *Psychological Review*, 125(6):969–984.
- Ekta Sood, Prajit Dhar, Enrica Troiano, Rosy Southwell, and Sidney K. D’Mello. 2025. [ScanEZ: Integrating cognitive models with self-supervised learning for spatiotemporal scanpath prediction](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1132–1142.
- Ekta Sood, Fabian Kögel, Philipp Müller, Dominike Thomas, Mihai Băce, and Andreas Bulling. 2023. [Multimodal integration of human-like attention in visual question answering](#). In *Proceedings of the IEEE/CVF Confer-*

- ence on Computer Vision and Pattern Recognition Workshops (CVPRW).
- Ekta Sood, Simon Tannert, Diego Frassinelli, Andreas Bulling, and Ngoc Thang Vu. 2020a. Interpreting attention models with human visual attention in machine reading comprehension. In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 12–25.
- Ekta Sood, Simon Tannert, Philipp Mueller, and Andreas Bulling. 2020b. Improving natural language processing tasks with human gaze-guided neural attention. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, pages 6327–6341.
- Adrian Staub. 2011. The effect of lexical predictability on distributions of eye fixation durations. *Psychonomic Bulletin & Review*, 18(2):371–376.
- Adrian Staub. 2021. How reliable are individual differences in eye movements in reading? *Journal of Memory and Language*, 116:104190.
- Adrian Staub, Sarah J. White, Denis Drieghe, Elizabeth C. Hollway, and Keith Rayner. 2010. Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5):1280–1293.
- Ben Steichen, Wilsen Kosasih, and Christian Berra. 2022. Inferring search user language proficiency from eye gaze data. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval (CHIIR)*, pages 211–220.
- Michalina Strzyz, David Vilares, and Carlos Gómez-Rodríguez. 2019. Towards making a dependency parser see. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1500–1506.
- Ece Takmaz, Sandro Pezzelle, Lisa Beinborn, and Raquel Fernández. 2020. Generating image descriptions via sequential cross-modal alignment guided by human gaze. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 4664–4677.
- Abishek Thamma and Micha Heilbron. 2026. Human-like fleeting memory improves language learning but impairs reading time prediction in transformer language models. *Transactions of the Association for Computational Linguistics*, 14:877–892.
- Andrea Gregor de Varda and Marco Marelli. 2024. Locally biased transformers better align with human reading times. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2024)*, pages 30–36.
- Xintong Wang, Xiaoyu Li, Xingshan Li, and Chris Biemann. 2024. Probing large language models from a human behavioral perspective. In *Proceedings of the Workshop on Bridging Neurons and Symbols for Natural Language Processing and Knowledge Graphs Reasoning (Neusym-Bridge) at LREC-COLING 2024*, pages 1–7.
- Alex Warstadt, Alicia Parrish, Haokun Liu, Anhad Mohanney, Wei Peng, Sheng-Fu Wang, and Samuel R. Bowman. 2020. BLiMP: The benchmark of linguistic minimal pairs for English. *Transactions of the Association for Computational Linguistics*, 8:377–392.
- Ethan G. Wilcox, Cui Ding, Mrinmaya Sachan, and Lena A. Jäger. 2024a. Mouse tracking for reading (MoTR): A new naturalistic incremental processing measurement tool. *Journal of Memory and Language*, 138:104534.
- Ethan G. Wilcox, Richard Futrell, and Roger P. Levy. 2024b. Using computational models to test syntactic learnability. *Linguistic Inquiry*, 55(4):805–848.
- Ethan G. Wilcox, Tiago Pimentel, Clara Meister, Ryan Cotterell, and Roger P. Levy. 2023. Testing the predictions of surprisal theory in 11 languages. *Transactions of the Association for Computational Linguistics*, 11:1451–1470.
- Ethan G. Wilcox, Pranali Vani, and Roger P. Levy. 2021. A targeted assessment of incremental processing in neural language models and humans. In *Proceedings of the 59th Annual Meeting of the Association for Computational Lin-*

guistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 939–952.

Weijie Xu, Brian Dillon, and Richard Futrell. 2026. [Memory efficiency and resource-rational encoding in sentence processing](#). In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 33603–33618.

Duo Yang and Nora Hollenstein. 2023. [PLM-AS: Pre-trained language models augmented with scanpaths for sentiment classification](#). In *Proceedings of the 4th Northern Lights Deep Learning Conference*.

Jie Yu, Wenya Kong, and Fangfang Liu. 2024. [CeER: A nested name entity recognition model incorporating gaze feature](#). In *Proceedings of the Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data*, pages 32–45.

Sina Zarrieß, Sebastian Loth, and David Schlangen. 2015. [Reading times predict the quality of generated text above and beyond human ratings](#). In *Proceedings of the 15th European Workshop on Natural Language Generation (ENLG)*, pages 38–47.

Dongsen Zhang, Peipei Li, Zekun Li, Yiwei Ru, Huijia Wu, and Zhaofeng He. 2025. [Eye movements as images: A multimodal framework for eye movements representation](#). In *Proceedings of the 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5.

Leran Zhang and Nora Hollenstein. 2024. [Eye-tracking features masking transformer attention in question-answering tasks](#). In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 7057–7070.