A Comparative Study Between Mouse and Eye Tracking Signals for Long Romanian Texts

Alexandru Bogdan Gheorghe Sergiu Nisioi

Human Language Technologies Research Center
Faculty of Mathematics and Computer Science
University of Bucharest

ghrghbogdan@gmail.com, sergiu.nisioi@unibuc.ro

Abstract

Understanding human language processing via eye-tracking (ET) is precise but limited by scalability. Mouse-Tracking (MoTR) offers a cost-effective alternative, yet its viability for long-form reading in languages like Romanian remains underexplored. The primary challenge lies in the motor-induced noise and biomechanical discrepancies between hand and eye movements. Here we show that combining targeted technical enhancements with a Hertz-based velocity transformation suggests that MoTR can capture significant cognitive signals comparable to ET. We evaluate this by training a BERT-enhanced Fusion Model that integrates semantic context to bridge the mechanical gap, achieving an internal consistency of $\rho \approx 0.58$ and a cross-modal correlation of $\rho \approx 0.22$ in the velocity domain. These results indicate that when properly normalized, manual tracking captures similar cognitive constraints as gaze, with predictive accuracy approaching the empirical bounds of human behavioral variance.

Keywords: Mouse-tracking, Eye-tracking, Cognitive Modeling, BERT Fusion, Gaze Prediction

1. Introduction

Although eye-tracking (ET) remains the gold standard for capturing the 'Eye-Mind' connection, the prohibitive cost of hardware and the constraints of laboratory environments pose significant barriers to data collection, particularly for low-resource languages. While initiatives like the MultiplEYE project[1] aim to provide standardized multilingual corpora, the data available for the Romanian language remain scarce compared to English (Kennedy et al., 2003; Luke and Christianson, 2018). On the one hand, Romanian shares certain structural similarities with other Romance languages, but on the other hand it presents particular challenges: the widespread use of diacritics increases visual crowding and decoding difficulty, a relatively flexible word order that complicates incremental syntactic integration and increases word-level surprisal.

Mouse-tracking while reading (MoTR) has emerged as a promising proxy, contributing to large-scale browser-based experiments Wilcox et al. (2024). Despite its potential, the transition from ocular to manual tracking introduces significant motor noise. We identify two critical bottlenecks: first, the "noise" generated during return sweeps, where the cursor inadvertently records data while transitioning between lines; and second, the variability in participants' manual dexterity. We observed that maintaining the cursor precisely on a text line is a demanding task that varies significantly across individuals, often leading to vertical "drifting" that corrupts the signal .

To address these challenges, we propose an Enhanced MoTR Framework (illustrated in Figure 1) featuring structural interface modifications: a Row-Level Gatekeeper to isolate line-specific reading and a Vertical Axis Constraint (vertical lock) to compensate for varying dexterity. Due to the restrictions of the gatekeeper system, we implement a click-to-release mechanism that enables intentional regressions, allowing users to manually bypass line constraints to revisit previous segments without compromising the system's automated tracking integrity.

Beyond structural interface modifications, we consider the mathematical representation of reading behavior. In our observations, reading durations in MoTR typically follow a more pronounced right-skewed, long-tail distribution, which can be suboptimal for linear analyses such as Pearson's correlation. This is problematic because it may distort the perceived alignment between modalities by over-emphasizing high-duration motor outliers. To mitigate this, we explore projecting these durations into the velocity domain (Hertz). Such transformation provides a natural normalization of the data, potentially stabilizing the signal and leading to a more consistent comparison between the different mechanical scales of ocular and manual movements.

Finally, we show how surface-level linguistic features (word length or frequency) can be augmented

All authors are corresponding authors.

[1] https://multipleye.eu

Proceedings of The Second International Workshop onEye-Tracking Resources and Evaluation for Human-Aligned NLP, pages 41–49
May 12, 2026.

to better capture the cognitive effort. Thus, we integrate contextual embeddings into our predictive engine. Observing how latent semantic difficulty influences reading behavior provides a more nuanced lens through which to evaluate the alignment between manual and ocular signals, rather than relying solely on surface-level statistics.

2. Related Work

The foundation of eye-tracking research in psycholinguistics rests on the Eye-Mind Hypothesis (Just and Carpenter, 1980), which posits that gaze duration is a direct proxy for cognitive processing time. Landmark studies by Rayner (1998); Rayner et al. (2003) established how lexical features, such as word length and frequency, systematically influence fixation patterns. More recently, the MultiplEYE initiative (Jäger et al., 2026) has expanded these by developing standardized multilingual corpora. However, collecting high-fidelity ocular data for the Romanian language remains a significant challenge, with the MultiplEYE project standing as the sole major initiative currently addressing this lack of standardized resources.

