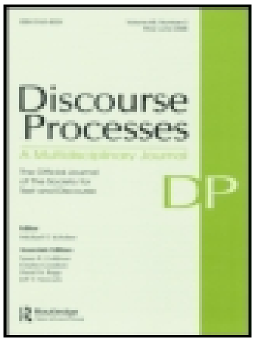

Discourse Processes

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hdsp20

Lost in a Story, Detached from the Words

Lynn S. Eekhof , Moniek M. Kuijpers , Myrthe Faber , Xin Gao , Marloes Mak , Emiel van den Hoven & Roel M. Willems

To cite this article: Lynn S. Eekhof , Moniek M. Kuijpers , Myrthe Faber , Xin Gao , Marloes Mak , Emiel van den Hoven & Roel M. Willems (2021): Lost in a Story, Detached from the Words, Discourse Processes, DOI: 10.1080/0163853X.2020.1857619

To link to this article: https://doi.org/10.1080/0163853X.2020.1857619

Published online: 06 Jan 2021.

DISCOURSE PROCESSES
https://doi.org/10.1080/0163853X.2020.1857619

Routledge
Taylor & Francis Group

OPEN ACCESS

Check for updates

Lost in a Story, Detached from the Words

Lynn S. Eekhof[a], Moniek M. Kuijpers[b], Myrthe Faber[c,d], Xin Gao[a], Marloes Mak[a], Emiel van den Hoven[e], and Roel M. Willems[a,c,f]

[a]Centre for Language Studies, Radboud University, Nijmegen, The Netherlands; [b]Digital Humanities Lab, University of Basel, Basel, Switzerland; [c]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands; [d]Department of Communication and Cognition, Tilburg Center for Cognition and Communication, Tilburg University, Tilburg, The Netherlands; [e]Linguistics Department, University of Potsdam, Potsdam, Germany; [f]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

ABSTRACT

This article explores the relationship between low- and high-level aspects of reading by studying the interplay between word processing, as measured with eye tracking, and narrative absorption and liking, as measured with questionnaires. Specifically, we focused on how individual differences in sensitivity to lexical word characteristics—measured as the effect of these characteristics on gaze duration—were related to narrative absorption and liking. By reanalyzing a large data set consisting of three previous eye-tracking experiments in which subjects (N = 171) read literary short stories, we replicated the well-established finding that word length, lemma frequency, position in sentence, age of acquisition, and orthographic neighborhood size of words influenced gaze duration. More importantly, we found that individual differences in the degree of sensitivity to three of these word characteristics, i.e., word length, lemma frequency, and age of acquisition, were negatively related to print exposure and to a lesser degree to narrative absorption and liking. Even though the underlying mechanisms of this relationship are still unclear, we believe the current findings underline the need to map out the interplay between, on the one hand, the technical and, on the other hand, the subjective processes of reading by studying reading behavior in more natural settings.

Introduction

Reading is a complex and multidimensional process that can take the reader from the processing of orthographic units to the construction of entire story worlds. Both the low-level aspects, e.g., word identification, and the high-level aspects of reading, e.g., situation model building and forms of narrative engagement, have been studied extensively in previous studies. However, as researchers have noted, not much attention has been dedicated to combining these two lines of research and linking these two aspects of reading (e.g., Perfetti & Stafura, 2014; Radach et al., 2008). This is surprising, given the fact that both aspects are crucial for a pleasurable reading experience, and we are, as of yet, unaware of how they interact to lead to pleasurable reading. In this article we therefore explore the relationship between low-level and high-level aspects of reading by focusing on word processing in the context of literary narratives. In what follows, we will focus on (a) the role of lexical characteristics in word processing, (b) sources of variability in word processing as a result of high-level aspects of reading, and (c) high-level processes specifically related to the reading of literary narratives.

CONTACT Lynn S. Eekhof l.eekhof@let.ru.nl Centre for Language Studies, Postbus 9103, Nijmegen 6500 HD, The Netherlands.

Supplemental data for this article can be accessed on the publisher's website.

Low-level aspects of reading: Word processing

Word processing is at the core of reading (Perfetti & Stafura, 2014) and has been studied in great detail. One important line of research has investigated how lexical characteristics influence the time a reader needs to identify a word. For example, it has been shown that shorter words, words that are used more frequently, and words with a more concrete meaning are read faster than long, low-frequency, and abstract words (Juhasz & Rayner, 2003, 2006; Kliegl et al., 2004; Rayner & Duffy, 1986). Age of acquisition also modulates reading speed: Words acquired later in life are generally read at a slower rate (Juhasz & Rayner, 2003). Similarly, orthographic neighborhood size (i.e., the number of word types that differ from the target word by only one letter) has been found to have an inhibiting effect on reading. That is, readers spend more time looking at words with many neighbors (Pollatsek et al., 1999; but see Perea & Rosa, 2000, for an overview of conflicting findings from reading and lexical decision studies).

The influence of high-level aspects of reading

Interestingly, there is variability in the degree to which word characteristics influence reading behavior. Some of this variability can be traced back to high-level aspects of reading related to both readers and texts. We will now discuss some of these high-level sources of variation.

Reading skill

Low-level reading behavior has been found to be influenced by individual differences in reading skill between readers.[1] For example, although low-frequency words are usually read slower compared to high-frequency words, readers with better reading skills (Ashby et al., 2005; Kuperman & Van Dyke, 2011; Perfetti et al., 1979), higher vocabulary scores (Mainz et al., 2017), increased print exposure (Chateau & Jared, 2000; Sears et al., 2008), and a higher educational background (Tainturier et al., 1992) display smaller word frequency effects. That is, differences in processing time between low- and high-frequency words are less pronounced in skilled readers, compared to relatively poor readers. There is some evidence to suggest that the effects of orthographic neighborhood size and word length also vary as a function of reading skill (Barton et al., 2014; Chateau & Jared, 2000; Sears et al., 2008; Spinelli et al., 2005). Effects of age of acquisition, on the other hand, do not appear to vary as much, as suggested by the lack of an interaction with vocabulary scores (albeit in a sample of only 12 subjects; Butler & Hains, 1979). The differences between readers of various skill have been argued to arise due to more automatized word processing in skilled readers: As exposure to words increases, more stable lexical representations develop, leading to faster recognition (Brysbaert et al., 2018; Kuperman & Van Dyke, 2011; Mainz et al., 2017; Yap et al., 2009). By contrast, less skilled readers' reading behavior is less automatic and shows greater reliance on both word characteristics and context (Kuperman & Van Dyke, 2011; Leinenger & Rayner, 2017; Perfetti et al., 1979).

Contextual richness

Besides an effect of reading skill, there is evidence that the richness of the context in which a word is read also influences how it is processed. For example, reading words embedded in isolated sentences leads to larger word frequency effects than reading words embedded in longer excerpts (Radach et al., 2008). