

# Los Angeles Reading Corpus of Individual Differences: Pilot distribution and analysis

Jesse A. Harris (jharris@humnet.ucla.edu)

Department of Linguistics, 3125 Campbell Hall  
Los Angeles, CA 90095 USA

Stephanie Rich (skrich@ucsc.edu)

Department of Linguistics, 1156 High Street  
Santa Cruz, CA 95064 USA

## Abstract

We introduce the LARC-ID, a pilot corpus of eye-movements obtained from subjects reading texts from a range of genres. Materials were presented in multiple paragraphs on the screen to more closely match naturalistic reading environments. Readers were encouraged to read for comprehension and enjoyment, engaging in various kinds of comprehension questions, including an open-ended reflection at the end of each text. Subjects also participated in a battery of individual difference measures, including those known to predict reading behavior in controlled experimental contexts, e.g., Rapid Automatized Naming (RAN; Denckla & Rudel, 1976), the author recognition (ART; Stanovich & West, 1989), and reading span (RSpan; Daneman & Carpenter, 1980) tasks. In addition to describing the central properties of the text and relationships between tasks in the battery, we present a sample analysis exploring how intrinsic lexical characteristics (length, frequency, and morphological complexity) interact with selected individual difference measures. The analysis provides the very first glimpse into what we hope will become a useful resource for reading researchers and educators. The entire corpus is freely available for unrestricted use.

**Keywords:** reading; corpus; individual differences

## Introduction

Much research on reading records eye movement measures in a necessarily artificial environment, presenting highly controlled, and often challenging, sentences with characteristics of specific interest to the researchers' hypotheses. While certainly a valid method for exploring reading and language comprehension, these materials tend to make little sense in isolation. Such a method is well-suited for measuring the subtle effects of linguistic variables on the comprehension process, and assumes that the kind of reading that subjects engage in during the experiment represents, on the whole, the reading process in general. Yet, early work on eye movements has shown that subjects adapt to the presumed goal of the experiment (Yarbus, 1967), to the point that even the presence or absence of section titles can radically alter the reading profile of an individual (Bransford & Johnson, 1972). Readers adopt task-driven strategies in response to differences in such factors such as attention, cognitive reserve, required engagement, and task instructions (e.g., Wotschack, 2009, for review).

Rayner & Raney (1996) observed, for example, that the *frequency effect*, in which words with higher frequency elicited shorter fixation times on early eye-movement measures (e.g., Rayner & Duffy, 1986; Inhoff & Rayner, 1986, among others), was eliminated when subjects were instructed to search

the text for specific word forms. Readers also appear to engage in more careful reading when presented with text perceived as more difficult, which may influence low-level decisions regarding the landing position of the eye within the word (O'Regan, 1990). Readers may also simply differ globally with respect to how they progress through text and to what kinds of information they attend (Hyönä et al., 2002).

The first goal of this pilot project was to create a corpus that captures reading of long-form, connected text from various genres and topics as a measure of engaged, naturalistic reading. Texts that college-aged readers were likely to encounter on their own were selected by consulting a focus-group consisting of college-aged students who were not involved in the study. The experimental subjects were students attending an undergraduate seminar on reading and vision taught by the first author, and were internally motivated to learn how eye movements research was conducted. The second goal addressed the fact that readers adopt different strategies in progressing through a text: some readers progress quickly at first, and later returning the more difficult sections, while others progress more slowly (Rayner et al., 2012). These patterns are generally implicit: readers cannot accurately introspect on the rapid movements of their eyes. The study presented here reports a battery of individual difference measures in order to correlate individual reading profiles with other cognitive measures thought to influence comprehension.