As a scalable alternative to expensive eye-tracking hardware, Wilcox et al. (2024) introduced the MoTR paradigm, demonstrating that cursor trajectories can successfully capture fundamental psycholinguistic effects. Despite its potential, existing literature (Wilcox et al., 2024; Huang et al., 2012) highlights that manual signals are inherently "noisier" than ocular ones. Challenges such as varying manual dexterity and the noise generated during "return sweeps" (transitions between lines) remain significant hurdles for signal precision.

Most existing studies rely on single-sentence stimuli or short fragments to minimize motor variance (Oğuz et al., 2025; Popescu and Nisioi, 2025). While it is effective for isolating lexical effects, these approaches do not capture the complexity of long-form naturalistic reading. As the text length increases, the cumulative effects of fatigue become more pronounced, leading to errors. Recent task-oriented datasets like PreferRead (de Langis et al., 2025) have begun using MoTR for longer texts to monitor LLM annotators, but they focus on evaluation rather than naturalistic reading.

Initially tested on English passages from the Provo Corpus (Wilcox et al., 2024), the method has recently been expanded to other languages. For instance, Oğuz et al. (2025) utilized MoTR to probe gender agreement in Russian, while Haveriku et al. (2025) explored its feasibility for Albanian. In the Romanian context, Popescu and Nisioi (2025) provided the first application of MoTR using isolated sentences from the MLSP dataset.

Traditional predictive models of reading behavior have relied on surface-level features, such as word length or Zipf frequency (Kliegl et al., 2004a). The integration of pre-trained language models like BERT (Devlin et al., 2019) provides a means to represent semantic difficulty and "spillover effects". By utilizing PCA for dimensionality reduction (Pedregosa et al., 2011), we aim to focus the model on the most significant semantic patterns while facilitating a more stable training process.

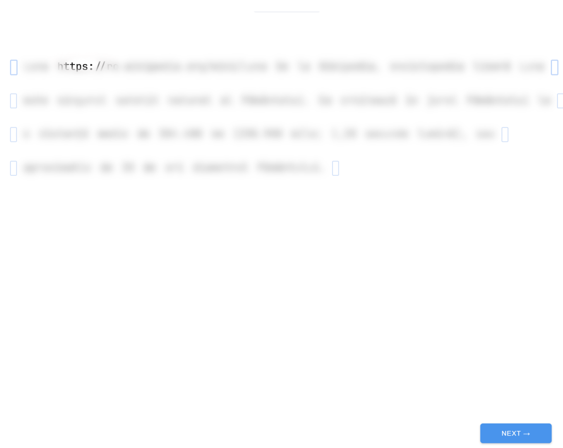


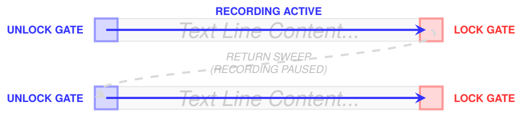
Figure 1: The enhanced MoTR interface featuring the Row-Level Gatekeeper system and the spotlight mechanism.

3. Methodology

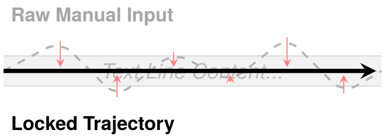
We propose a multi-stage pipeline that integrates structural modifications to the Mouse-Tracking while Reading (MoTR) paradigm with a specialized data-processing workflow. These enhancements are specifically made to minimize motor-induced noise and enforce a tighter coupling between manual displacement and cognitive processing. The process is organized into two primary stages: an experimental setup involving interface constraints and dual-modality data collection, followed by a computational pipeline that integrates speed-based normalization and contextual modeling.

3.1. Experimental Setup and Technical Enhancements

The study involved 18 participants, divided into two equal groups: 9 subjects for the Mouse-Tracking (MoTR) tests and 9 subjects for the Eye-Tracking (ET) baseline. The MoTR group consisted of university-level students aged between 20 and 29, all of whom are native Romanian speakers. This specific age range was selected to ensure that all participants had similar computer skills and reading habits, reducing the risk of motor-skill differences affecting the data.



(a) Row-Level Gatekeeper mechanism.



