What determines where readers fixate next? Leveraging NLP to investigate human cognition

Adrielli Tina Lopes Rego
Department of Education Sciences
Vrije Universiteit Amsterdam
a.t.lopesrego@vu.nl

Joshua Snell
Department of Experimental
and Applied Psychology
Vrije Universiteit Amsterdam
j.j.snell@vu.nl

Martijn Meeter
Department of Education Sciences
Vrije Universiteit Amsterdam
m.meeter@vu.nl

Abstract

During reading, readers perform rapid forward and backward eye movements through text, called saccades. How these saccades are targeted in the text is not yet fully known, particularly regarding the role of higher-order linguistic processes in guiding eye-movement behaviour in naturalistic reading. Current models of eye movement simulation in reading either limit the role of high-order linguistic information or lack explainability and cognitive plausibility. In this study, we investigate the influence of linguistic information on saccade targeting, i.e. determining where to move our eyes next, by predicting which word is fixated next based on a limited processing window that resembles the amount of information humans readers can presumably process in parallel within the visual field at each fixation. Our preliminary results suggest that, while word length and frequency are important factors for determining the target of forward saccades, the contextualized meaning of the previous sequence, as well as whether the context word had been fixated before and the distance of the previous saccade, are important factors for predicting backward saccades.

1 Introduction

The eye movements of readers can reveal aspects of the cognitive mechanism that underlies language processing during reading. Decades of research have explored the explanatory power of eye movements to better understand which factors play a role in text comprehension (Rayner, 1998; Rayner et al., 2006). One well-established phenomenon in reading is the idea that lexical information, such as word length, frequency, and surprisal, influences the durations and locations of fixations in text (Kliegl et al., 2004). However, the influence of higher-level language processing on saccade targeting is less well known (Warren et al., 2011; Vasishth et al., 2013). Cognitive models of eye movements in reading vary in how saccade programming is simulated, and most models leave postlexical information implicit (e.g. Reichle et al., 2009). Furthermore, current machine learning approaches for predicting fixation location in reading are limited in shedding light on human language processing. They do little to explicate what drives saccade decisions and have few parallels to psycholinguistic theories and to behavioural evidence about the human cognitive systems engaged in reading.

Here we investigate to what extent we can successfully leverage deep learning methods to investigate a fundamental question about human language processing: what determines saccade programming during reading? We approach the prediction of the next fixation location as a classification problem at the word level, spanning a window of words that approximates the parallel processing of words in the human visual field (n-3 to n+3) (Snell and Grainger, 2019). In addition, we tailor the input of the model according to what information is likely to be available to the reader at each fixation. To represent the low-level linguistic information available for all words in the processing window, word length, frequency, and surprisal (i.e. negative log-probability) are employed. To represent higher-level linguistic information available on each previous word and the currently fixated word in the input sequence, we employ contextualized word embeddings from GPT-2, a unidirectional large language model (Rad-

Proceedings of First International Workshop on Gaze Data and NLP associated with RANLP 2025, pages 1–6, Varna, Bulgaria, Sep 12, 2025.
https://doi.org/10.26615/978-954-452-104-2-001

ford et al., 2019). Finally, previous fixation information is included to capture some of the dynamics of the sequential nature of eye movements. In sum, we attempt to combine the mapping power of neural networks with a more cognitively plausible set-up to understand what determines the next fixation target in human reading.

2 Related Work

The task of predicting fixation locations in reading has been mainly addressed with one of the two modeling strategies: theory-driven or data-driven. Theory-driven models are cognitive models that simulate eye movements in reading by computationally implementing psycholinguistic theories of reading with the goal of revealing the cognitive mechanisms involved in reading. The next fixation location is explicitly determined, i.e., it is clear how the model arrives at each saccadic decision. However, they are hardly ever evaluated on unseen texts and readers, and are limited in explaining the role of high-order linguistic information in saccade targeting. In E-Z reader (Reichle et al., 2009), for example, saccade targeting is limited to the range of word n-1 to word n+2, and is mainly determined by word length, frequency and predictability. Regressions occur randomly with a certain probability set by the modeler. Perhaps a more elegant mechanism is proposed by SWIFT (Engbert et al., 2005), in which the probability of each word in the model's four-word processing window to be the next saccade target is proportional to its relative word activation in an attention gradient. However, SWIFT is limited in that higher-level language processing is not accounted for. SEAM (Rabe et al., 2024) partially addresses this limitation by having sentence-level dependencies indirectly affect word activations, but this effect only occurs between verbs and subjects.