There are already several existing corpora of eye-movements while reading. Some consist of texts not created by the experimenters, while others contain texts especially designed for the corpus. An example of the former is the Dundee corpus, in which readers (10 English, 10 French) were presented with newspaper articles in paragraph format totaling approximately 50,000 words (Kennedy, 2003; Kennedy & Pynte, 2005). Another corpus with multilingual readers (14 monolingual British English speakers; 19 Dutch-English bilinguals) is the GECO corpus (Cop et al., 2017), which obtained eye-movements recorded while subjects read an entire novel consisting of over 5,000 sentences. The Provo corpus (Luke & Christianson, 2018) presented 55 short paragraphs from various sources to 84 native speakers of English. Predictability norms for each word were collected in a separate cloze task, and paired with their eye-movement data. Other corpora have constructed the materials to ensure that constructions of interest to researchers are present, e.g., the

Potsdam corpus of German sentences (Kliegl et al., 2004) and the DEMONIC corpus in Dutch (Kuperman et al., 2010).

The GECCO corpus collected several measures of language proficiency: scores from LexTALE (Lexical Test for Advanced Learners of English; Lemhöfer & Broersma, 2012), a spelling test, and a lexical decision test, as well as a self-reported linguistic background questionnaire (adapted from LEAP-Q; Marian et al., 2007) to probe bilingual proficiency. Another large-scale study of 546 college-aged readers examined the relationship between reading performance and two individual difference measures (Author Recognition Task – ART, and Rapid Automatized Naming – RAN) on various isolated filler sentences presented in several experiments (Gordon et al., 2019).

However, we know of no previous corpus that pairs eye-movements on naturalistic, long-form text with measures of individual differences from both linguistic and domain-general tasks, despite the many studies implicating an association between the two in experiments with controlled materials. For example, Kuperman & Van Dyke (2011) explore the effect of individual performance on several verbal and cognitive tasks in reading in non-college bound individuals reading isolated sentences. They observed that scores on the RAN task, described below, and word identification tests were the only individual difference measures that predicted reading. These measures were found to interact with effects of word length and frequency, often observed in reading time studies.

As the present corpus offers many options for exploration, we have constrained our central analysis to the effect of individual performance on intrinsic lexical characteristics. The analysis concentrates on eye-movement measures that appear relatively early in the eye-movement record, as they have routinely shown sensitivity to lexical-level effects. The corpus and the methods are described in more detail below.

## Methodology

The corpus consists of eye movements from fifteen undergraduate students (10 female) who read texts for comprehension, paired with a battery of individual difference measures. Data was collected in two sessions, conducted a week apart. Subjects were invited to take breaks as needed. In the first session, participants read long form texts and answered simple two-alternative forced-choice comprehension questions intermittently throughout the experiment. In addition, subjects were asked to reflect on the meaning of the text with the experimenter via open discussion prompts to encourage more careful and thoughtful comprehension. For example, subjects were asked to describe the relationship between the main characters, speculate on implications of the passage, or to reflect on the author's intention. Questions can be found with the texts on the OSF page hosting the data.

Five texts were used in the corpus, classified into information seeking, fiction, and opinion genres. Three short articles from Wikipedia comprised the information seeking category, a short story (“No Kick from Champagne” by Lowry Pei)

represented the fiction genre, and a piece on internet security (“Peak Indifference” by Cory Doctorow, about Internet privacy) represented the opinion genre. Texts are distributed under a Creative Commons License ShareAlike and may be distributed freely as long as attribution is retained. Key characteristics of the texts are summarized in Table 1, along with the number of fixations (average per reader and total per text) collected for each piece.

The individual differences battery sampled from well-known, standard, and previously validated tests that have been independently hypothesized to reference qualitatively different aspects of cognition. The tasks in the battery were selected to cover a heterogeneous range of cognitive and attentional abilities, while minimizing redundancy between tests. The tasks are briefly described below.

## Methods

Texts were presented in complete paragraph form using SR Experiment Builder software. Paragraphs did not span across pages in order to preserve coherence within each screen. The text was presented in 11 point monospaced Monaco style font with 2.5 linespacing. Eye movements were recorded with a SR EyeLink 1000 Plus, and were sampled at 1000 Hz. Subjects' heads were stabilized in a tower mount with the eye tracking camera mounted above their heads. Viewing was binocular but only the right eye was recorded. Subjects were positioned approximately 55cm from a 24-inch Dell UltraSharp U2410 LCD monitor (55.88cm width × 49.27cm height) with a 1024×768 resolution. Approximately 2 characters subtended 1 degree of visual angle. A 9 point calibration was performed at the beginning of the experiment, before a new text, after a break, and as needed. In addition, a drift correct was performed before each new page of text was presented. Blinks and other artefacts were removed automatically, and cases of minor vertical drift, typically occurring towards the bottom of the screen, were adjusted by moving fixations onto the appropriate line. Horizontal position was unaffected.