(b) Vertical Axis Constraint (OY lock).

Figure 2: Functional enhancements to the MoTR interface for noise reduction.

The reading materials were selected from the Romanian sub-corpus of the MultiplEYE project (Jakobi et al., 2026; Nisioi et al., 2026; Kasperė et al., 2026). We used these texts because they are standardized and provide a consistent mix of different genres and difficulty levels. To ensure that the data reflects genuine cognitive effort, each reading session was followed by six comprehension questions with four response options. These questions served as a control mechanism.

Furthermore, the questions were designed by the MultiplEYE team (Hollenstein et al., 2026) to assess the active reading of the text. Instead of focusing on simple surface-level facts, they required participants to integrate information and process the deeper meaning of the content, ensuring that the tracking data reflects high-level cognitive engagement.

To minimize motor-induced noise and improve data quality, we introduced three critical functional improvements to the original MoTR paradigm. These modifications were specifically designed to ensure that the mouse movements reflect the user's mental focus as accurately as possible.

Row-Level Gatekeeper System: We implemented this system (see Figure 2a) to separate actual reading time from the "noise" created during line transitions. In standard reading, the time spent moving the cursor from the end of one line to the start of the next (the "return sweep") does not represent text processing and implies a significant amount of noise. By using neutral-colored trigger boxes to "unlock" and "lock" each row, we ensure that we only record the time when the user is actively engaged with the words on that specific line.

Vertical Axis Constraint (OY lock): This feature (visualized in Figure 2b) was added to prevent the cursor from drifting vertically between rows. During our initial observations, we noticed that participants have different levels of manual dexterity; without a lock, the cursor would often slip off the line, making the data messy and adding cognitive effort to the participant. By locking the vertical axis, we force the "spotlight" to stay perfectly centered on the text, ensuring that the recorded movement reflects the user's cognitive progress rather than accidental hand instability.

Metric	Raw	Z-score	Log	Hertz
Total Duration	0.13	0.13	0.07	**0.23**
First Duration	0.08	0.08	0.07	**0.17**

Table 1: Pearson correlation coefficients across different normalization techniques on MoTR and ET means.

Nonlinear Navigation and Snapping: To allow for natural reading habits, such as going back to re-read a word, we developed a click-and-drag mechanism. When a user wants to re-read a part of the text, they simply hold the click and drag to the word that they want to read again. When the mouse button is released on the specific word, the cursor automatically "snaps" to the row and reactivates the vertical lock. This modification helps the user to move freely through the text while keeping the data structured and perfectly aligned with the rows.

3.2. Data Pipeline and Predictive Modeling

Both MoTR and ET recordings were processed using the same pipeline to ensure they could be compared directly. We used unique word identifiers to align the tracking signals with each specific token in the text. To model genuine reading behavior, we filtered out data points that were too short to represent cognitive effort—specifically durations under 160 ms for MoTR and 80 ms for ET. These thresholds account for the natural physical limits of the hand and the eye (Rayner, 1998; Card et al., 1983), removing "noise" from the dataset.

Statistical analyses of reading times often encounter 'long-tail' distributions in raw millisecond data, which can distort linear correlations (Kliegl et al., 2004b). While log-transformations are a standard approach for variance stabilization and Z-score normalization is frequently used for scale alignment, the latter remains a linear transformation that fails to address the underlying non-normality and skewness of the data, we explore as an alternative the projection of these durations into the velocity domain (Hertz). This transition is achieved by inverting the temporal duration (d) recorded for each word using the formula $Hz = 1000/d$, effectively converting 'time spent' into 'processing speed'. As demonstrated in Table 1, this approach aims to provide a more robust basis for cross-modal com-

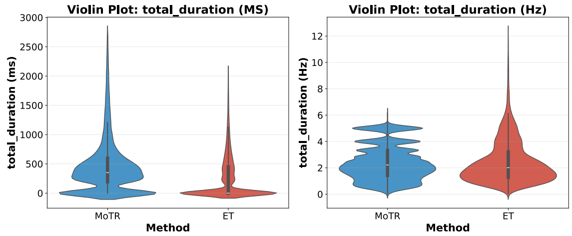


Figure 3: MoTR and ET duration distributions in ms (left) and Hz (right). The Hz projection aligns core densities and stabilizes MoTR variance, while ET's extended tail reflects superior ocular velocity.

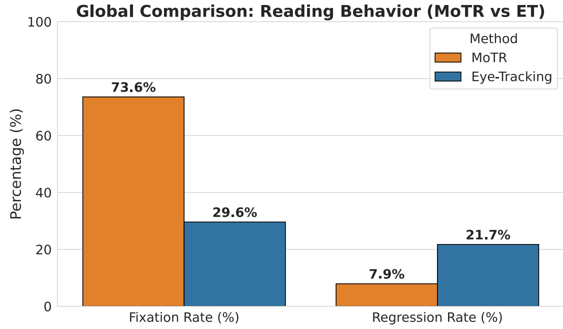


Figure 4: Global reading metrics comparison between MoTR and ET. The manual modality (MoTR) exhibits a significantly higher fixation rate (73.6%), while spontaneous regressions are more frequent in eye-tracking (21.7%), reflecting the mechanical differences between cursor-based and ocular reading behaviors.

parison—referring here to the statistical alignment between the ocular (ET) and manual (MoTR) tracking signals. By doing so, we seek to normalize the mechanical lag between hand and eye movements into a unified, speed-based metric that may better reflect the underlying cognitive pace.

Converting to Hertz (speed) balances these values, making the data more consistent and easier for our models to process. Also, projecting the data into a velocity space (speed) makes the mouse and eye signals look more alike (Figure 3). This transformation helps the model to focus on the overall "pace" of reading, which we found to be a more stable signal for predicting eye movements from mouse data.

The predictive **Baseline** employs a Random Forest Regressor with 100 estimators and a minimum leaf size of 10 selected for its robustness in handling heterogeneous features and its capacity to capture non-linear interactions between lexical attributes without assuming a specific data distribution. We compared two distinct configurations to evaluate how different types of information affect the prediction of reading speed. This model uses basic word features, such as Zipf frequency (Speer and Chin, 2021), which provides a logarithmic measure of word commonality on a scale (typically 0-8) rather than raw occurrence counts, word length, and syllable count (Kozea, 2025) that utilizes **Hunspell** hyphenation rules to account for the phonological complexity of each token.

To capture "spillover effects", where the difficulty of a previous word affects the current one, the model considers a window of two words before and one word after the target token. While the semantic embeddings are decontextualized, the situational context is explicitly modeled through this sliding window of linguistic features. Specifically, the regressor is provided with the left context (frequency and length of t-1,t-2) and right context (the immediate succeeding word t+1), alongside the relative position of the word within the paragraph. To enhance the baseline by adding semantic context we decided to make a **Fusion Model** with our Random Forest Regression and BERT (`bert-base-multilingual-cased`). To handle tokenization, each word was processed as an independent sequence from which we extracted the hidden state of the [CLS] token from the final hidden layer. This captures the highest level of semantic abstraction and serves as a pre-computed aggregate representation for all sub-word units generated by the tokenizer for that specific word. To maintain model efficiency and robustness, we utilized Principal Component Analysis (PCA) to condense the 768-dimensional BERT embeddings into 32 principal components, retaining 95% of the semantic variance while significantly reducing the feature space (Takeshita et al., 2025). This dimensionality reduction prevents the Random Forest Regressor from overfitting to the high-dimensional noise often found in large embeddings, thereby ensuring the model generalizes better to unseen texts rather than simply memorizing the training data. The models were trained and tested using an 80/20 split. Performance is reported using Spearman's rank correlation (ρ) and Pearson's (r) to measure the alignment between predicted and real reading speeds, along with Mean Absolute Error (MAE) to track the average prediction gap.

4. Results and Discussion

4.1. Behavioral Analysis and Signal Validations

The fundamental behavioral discrepancies between the two tracking modalities are illustrated in Figure 4. In this context, we define the fixation rate as the percentage of unique tokens receiving at least one tracking event, while the regression rate represents the proportion of words revisited through backward movements. MoTR and ET exhibit divergent reading signatures: MoTR shows a significantly higher fixation rate (73.6%) compared to ET (29.6%), whereas ET reveals a much higher regression rate (21.7%) than MoTR (7.9%). The data indicates that the MoTR paradigm enforces a linear reading trajectory, which directly accounts for the higher fixation rate observed. While ocular reading is rapid and almost effortless for text scanning, the manual interface requires a deliberate line-by-line progression. Consequently, the regression rate is significantly suppressed in MoTR because manual cursor repositioning imposes a substantial physical and cognitive load on the participant. Because the overall time spent on each token is inherently higher in the mouse-tracking modality, users may achieve deeper initial processing, resorting to backward movements as a last resort only when comprehension is critically challenged. In essence, the mechanical effort of the hand serves as a filter that occurs only in the most cognitively necessary regressions.