In contrast, data-driven models of eye movement simulations solely focus on accurately predicting eye movements by harnessing advanced machine learning methods while using previous (and future) fixations and/or linguistic information as input. These models rely on the predictive power of machine learning methods to achieve accurate prediction of fixations on a variety of texts, with different reading goals and reader profiles. They do this without guidance of theories of reading and little to no parallel with human cognition with respect to the model input and/or architecture. The first successful data-driven model (Nilsson and Nivre, 2009) employed logistic regression with manually engineered features extracted from the text stimuli and the previous eye movements of readers to predict the next saccade target location within a five-word window around the currently fixated word; feature importance was not reported. Wang et al. (Wang et al., 2019) combined CNN, LSTM and CRFs to predict next fixation location, based on word length, part-of-speech, and bag-of-words representations, but no regressions nor refixations were produced by the model. Dweng et al. (Deng et al., 2023) proposed Eyettention, which combines the fixation sequence (represented by non-contextualized BERT embeddings, fixation duration and landing position) and the word sequence (represented by contextualized BERT embeddings and word length) using two (bi-)LSTMs and a cross-attention layer. This model was surpassed in performance by ScanDL (Bolliger et al., 2023), a sequence-to-sequence diffusion model that generates synthetic scanpaths by also combining the fixation sequence and the word sequence (both represented by BERT embeddings). While data-driven models have been so far more successful than theory-driven ones in accurately predicting fixation locations, they still lack explanatory power and cognitive plausibility to be useful models to investigate human cognition in reading: Much of the information driving prediction is left implicit (e.g. predicting upcoming fixations based on previous fixations does not explain what underlies saccade targeting), and most information used in the input is not plausibly available to a human reader at each fixation step (e.g. future fixations, and many/all upcoming words).

3 Method

We formulated saccade targeting in reading as a classification problem, where the model has to decide which word to fixate next given a set of candidate words in the input sequence. The classifier was a shallow fully connected neural network, with one hidden layer of 128 nodes, ReLu activation and a drop-out layer [1]. The input sequence consisted of a window of seven words, i.e. the fixated word plus three words before and three words after, to approximate the limited amount of information a human reader can likely take in the visual field at each

[1] Pilot studies were performed with CNNs and LSTMs to preserve the word structure in the input, but, surprisingly, the fully-connected neural network yielded the best results.

fixation. To represent lexical information on each word in the input sequence, we used word length, frequency, and surprisal, which are assumed to be available to the reader to some degree through either past word recognition or current parafoveal processing. To represent higher-order language information, we used the contextualized word embedding of the fixated word from GTP-2, which is assumed to encode the meaning constructed from the text up to the fixated word. Finally, to capture some of the dynamics inherent to the sequential nature of eye movements, we added information on whether each word in the input sequence has been fixated before, the previous fixation duration and the previous saccade length. All features were z-normalized, except for the word embedding and the binary feature encoding whether or not the word had been fixated before.

We trained the classifier on the L1-English part of the MECO corpus (Siegelman et al., 2022), using 5-fold cross-validation with a 80/20 split based on text ids. The material consists of the first 10 texts of the corpus, structured similarly to Wikipedia-style encyclopaedic entries, covering a diverse range of topics. Each text had approximately 200 words and 10 sentences. All participants (n = 46) were native speakers of English and university students. They were instructed to read the texts silently and answer (four) comprehension questions after each text. We used the fixation dataset available in the "fixation report" folder, in the path "release 1.0/version 1.2/primary data/eye tracking data/fixation report", in the OSF directory of the MECO corpus. We only included the fixations on words that had three words to the left and three words to the right, resulting in 66,383 fixations in total. Around 34% of these fixations were to word n+1, followed by 25% to word n+2, 18% to word n, 10% to word n-1, 7% to word n+3, 3% to word n-2, and 1% to word n-3.

Model evaluation consisted of measuring the F1 scores ($2 * (precision * recall)/precision + recall$) for each word position in the input sequence (seven words, including currently fixated word) and the macro-averaged F1 score across word positions. We compare the model performance with three baselines: OB1-reader (Snell et al., 2018), a cognitive model of eye movement control in reading, in which saccade targeting is determined by word recognition and visual attention; the same model trained on random input vectors; and a majority baseline, which always predicts the majority class (word n+1). To evaluate OB1-reader, we ran 10 simulations on the corpus texts and, for each simulation, we selected the fixations that overlapped between the model simulation and the corpus, and checked whether the next fixation target was the same. We then reported the resulting F1 score averaged over simulations.