After the reading portion of the first session, participants completed a Rapid Automatized Naming (RAN) task while their eye movements were recorded (Denckla & Rudel, 1976; Jones et al., 2013). Developed as a diagnostic task for dyslexic readers, RAN performance has been shown to correlate with reading ability in multiple stages of development (Mol & Bus, 2011). Scores from the RAN task include accuracy, and time to completion for eight 6x6 arrays of letters, numbers, colors, and shapes. The Eye-Voice Span (EVS), i.e., the distance between the voice and the fixation in reading aloud (Levin & Addis, 1979; Jones et al., 2013), was also recorded for each array. Measures of performance on the RAN task that include EVS (rather than accuracy or time to completion) have been shown to predict numerous eye-movement measures (Gordon & Hoedemaker, 2016). Prior to viewing each array, subjects were introduced to the items in the array and given labels for naming in a practice section.

Subjects completed the Moore & Gordon (2015) revision

Texts	Genre	Author	Sentences		Word	Length	Log-Frequency	Fixations
			Count		Tokens; Types	Mean [Range]	Mean [Range]	Average; Total
Butterfly	Info	Wikipedia	10		291; 157	6.21 [1,15]	3.45 [0.60, 6.18]	124; 1,861
Turing bio	Info	Wikipedia	13		288; 191	6.23 [1, 13]	3.42 [0.70, 6.18]	109; 1,636
Turing test	Info	Wikipedia	13		313; 162	6.27 [1, 17]	3.69 [0.60, 6.18]	119; 1,781
Champagne	Fiction	L. Pei	211		3535; 1020	5.63 [1, 13]	3.49 [0.30, 6.33]	1,172; 17,577
Peak Indifference	Opinion	C. Doctorow	43		1134; 530	6.17 [1, 13]	3.40 [0.30, 6.33]	466; 6,992

Table 1: Properties of texts (sentence count, word count, length and log frequency) and number of fixations from readers (average number of fixations per individual reader and total number of fixations recorded for each text).

of the Author Recognition Test (Stanovich & West, 1989), which measures how many names a participant recognizes as authors, e.g., *Margaret Atwood*, from a list containing non-author foils, e.g., *Frances Fincham*. Finally, subjects answered a short questionnaire about their academic and linguistic history. Variables such as age, year in college, major, gender, and native language, among others, were collected. Subjects also provided a more subjective profile of their reading habits (similar in spirit to Acheson et al., 2008), including how much they enjoyed reading, how often they read for pleasure, and when was the last book they read for pleasure. They also provided as many names of authors in multiple genres as they could within a fixed time span. These materials have been made available on an Open Science Foundation page (<https://osf.io/gnvbu/>) hosting the corpus, materials, and guided tutorial.

In the second session, subjects completed three tasks meant to capture distinct aspects of individual variation: (i) *verbal working memory* capacity, as indexed by the Reading span task (Daneman & Carpenter, 1980), (ii) *non-verbal working memory and attention*, as measured by the N-back task (successful recall of an element in a sequence like 7 . . . T . . . R . . . 8 at a particular position, e.g., 2 positions back from 8 is T), and (iii) *general problem solving capacity*, as indexed by performance on a sampling of 36 Raven’s Advanced Progressive Matrices across all difficulty levels (Raven, 1989).

## Variables used in dataset

**Eye movement variables** Two common early eye movement measures are explored here (Rayner, 1998): *first fixation duration*, the duration of the initial fixation when first encountering a word, *first pass time*, the sum of fixations on a word before moving to the right or left. We also present the *number of fixations* made on a word during first pass reading. Other reading measures are included in the dataset, but are omitted here for space limitations.