To quantify the initial alignment between the two modalities, we calculated Pearson correlations based on the means of all participants. As detailed in Table 1, projecting reading metrics into the velocity domain (Hz) almost doubles the alignment for total duration, rising from $r = 0.13$ to $r = 0.23$. This improvement suggests that processing speed provides a stable basis for cross-modal comparison than raw temporal data, as it successfully compensates for the mechanical inertia and the inherent gap between hand and eye movements. This alignment is further reinforced by analyzing the Word Length Effect, a standard psycholinguistic benchmark (Rayner, 1998). As illustrated in Figure 6, the correlation in the velocity domain reaches $r = 0.95$ for the total duration and $r = 0.89$ for the first duration, providing strong empirical evidence that our enhanced MoTR framework reflects the same fundamental linguistic processing speeds as high-fidelity eye-tracking.

4.2. Predictive Performance and Behavioral Consistency

To establish a framework for evaluating our results, we first defined an empirical upper bound through an inter-participant consistency analysis. We opted for this pairwise consistency measure over traditional split-half reliability to more effectively assess the robustness of the MoTR signal across the participant pool. Given our sample size, the pairwise approach ensures that the consistency estimate remains independent of arbitrary data partitioning, whereas split-half reliability would have been susceptible to selection bias. As illustrated in Figure 5, the MoTR signal maintains an average inter-reader

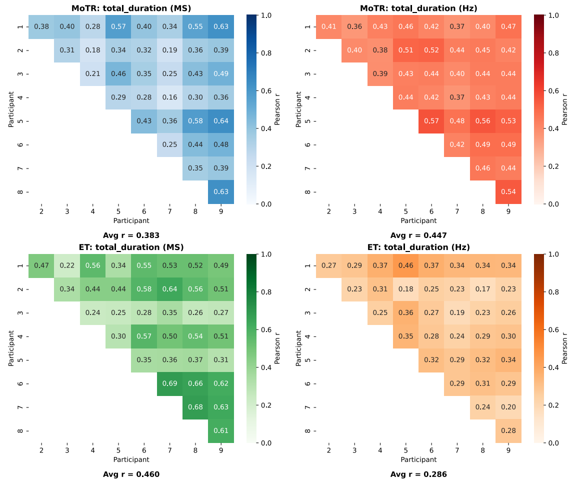


Figure 5: Intra-method Consistency: Inter-participant correlation matrices showing the behavioral upper bound for both modalities. The correlation has been measured between shifted subgroups (Subjects 1-8 vs. Subjects 2-9) to avoid self-correlation.

correlation of $r \approx 0.46$. Since the correlation between two human readers within the same modality represents the natural ceiling of behavioral similarity, it is statistically unrealistic to expect a cross-modal predictive model to significantly exceed this threshold. When viewed against this benchmark, our model's performance—reaching a Spearman $\rho = 0.34$ for the ET $\rightarrow$ MoTR direction and $\rho = 0.22$ for the MoTR $\rightarrow$ ET direction demonstrates that our approach is capturing a meaningful portion of the shared cognitive signal. The predictive results, summarized in Table 2, further reveal that the Fusion Model consistently outperforms the linguistic baseline across all scenarios and normalization strategies. While the standard log-transformation stabilizes variance and improves alignment over raw millisecond data (e.g., increasing the MoTR $\rightarrow$ ET correlation from 0.17 to 0.20), the best results are observed in the velocity domain (Hz). This superiority confirms that semantic context, provided by BERT embeddings, is important for bridging the inherent gap between motor and ocular behaviors. While the baseline model relies on surface-level features like word length, the Fusion Model uses contextual information to account for the "cognitive lag" between the eye's near-instantaneous movement and the hand's more deliberate cursor control. This integration is particularly evident in the velocity domain (Hz), where the alignment reaches its peak. By projecting both signals into a unified velocity space, we enable a more effective alignment between BERT-derived semantic representations and the observed reading pace, successfully bridging the mechanical gap between hand and eye movements. These results support the use of MoTR as

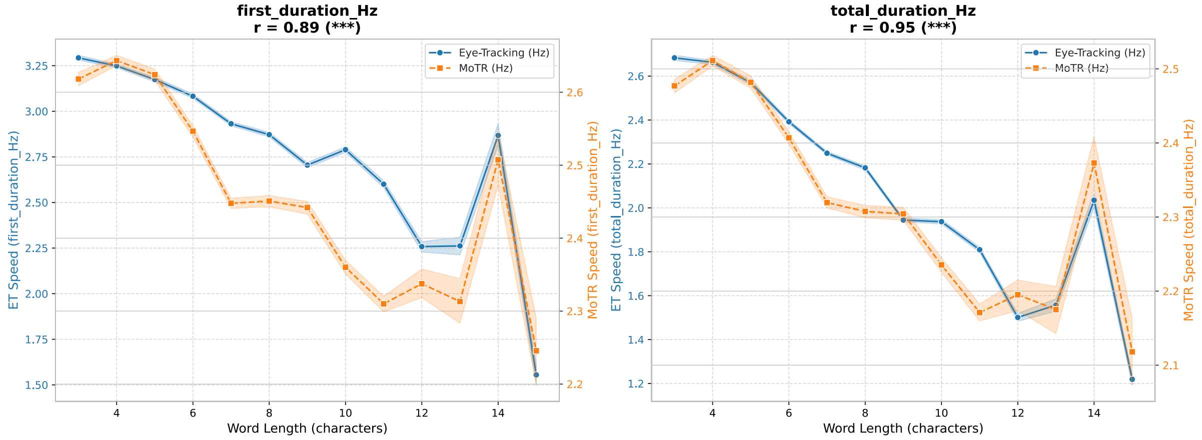


Figure 6: Psycholinguistic Validation: Word Length Effect comparison between ET and MoTR in the Velocity Domain (Hz), showing a near-perfect alignment ($r = 0.95$).

	Raw Temporal (ms)		Log-Transformed		Velocity Domain (Hz)	
Validation Scenario	**Base (ρ)**	**Fusion (ρ)**	**Base (ρ)**	**Fusion (ρ)**	**Base (ρ)**	**Fusion (ρ)**
MoTR → MoTR (Internal)	0.54	0.53	0.57	**0.58**	0.56	**0.58**
MoTR → ET (Cross-Modal)	0.16	0.17	0.19	0.20	0.20	**0.22**
ET → ET (Internal)	0.40	0.41	0.42	**0.44**	0.41	0.43
ET → MoTR (Cross-Modal)	0.26	0.34	0.27	**0.35**	0.27	0.34

Table 2: Performance comparison (Spearman ρ) across raw temporal (ms), log-transformed, and velocity (Hz) domains. Bold values indicate the best performance for each scenario.

a viable proxy for eye-tracking, providing a scalable alternative for capturing cognitive signals in NLP research.

5. Conclusion

In this study, our findings suggest that Mouse-Tracking (MoTR), when augmented with specific technical and methodological enhancements, shows potential as a scalable alternative to Eye-Tracking (ET) in cognitive NLP tasks. The implementation of the Row-Level Gatekeeper system and the vertical axis constraint were designed to isolate cognitive signals from motor-induced noise. Participants enjoyed these enhancements, who reported an intuitive and natural reading experience. These architectural modifications ensured that the recorded cursor trajectories reflect the reader's mental state rather than accidental hand instability or mechanical artifacts from row transitions. Our findings highlight that Hertz transformation can attenuate the skewness of the data and improve the linear relationship between reading times. By projecting raw durations into a unified velocity domain, we achieved an alignment regarding the Word Length Effect ($r = 0.95$), providing additional proof that MoTR captures similar underlying linguistic processing speeds as professional-grade eye-tracking hardware. Furthermore, the integration of semantic context through the BERT-enhanced Fusion Model was intended to capture additional linguistic nuances and incorporate contextual information that surface-level features alone may not fully represent. The proximity of our model's performance ($\rho = 0.34$) to the empirical ceiling of inter-participant consistency ($r \approx 0.46$) suggests that the predictive results are nearing the upper bound defined by inherent human behavioral variance. Ultimately, this research adds towards to Mouse Tracking framework for democratizing access to high-resolution cognitive signals.

By removing the financial and logistical barriers associated with traditional eye-tracking, our proposed changes to the MoTR paradigm can further aid the collection of large-scale, crowdsourced cognitive datasets. Future work will focus on expanding this validation to cross-linguistic settings and investigating the utility of these signals in low-resource language modeling.

6. Limitations

Several limitations must be acknowledged. First, the participant pool consisted of 9 subjects per modality; a larger and more diverse demographic sample would further strengthen the generalizability of these findings. Second, mechanical hand inertia remains a significant factor. The hand is fundamentally slower than the eye, as evidenced by higher fixation rates (73.6%) and lower regression rates (7.9%) in MoTR compared to ET. Manual regressions are mechanically unnatural and require more physical effort, while the necessity of synchronizing hand and eye movement slows the reading pace, potentially facilitating better initial comprehension and reducing the need for backward movements.

Furthermore, it should be noted that this study did not include a direct performance comparison between the original, unmodified MoTR framework and our enhanced version. Consequently, while our results are promising, the specific contribution of each individual refinement—such as the vertical axis constraint or the gatekeeper system—remains to be empirically quantified in future ablation studies. Finally, this study focused exclusively on Romanian texts from the MultiplEYE corpus. Expanding this methodology to different writing systems, such as right-to-left (RTL) languages, would necessitate structural adaptations to the Row-Level Gatekeeper system to accommodate reversed reading directions and navigation patterns.

7. Ethical Considerations

Data collection followed the Declaration of Helsinki and GDPR standards. Participants provided written informed consent and were briefed on their right to withdraw. Raw data were pseudonymized at the point of collection, and no personally identifiable information was utilized during the modeling. Stimuli consisted of standardized MultiplEYE texts, appropriate for research and free from harmful material.

8. Acknowledgements

This work was supported by the European Cooperation in Science and Technology under COST Action CA21131 (MultiplEYE), by the project InstRead: Research Instruments for the Text Complexity, Simplification, and Readability Assessment CNCS - UEFISCDI project number PN-IV-P2-2.1-TE-2023-2007.

9. Bibliographical References

Stuart K. Card, Thomas P. Moran, and Allen Newell. 1983. The psychology of human-computer interaction.

Karin de Langis, William Walker, Khanh Chi Le, and Dongyeop Kang. 2025. Tracing how annotators think: Augmenting preference judgments with reading processes.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. 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 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Alba Haveriku, Sara Bedulla, Nelda Kote, and Elinda Kajo Meçe. 2025. *Understanding Reading Patterns of Albanian Native Readers Through Mouse Tracking Analysis*, pages 433–443. Springer, Cham.

Nora Hollenstein, Marie-Luise Müller, Deborah N. Jakobi, Cui Ding, Maja Stegenwallner-Schütz, Ana Matić, Eva Pavlinušić Vilus, Ramunė Kasperė, Anna Bondar, Maroš Filip, Stefan Frank, Jana Hofmann, Thyra Krosness, Kaidi Lõo, Johanne Nedergaard, Chiara Tschirner, and Lena A. Jäger. 2026. *MultiplEYE Data Collection Guidelines*.

Jeff Huang, Ryen White, and Georg Buscher. 2012. User see, user point: gaze and cursor alignment in web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, page 1341–1350, New York, NY, USA. Association for Computing Machinery.

Deborah N. Jakobi, Andreas Säuberli, Jana M. Hofmann, Carlson Büth, Anastassia Shaitarova, Daniel Krakowczyk, and Lena A. Jäger. 2026. The multipleye preprocessing pipeline. `https://github.com/MultiplEYE-COST/multipleye-preprocessing`. GitHub repository.

Marcel A. Just and Patricia A. Carpenter. 1980. A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4):329–354.

