Inferring sentence comprehension from eye movements in reading

Yunyan Duan (Northwestern U.: yduan@u.northwestern.edu), Yevgeni Berzak (MIT),
Klinton Bicknell (Northwestern U.; Duolingo), Roger Levy (MIT)

A large body of work in sentence processing suggests that eye movements in reading reflect how readers process the text and engage with it mentally ([1,2], among many others). This conclusion is largely supported by evidence from the effect of systematic targeted manipulation of linguistic characteristics of the text on reading time measures across multiple subjects and items. However, little is currently known about how much information can be obtained about the cognitive state of a *specific reader* from their eye movements over an *individual sentence*. Considering that reading is a multifaceted and text contingent process which requires dynamic integration of information from multiple sources, the link between eye movements and readers' cognitive state may be stronger than currently known.

In this study we focus on the connection between eye movements and reading comprehension. Prior work has demonstrated a potential link between the two by using supervised classification methods to predict comprehension from eye movements with above chance accuracy [3,4]. Here we adopt a similar classification based methodology, but differently from that work we predict comprehension on the granularity level of an individual sentence rather than a page length text sample, and crucially, we propose a challenging evaluation setup in which both the test reader and the test sentence are unseen in training.

To this end, we introduce a new dataset comprising eye movement records of 54 native English speakers reading 58 sentences. The sentences contain a target word (Target) which is crucial for comprehending the sentence, and has a high-frequency neighbor (HFN) [5] which is also contextually more plausible. Sentence comprehension is tested using a multiple choice question with four answers: one for correct comprehension of the Target (chosen in 74.8% of the trials), one consistent with the HFN (20.2%), and two unrelated answers (5% combined).

Our classification task consists of predicting whether a reader answered the comprehension question correctly based on their eye movements while reading the sentence. We experiment with two classifiers. The first is a Logistic Regression (LR) model which uses word properties and eye movement interaction features for the Target and HFN as well as global text and reading time statistics commonly used for this task. We contrast this model with a state of the art Convolutional Neural Network (CNN) [6] classifier which encodes text and reading times for each individual word in the sentence without information about the Target and its HFN. We test the informativity of the eye movement signal for our prediction task by comparing each model to baseline models which have access only to text and reading speed information. We perform Leave One Out (LOO) cross validation evaluation, in which the training set of each split does not contain trials from the test subject nor the test sentence, using 60% of the subjects and 60% of the items (1,060 trials) of our dataset.

Our results are summarized in Table 2 and Table 3. First, in the LR model, both Target specific and global eye movement features improve performance over the baselines, with the combination of the two yielding the strongest results. Similarly, eye movement features improve the performance of the CNN. Second, the CNN performance is substantially better than LR, highlighting the importance of an architecture that encodes the entire sentence and effectively captures feature interactions. Overall, these results suggest that eye movements contain substantial signal about reading comprehension at the individual sentence and subject level.

Table 1: Example Item. The Target is marked in bold. The Target's HFN is "minute".

<u>Sentence</u>: I'm really glad that the last **minuet** went by so quickly and I could finally go home.
<u>Question</u>: What was I probably watching to look for cues that I could leave?

A. an orchestra (correct) B. a clock (incorrect; HFN consistent)

C. the president (incorrect; unrelated) D. a weather report (incorrect; unrelated)

Figure 1. Structure and features of Logistic Regression and Convolutional Neural Network.

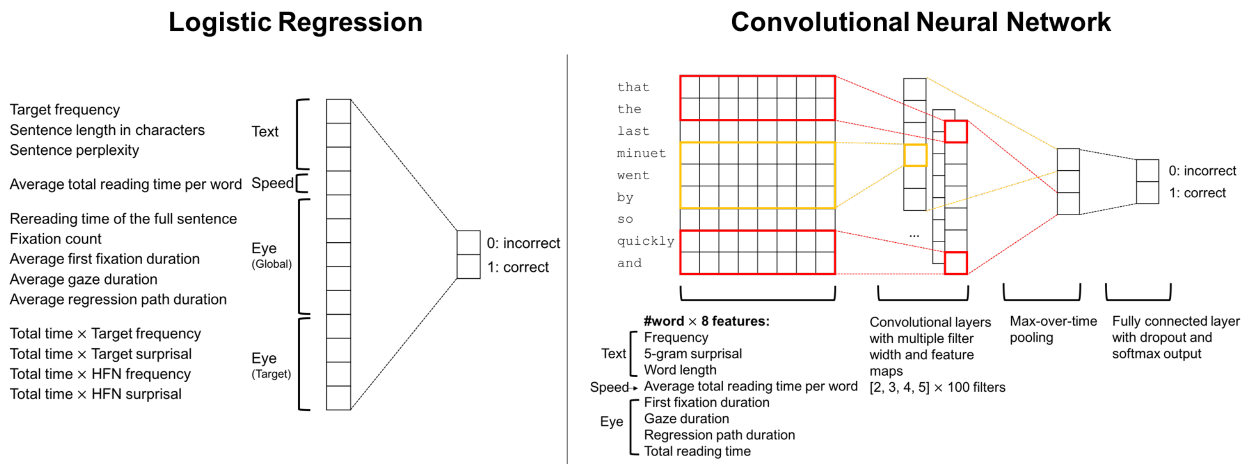

Table 2. ROC-AUC of Logistic Regression and Convolutional Neural Network.

	Logistic Regression		Convolutional Neural Network
Text	0.56		0.75
Text + Speed	0.55		0.78
Text + Speed + Eye	Target	0.57	
	Global	0.61	
	Both	**0.63**	**0.80**

Table 3. Comparison of different models' likelihood using Wilcoxon signed-rank test. Bonferroni-corrected *p*-values are reported.

	CNN vs. LR		
	Text	Text + Speed	Text + Speed + Eye
p	< 0.001 ***	< 0.001 ***	< 0.001 ***

	Text+ Speed vs. Text		Text + Speed + Eye vs. Text		Text + Speed + Eye vs. Text + Speed	
	CNN	LR	CNN	LR	CNN	LR
p	n.s.	n.s.	< 0.001 ***	0.007 **	< 0.001 ***	0.01*

References
[1] Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, *87*(4), 329. **[2]** Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372. **[3]** Copeland, L., & Gedeon, T. (2013, December). Measuring reading comprehension using eye movements. In *Cognitive Infocommunications (CogInfoCom), 2013 IEEE 4th International Conference on* (pp. 791-796). IEEE. **[4]** Martínez-Gómez, P., & Aizawa, A. (2014). Recognition of understanding level and language skill using measurements of reading behavior. In *Proceedings of the 19th international conference on Intelligent User Interfaces* (pp. 95-104). ACM. **[5]** Slattery, T. J. (2009). Word misperception, the neighbor frequency effect, and the role of sentence context: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1969. **[6]** Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.