**Individual difference variables** In addition to the self-explanatory categories of *age*, *year in college*, *gender*, and *major*, a battery of individual difference measures obtained during the second testing section are included in the dataset. We discuss a sample of these variables, including the mean correct response to Raven’s progressive matrices, participant score on the Author Recognition Task, the average total time

a participant spent on the RAN task, along with the average temporal span between the voice and the fixation point in the RAN task, and the participant’s Reading span score. The measures are summarized in Table 2.

Most individual level factors showed no correlation with other factors. For example, year in college was uncorrelated with ART ( $r = 0.08$ ), which is perhaps surprising under the assumption that ART is an indirect indicator of print exposure and that students are exposed to more text in college. However, other measures were correlated. There was a numerical trend towards negative correlation between each subject’s total accuracy on the RAN task and their eye-voice span [ $r = -0.47$ ,  $t = -1.92$ ,  $p = 0.08$ ]. A positive correlation between the time to complete the RAN tasks and the eye-voice span was observed [ $r = 0.89$ ,  $t = 7.03$ ,  $p < 0.001$ ], indicating that longer eye voice spans resulted in longer overall naming times.

Moore & Gordon (2015) found a small negative correlation between ART scores and average RAN durations, as well as a positive correlation between ART scores and accuracy on item naming in RAN. In our study, there was a negative correlation between ART scores and eye-voice span on letter arrays: individuals with greater print exposure also had shorter eye-voice spans [ $r = -0.53$ ,  $t = -2.27$ ,  $p < 0.05$ ]. However, no significant correlations between an individual’s ART score and eye-voice span on other arrays (numbers, shapes, colors) were detected. As Moore & Gordon (2015) note, the relationship between RAN and ART appears to be relatively weak for college-aged readers (see also Gordon et al., 2019).

**Lexical level characteristics** Lexical level characteristics were calculated for every word included in the corpus from the English Lexicon Project (Balota et al., 2007). Variables explored here include (a) the length of the word in characters, (b) the log frequency of the word (LgSUBTLWF) as calculated in the SUBTLEX corpus from American subtitles, and (c) morphological complexity as indicated by the number of morphemes (NMorph) in a word. Other variables in the dataset include other measures of frequency, orthographic and phonological neighborhood size, number of higher frequency orthographic neighbors, part of speech, and behavioral results on lexical decision and naming tasks.

**Model fitting procedure** All models presented below are linear mixed effect regression models with fixed effects nor-

Variable	Name	Description	Range	Median	Mean
Age	Age	Age in years	[18,22]	19	19
Gender	Gen	Gender [optional]	—	F (10)	—
Year in college	YCol	Year in college	[1,4]	2	2
Raven’s progressive matrices	RM	The percent correct answers to Ravens Progressive Matrices.	[33%,89%]	78%	73%
Author recognition task	ART	The number of names the subjects recognized as authors, after misidentifications were removed from the score.	[5, 33]	15	17
Rapid Automated Naming Duration	RAN.dur	The total time in seconds to complete the RAN task, averaged across the six arrays.	[15, 24]	19	19
Eye-Voice span	EVS	The average temporal span between the location of the eye and onset of speech for the item being named.	[524, 796]	630	648
Reading span	RSpan	The number of items that a subject was able to correctly recall in order on a reading span task before making an error.	[0,5]	3	3

Table 2: Summary of sample individual difference variables included in corpus.

malized by a z-transformation. As not all models with maximal random effect structures converged, by-subjects and by-text random intercepts were used (comparable to other corpus analyses, e.g., Dirix & Duyck, 2017; Gordon et al., 2019). The  $p$ -values reported in the text were estimated using `lmerTest` (Kuznetsova et al., 2017). We explored the data with hierarchical linear modeling in three stages. First, we fit the data with a selection of intrinsic lexical features (length, frequency, and morphological complexity) known to influence early reading measures. For each of the three measures, a simple model with word length as the only fixed-effect predictor was initially fit. Increasingly complex models were computed by adding log frequency (LgSUBTLWF) and word complexity (NMorph) as interactive predictors until the best-fitting model with the lowest AIC value was obtained. Predictors were standardized in a z-transformation. To reduce excessive multicollinearity, predictors with the highest variance inflation factor (VIF) above 5 were removed in succession, and the model was re-fit. Since the interaction between word length and number of morphemes is highly correlated, this term was the first to be removed from all models. Non-significant effects were also removed. The VIF for any condition and the kappa for each reported model were both below 5. Second, a separate set of models were created with (a) each of the individual difference measures as the sole fixed-effect predictor, and (b) all of the individual difference measures in a single additive model. Third, each individual difference measure was added to the first set of models, set to interact with word frequency.