Ramunė Kasperė, Anna Bondar, Sergiu Nisioi, Maja Stegenwallner-Schütz, Hanne B. Søndergaard Knudsen, Ana Matić, Eva Pavlinušić Vilus, Dorota Klimek-Jankowska, Chiara Tschirner, Not Battesta Soliva, Deborah N. Jakobi, Cui Ding, Dima Abu Romi, Cengiz Acarturk, Matilda Agdler, Anton Marius Alexandru, Mohd Faizan Ansari, Annalisa Arcidiacono, Elizabete Ausma Velta Barisa, Ana Bautista, Lisa Beinborn, Yevgeni Berzak, Nedeljka Bjelanović, Anna Isabelle Bothmann, Jan Brasser, Caterina Cacioli, Anila Çepani, Ilze Ceple, Adelina Çerpja, Dalí Chirino, Jan Chromý, Alessandro Corona Mendozza, Iria de Dios-Flores, Nazik Dinçtopal Deniz, Ana Došen, Kristian Elersic, Inmaculada Fajardo, Zigmunds Freibergs, Angelina Ganebnaya, Shan Gao, Jéssica Gomes, Annjo Klungervik Greenall, Alba Haveriku, Miao He, Anamaria Hodivoianu, Yu-Yin Hsu, Amanda Isaksen, Andreia Janeiro, Kristine Jensen de López, Aleksandar Jevremovic, Vojislav Jovanović, Hanna Kędzierska, Nik Kharlamov, Sara Košutar, Nelda Kote, Vanja Kovic, Izabela Krejtz, Thyra Krosness, Oleksandra Kuvshynova, Eilam Lavy, Ella Lion, Marta Łockiewicz, Kaidi Lõo, Paula Luegi, Mircea Mihai Marin, Clara Martin, Svitlana Matvieieva, Diane C. Mézière, Xavier Mínguez-López, Valeriia Modina, Jurgita Motiejunienė, Marie-Luise Müller, Tolgonai Nasipbek kyzy, Jamal Abdul Nasir, Johanne S. K. Nedergaard, Ayşegül Özkan, Patrizia Paggio, Marijan Palmović, Maria Christina Panagiotopoulou, Alberto Parola, Helena Pérez, Klaudia Petersen, Anja Podlesek, Eva Pospíšilová, Marta Prauliņa, Mikuláš Preininger, Loredana Pungă, Diego Rossini, Špela Rot, Habib Sani Yahaya, Irina A. Sekerina, Anne Gabija Skadină, Jordi Solé-Casals, Lonneke van der Plas, Saara M. Varjopuro, Spyridoula Varlokosta, João Veríssimo, Oskari Juhapekka Virtanen, Nemanja Vračar, Mila Vulchanova, Ahmad Mustapha Wali, Peizheng Wu, Nilgün Yücel, Stefan Frank, Nora Hollenstein, and Lena A. Jäger. 2026. The multipleye text corpus: Towards a diverse and ever-expanding multilingual text corpus. In *Proceedings of the 2026 International Conference on Language Resources and Evaluation (LREC 2026)*, Rabat, Morocco. European Language Resources Association and International Committee on Computational Linguistics.

Reinhold Kliegl, Ellen Grabner, Martin Rolfs, and Ralf Engbert. 2004a. Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, 16(1-2):262–284.

Reinhold Kliegl, Ellen Grabner, Martin Rolfs, and Ralf Engbert. 2004b. Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology - EUR J COGN PSYCHOL*, 16:262–284.

Kozea. 2025. Pyphen: A pure python module to hyphenate text.

Sergiu Nisioi, Anna Bondar, Ramuné Kasperé, and Maja Stegenwallner-Schütz. 2026. The multipleye text corpus data and materials.

Metehan Oğuz, Cui Ding, Ethan Gotlieb Wilcox, and Zuzanna Fuchs. 2025. Using motr to probe agreement processing in russian. *Open Mind*, 9:1682–1710.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830.

Cristina Maria Popescu and Sergiu Nisioi. 2025. Exploring mouse tracking for reading on Romanian data. In *Proceedings of the First International Workshop on Gaze Data and Natural Language Processing*, pages 44–51, Varna, Bulgaria. INCOMA Ltd., Shoumen, BULGARIA.

Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124 3:372–422.

Keith Rayner, Alexander Pollatsek, and Erik Reichle. 2003. Eye movements in reading: Models and data. *Behavioral and Brain Sciences*, 26.

Robyn Speer and Joshua Chin. 2021. rspeer/wordfreq: v3.2.

Sotaro Takeshita, Yurina Takeshita, Daniel Ruffinelli, and Simone Paolo Ponzetto. 2025. Randomly removing 50% of dimensions in text embeddings has minimal impact on retrieval and classification tasks. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 27705–27726, Suzhou, China. Association for Computational Linguistics.

Ethan Gotlieb Wilcox, Cui Ding, Mrinmaya Sachan, and Lena Ann Jäger. 2024. Mouse tracking for reading (motr): A new naturalistic incremental processing measurement tool. *Journal of Memory and Language*, 138:104534.

10. Language Resource References

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. Multipleye: Enabling multilingual eye-tracking data collection for human and machine language processing research.

Alan Kennedy, Robin Hill, and Joël Pynte. 2003. The dundee corpus. In *Proceedings of the 12th European Conference on Eye Movements (ECEM)*.

Steven G. Luke and Kiel Christianson. 2018. The Provo Corpus: A large eye-tracking corpus with predictability norms. *Behavior Research Methods*, 50(2):826–833. Available at https://osf.io/sjefs.