4 Results

As can be seen in Table 1, our model outperforms the baselines, including the OB1-reader model, although the difference in macro-averages is small. The easiest saccade to predict is to word n+1, which is also the most frequent. Backward saccades are the most difficult to predict, and the farther away from the current fixation, the lower the performance in predicting saccade targeting. OB1-reader performs remarkably well compared to our model, especially at one-word regressions and refixations. Overall, our model improves saccade targeting prediction compared to the baselines, but still performs below chance for word skips and refixations, and poorly for backward saccade targeting.

To determine feature importance, we replaced one feature at a time by its average over the dataset and retrained the model with the ablated feature. Table 2 shows the model performance when removing each feature. When word length is ablated, the model performance especially drops in predicting word skips (word positions 2 and 3). Word frequency also seemed to affect two word-skipping (word position 3). Whether or not the context word has been fixated before is predictive of backward saccades (word positions -1, -2, and -3), as well as refixations and two-word skipping (word positions 0 and 3). Embeddings seems to be informative for backward saccades, but not for word skipping (word positions 2 and 3). Finally, while the previous fixation duration does not seem to be an informative feature in general, the previous saccade distance supports to some extent the prediction of backward saccades (word positions -3 and -2) as well as two-word skipping (word position 3). In sum, word length and frequency were important features for the prediction of forward saccades, while the fixated word's contextualized embedding, whether the word has been fixated before and the previous saccade length were mainly informative of backward sacacades.

	-3	-2	-1	0	1	2	3	macro-avg
Classifier	.002 ±.005	.001 ±.002	.05 ±.017	.24 ±.018	.56 ±.024	.46 ±.018	.12 ±.038	.20 ±.006
OB1-reader	0	0	.11 ± .002	.30 ± .01	.31 ± .01	.32 ± .004	.15 ± .006	.17 ± .002*
Random	0	0	.01 ±.008	.11 ±.006	.44 ±.016	.35 ±.011	.004 ±.006	.11 ±.004 *
Majority	0	0	0	0	.51 ±.017	0	0	.07 ±.002 *

Table 1: F1 scores averaged over cross-validation splits for each true word position target, as well as averaged over positions. * means that the score was significantly different from the classifier model.

	-3	-2	-1	0	1	2	3	macro-avg
Classifier	.002 ±.005	.001 ±.002	.05 ±.017	.24 ±.018	.56 ±.024	.46 ±.018	.12 ±.038	.20 ±.006
w/o word length	.002 ±.005	.001 ±.002	.04 ±.01	.23 ±.03	.54 ±.02	.42 ±.02	.07 ±.02	.19 ±.008 *
w/o word frequency	0	.003 ±.008	.05 ±.02	.25 ±.008	.55 ±.02	.45 ±.01	.07 ±.02	.19 ±.003 *
w/o word surprisal	0	.002 ±.003	.04 ± .01	.24 ± .01	.56 ± .02	.46 ± .01	.10 ± .03	.20 ± .003
w/o has-been-fixated	0	0	.01 ± .01	.21 ± .02	.55 ± .02	.45 ± .01	.06 ± .06	.18 ± .01*
w/o embedding	0	0	.02 ± .01	.24 ± .02	.59 ± .02	.50 ± .01	.17 ± .06	.22 ± .01
w/o previous fixation duration	.004 ± .006	.001 ± .002	.06 ± .02	.25 ± .02	.56 ± .03	.46 ± .01	.10 ± .03	.20 ± .005
w/o previous saccade distance	0	0	.04 ± .01	.26 ± .02	.56 ± .02	.46 ± .01	.09 ± .04	.20 ± .004

Table 2: Feature ablation. This table displays the F1 scores averaged over cross-validation splits for each true word position target, as well as averaged over positions, for each model version in which one feature is ablated. * means that the score was significantly different from the full classifier model.