## Results

### Models of intrinsic lexical features

The best-fitting models given our procedure are shown in Table 3. Replicating the classic frequency effect, more frequent

words elicited shorter first fixation durations and first pass times. In both measures, the frequency effect was moderated by an interaction with length, in that longer words tended to increase the effect that frequency imposed on reading times.

In addition, longer words received shorter first fixation times, which might be unexpected given the association between length and reading times observed in previous literature (Inhoff & Rayner, 1986). However, longer words elicited longer first pass times, as well as more fixations during first pass reading. The pattern suggests that readers were more likely to re-fixate longer words, and that the first fixation in such cases tended to be short. In support of this interpretation, there was a moderate positive correlation between the number of fixations and first pass times,  $r = 0.19, t = 22.64, p < .001$ , but a negative correlation between the number fixations and first fixation durations,  $r = -0.18, t = -21.47, p < .001$ .

This pattern is compatible with a possible trade-off in oculomotor control strategies that is in line with current models of reading, such as E-Z Reader (Reichle et al., 1998, 2006) and SWIFT (Engbert et al., 2005). For example, in later versions of E-Z Reader, a re-fixation program is initiated with a probability determined by word length. The re-fixation program may be cancelled if the word is recognized within a temporal deadline. Low frequency words take longer to recognize, thereby eliciting more re-fixations and slowing forward progress through the text (Reichle et al., 2003).

Finally, there was a three-way interaction between Length, frequency, and word complexity for first pass times,  $t = -4.65, p < .001$ , and the number of first pass fixations,  $t = -2.28, p < .05$ . Word complexity appeared to mediate the interaction between frequency and length in first pass times. As shown by the model fits in Panel A of Figure 1, single morpheme words exhibited a smaller interaction between frequency and length compared to multi-morphemic words. In

Parameter	First fixation durations			First pass times			First pass fixations		
	Estimate	SE	t-value	Estimate	SE	t-value	Estimate	SE	t-value
(Intercept)	251.62	7.89	31.87 *	287.70	8.75	32.88 *	1.59	0.04	37.40 *
Length	-5.23	1.28	-4.07 *	13.48	1.92	7.03 *	0.18	0.01	20.89 *
LgSUBTLWF	-5.82	1.20	-4.87 *	-8.07	1.66	-4.86 *	–	–	–
NMorph	–	–	–	-5.80	1.73	-3.36 *	–	–	–
Length:LgSUBTLWF	1.83	0.74	2.48 *	4.04	1.39	2.90 *	0.02	0.01	2.24 *
Length:LgSUBTLWF:NMorph	–	–	–	-2.43	0.52	-4.65 *	-0.01	0.00	-2.28 *

Table 3: Best fitting linear mixed effect regression models with z-transformed predictors and by-subject and by-item random intercepts. Effects were assumed to be significant if  $|t| > 2$ , signified with a '\*’.

the case of complex word forms, low frequency items elicited shorter first pass times on shorter forms compared to longer forms. However, higher frequency items elicited shorter first pass times on longer forms. The number of first pass fixations followed a similar pattern. Both lexical and post-lexical explanations of the three-way interaction are possible. Complex words might have slowed forward progress during lexical processing either by increasing the probability of re-fixation or by decreasing the speed of word recognition by engaging morphological decomposition. Alternatively, complex low frequency words might require a post-lexical process of morphemic combination to generate word meanings.

### Models of individual differences

Separate models with individual difference measures as the sole predictor were computed for each of the three eye movement variables. Although no significant effects were observed, there was a trend towards an advantage for higher ART scores on first pass times,  $\hat{\beta} = -3.12, SE = 1.35, t = -2.31, p = .05$ . Higher RSpan scores were also associated with reduced fixations during first pass reading  $\hat{\beta} = 0.05, SE = 0.02, t = 2.44, p = .05$ . Similar results were observed when all of the individual difference measures were specified as additive predictors within a single model.

### Models of intrinsic lexical features and individual differences

Five central individual difference measures (ART, RAN.dur, EVS, RSpan, RM) were added individually to the three models in Table 3. Two additional terms were included in each model: a simple effect of the individual difference measure and its interaction with frequency. Only the interactions are reported here.

In first fixation durations, higher ART scores interacted with frequency,  $\hat{\beta} = -2.11, SE = 0.74, t = -2.85, p < .001$ . Two groups were formed by a median split on ART scores (High ART: 7 subjects; Low ART: 8 subjects). Box plots for the two groups at 8 bins of roughly equal sizes are shown in Figure 1, panel B. The pattern indicates that the High ART group benefitted most from words at the very highest end of the frequency range, but the two groups were roughly equivalent for lower frequency bins. The results are compatible with

previous findings that a smaller frequency effect is obtained for High ART groups, who progress more quickly through text in general (Sears et al., 2006; Jared et al., 1999).

Readers who completed the RAN task more quickly were also subject to a smaller frequency effect,  $\hat{\beta} = -2.03, SE = 0.75, t = -2.70, p < .05$ . In addition, readers with more expansive eye-voice spans tended to be less affected by the frequency effect, though the trend was not significant,  $\hat{\beta} = -1.33, SE = 0.76, t = -1.75, p = .08$ . A median split analysis on participants was conducted as above, and revealed that readers with lower EVS made shorter first fixations on less frequent words [0.31–1.81 log frequency].

Performance on the Reading span and Raven’s progressive matrices tasks did not interact with intrinsic lexical features for any of the eye-movement measures explored here.

### Conclusion

We have introduced a freely available, high-quality pilot corpus of naturalistic reading paired with a battery of individual differences. The subjects and materials were selected to maximize reader engagement in order to capture the reading profiles exhibited on naturalistic text.

The corpus was subjected to an analysis of early eye movement measures standardly implicated in lexical recognition and retrieval during unrestricted reading. The analysis concentrated on the effects of three intrinsic lexical characteristics: length, word frequency, and word complexity. Classical effects of word length and frequency were replicated, and shown to interact with increased morphological complexity in measures obtained from multiple word fixations. In addition, there was evidence for a possible trade-off between first fixation measures and slightly later eye movement measures (first pass times and number of first pass fixations) for difficult to process words, compatible with the predictions of current models of oculomotor control in reading.

Individual difference measures were added to the best-fitting model of the relationship between each eye movement measures and intrinsic lexical properties. We found support for a reduced frequency effect for individuals showing better performance on the RAN task and increased knowledge of authors on the ART task, in line with previous studies (e.g., Kuperman & Van Dyke, 2011; Kuperman et al., 2016;

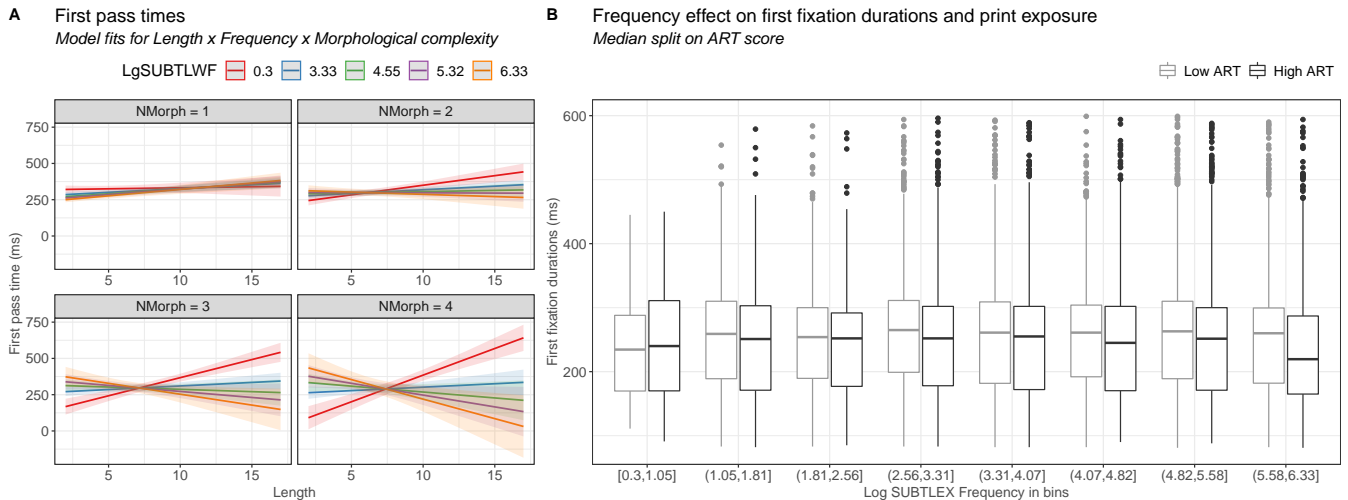


Figure 1: Model fits of first pass times depicting the three-way interaction between length and frequency (binned into at lower, median, and upper quartiles) at four levels of morphological complexity (Panel A). Frequency effect for High and Low ART groups in a median split analysis (Panel B).

Moore & Gordon, 2015; Sears et al., 2006). The patterns suggest that specific individual differences measures tap into processes employed while reading long-form naturalistic text for comprehension. Although the analyses presented here focused on how intrinsic lexical characteristics were modulated by individual differences in early eye-movement measures, later eye movement measures, such as patterns of rereading, might also provide insight into the factors that determine how readers progress through a text.

The design choices of eye-movement corpora naturally reflect the priorities of the researchers, as well as practical limitations of the data collection process, consequently serving some purposes better than others. This particular corpus was designed to address a series of broad questions concerning how highly motivated readers approach long-form naturalistic text given their general exposure to print and their performance on a battery of cognitive tasks. We were therefore most concerned with presenting engaging texts to readers who were likely to have advanced reading skills and who were internally motivated to read for comprehension.

Although the sampling method reflects our research priorities, we hope to develop this pilot corpus to include a larger pool of subjects with more diverse educational backgrounds, possibly recruiting readers outside of a college community. In addition, we hope to increase the sample size, which though small compared to most controlled experiments, is on par with the number of monolingual subjects included in the Dundee (10 monolinguals) and GECO (14 monolinguals) corpora. Finally, we hope to incorporate additional individual difference measures, including those specifically targeting vocabulary size, verbal proficiency, and subjective familiarity with words, to facilitate comparison with other corpora.

We believe that this initial exploration already shows that the corpus provides a rich dataset for exploration, even in its

current pilot form. There are many possibilities for future uses of this corpus. Examples include investigating other established reading effects, such as how a word's orthographic neighborhood density influences word recognition processes, the relationship between performance on the memory and problem solving tasks in later measures of reading, and the effect of contextual predictability or surprisal on reading times, to name a few.

Despite the growing research into individual differences in reading, we currently have relatively little understanding of what prompts different reading patterns among individuals, and what such strategies indicate about comprehension, attention, and general engagement, beyond of a broad description of the patterns. We hope that this publicly accessible tool will help fill an existing gap in the literature.

## References

- Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Beh Res Methds*, *40*(1), 278–289.
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Jrnl of Vrbl Lrng and Vrbl Behav*, *11*(6), 717–726.
- Cop, U., Dirix, N., Drieghe, D., & Duyck, W. (2017). Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Beh Res Methds*, *49*(2), 602–615.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Jrnl of Vrbl Lrng and Vrbl Behav*, *19*(4), 450–466.
- Denckla, M. B., & Rudel, R. G. (1976). Rapid 'automatized' naming (RAN): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, *14*(4), 471–479.

- Dirix, N., & Duyck, W. (2017). An eye movement corpus study of the age-of-acquisition effect. *Psych Bull & Rev*, 24(6), 1915–1921.
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psych Rev*, 112(4), 777–813.
- Gordon, P. C., & Hoedemaker, R. S. (2016). Effective scheduling of looking and talking during rapid automatized naming. *JEP:HPP*, 42(5), 742–760.
- Gordon, P. C., Moore, M., Choi, W., Hoedemaker, R. S., & Lowder, M. W. (2019). Individual differences in reading: Separable effects of reading experience and processing skill. *Mem & Cog*, 48, 1–13.
- Hyönä, J., Lorch Jr, R. F., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Jrnl of Ed Psych*, 94(1), 44–55.
- Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perceptn & Psychophys*, 40(6), 431–439.
- Jared, D., Levy, B. A., & Rayner, K. (1999). The role of phonology in the activation of word meanings during reading: evidence from proofreading and eye movements. *JEP:Gen*, 128(3), 219–264.
- Jones, M. W., Ashby, J., & Branigan, H. P. (2013). Dyslexia and fluency: parafoveal and foveal influences on rapid automatized naming. *JEP:HPP*, 39(2), 554–567.
- Kennedy, A. (2003). *The Dundee corpus [CD-ROM]*. Psychology Department, University of Dundee.
- Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vis Res*, 45(2), 153–168.
- Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Jrnl of Cog Psych*, 16(1-2), 262–284.
- Kuperman, V., Dambacher, M., Nuthmann, A., & Kliegl, R. (2010). The effect of word position on eye-movements in sentence and paragraph reading. *QJEP*, 63(9), 1838–1857.
- Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *JML*, 65(1), 42–73.
- Kuperman, V., Van Dyke, J. A., & Henry, R. (2016). Eye-movement control in ran and reading. *Scientific Studies of Reading*, 20(2), 173–188.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Jrnl of Stat Software*, 82(13), 1–26.
- Lemhöfer, K., & Broersma, M. (2012). Introducing Lex-TALE: A quick and valid lexical test for advanced learners of English. *Beh Res Methds*, 44(2), 325–343.
- Levin, H., & Addis, A. B. (1979). *The eye-voice span*. Cambridge, MA: MIT Press.
- Luke, S. G., & Christianson, K. (2018). The Provo Corpus: A large eye-tracking corpus with predictability norms. *Beh Res Methds*, 50(2), 826–833.
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Spch, Lang, and Hrng Res*, 50, 940–967.
- Mol, S. E., & Bus, A. G. (2011). To read or not to read: a meta-analysis of print exposure from infancy to early adulthood. *Psych Bull*, 137(2), 267–296.
- Moore, M., & Gordon, P. C. (2015). Reading ability and print exposure: item response theory analysis of the author recognition test. *Beh Res Methds*, 47(4), 1095–1109.
- O'Regan, J. K. (1990). Eye movements and reading. *Rev of Oculomotor Res*, 4, 395–453.
- Raven, J. (1989). The Raven Progressive Matrices: A review of national norming studies and ethnic and socioeconomic variation within the United States. *Jrnl of Ed Meas*, 26(1), 1–16.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psych Bull & Rev*, 124(3), 372–422.
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Mem & Cog*, 14(3), 191–201.
- Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C., Jr. (2012). *Psychology of Reading* (2nd ed.). Psychology Press.
- Rayner, K., & Raney, G. E. (1996). Eye movement control in reading and visual search: Effects of word frequency. *Psych Bull & Review*, 3(2), 245–248.
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psych Rev*, 105(1), 125–157.
- Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E-Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cog Sys Research*, 7(1), 4–22.
- Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Beh and Brain Sci*, 26(4), 445–526.
- Sears, C. R., Campbell, C. R., & Lupker, S. J. (2006). Is there a neighborhood frequency effect in English? Evidence from reading and lexical decision. *JEP:HPP*, 32(4), 1040–1062.
- Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Rdng Rsrch Qrtly*, 402–433.
- Wotschack, C. (2009). *Eye movements in reading strategies*. Unpublished doctoral dissertation, Universitätsverlag Potsdam, Postdam, Germany.
- Yarbus, A. L. (1967). *Eye movements during perception of complex objects*. Springer.