5 Discussion

In this study, we attempted to investigate the cognitive processes underlying saccade targeting in reading using deep learning. We sought to leverage machine learning while using input whose information content may resemble more closely what is plausibly available to human readers during saccade planning. Importantly, we attempted to fill a gap in understanding the role of high-order language information by investigating to what extent the text meaning, as represented by contextualized embeddings, supports where readers tend to fixate next, beyond lower-level lexical information. Our preliminary results indicated that forward saccades tend to be more driven by automatic, oculomotor cues, as well as low-level linguistic cues, such as word length and frequency, whereas backward saccades are more heterogeneous, with the semantics of the previous context playing a role, but also factors possibly related to oculomotor error, such as skipping a word due to overshooting, as suggested by the features "has-been-fixated" and "previous saccade amplitude". Our results are in line with well-established findings in the literature that support the major role of lower-order linguistic features in forward saccades (Rayner, 1998; Kliegl et al., 2004; Engbert et al., 2005) and the heterogeneous nature of backward saccades (Von Der Malsburg and Vasishth, 2011; Inhoff et al., 2019; Wilcox et al., 2024). Furthermore, refixations seemed to be driven by word length and whether the word had been fixated before, but, surprisingly, not by factors pertaining word meaning, such as frequency, surprisal and its contextualized embedding, suggesting that, at least in this dataset, most refixations were a result of oculomotor and low-level linguistic cues. Ultimately, our goal is to model the complex interplay between the oculomotor system and language processing that drives saccade targeting in reading. Combining the predictive power of machine learning methods with more cognitively plausible and interpretable modeling may shed light on the mechanisms behind this process.

6 Limitations and Future Work

The model proposed here fails to predict backward saccades with an acceptable level of accuracy. Previous correlational research has suggested PMI scores to be predictors of regression targeting in reading (Wilcox et al., 2024). A follow-up study may explore the potential of such measure in informing the prediction of backward saccade targeting in reading. In addition, the dynamics of eye movements is not fully explored in our model, as only information on the previous fixation is used. It is possible that information on more previous fixations is needed to capture the complex relation between the sequence of eye movements and the sequence of language input.

Finally, we assumed that word length, frequency and surprisal of the words in the upcoming context are fully available to the reader, which is a simplification. As a follow-up, this information will be modulated by OB1-reader's visual attention gradient, based on eccentricity and visual acuity. That is, the closer the words are to the fixation the more accurate the linguistic information available. Future work may investigate whether our neural network model can be merged with a cognitive model, such as OB1-reader, to use word activations generated by the cognitive model as a proxy of low-order visual and linguistic information, together with high-order linguistic information represented by contextualized embeddings, to predict saccade targeting. More of the dynamics of the relation between eye movements and language input might be indirectly captured by the cognitive model's word activations.

7 Acknowledgments

This research was funded by the "Nederlandse Organisatie voor Wetenschappelijk Onderzoek" (NWO) 672 Open Competition-SSH (Social Sciences and Humanities), grant number 406.21.GO.019 to MM. A big thanks to the MultiplEYE Cost Action (CA21131) for funding travel expenses of the first author to attend the Workshop.

References

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. *arXiv preprint arXiv:2310.15587*.

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

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.

Albrecht W Inhoff, Andrew Kim, and Ralph Radach. 2019. Regressions during reading. *Vision*, 3(3):35.

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.

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.

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.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

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

Keith Rayner, Kathryn H Chace, Timothy J Slattery, and Jane Ashby. 2006. Eye movements as reflections of comprehension processes in reading. *Scientific studies of reading*, 10(3):241–255.

Erik D Reichle, Tessa Warren, and Kerry McConnell. 2009. Using ez reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic bulletin & review*, 16:1–21.

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

Joshua Snell and Jonathan Grainger. 2019. Readers are parallel processors. *Trends in Cognitive Sciences*, 23(7):537–546.

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.

Shravan Vasishth, Titus von der Malsburg, and Felix Engelmann. 2013. What eye movements can tell us about sentence comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(2):125–134.

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.

Xiaoming Wang, Xinbo Zhao, and Jinchang Ren. 2019. A new type of eye movement model based on recurrent neural networks for simulating the gaze behavior of human reading. *Complexity*, 2019(1):8641074.

Tessa Warren, Erik D Reichle, and Nikole D Patson. 2011. Lexical and post-lexical complexity effects on eye movements in reading. *Journal of Eye Movement Research*, 4(1):1.

Ethan Gotlieb Wilcox, Tiago Pimentel, Clara Meister, and Ryan Cotterell. 2024. An information-theoretic analysis of targeted regressions during reading. *Cognition*, 249:105765.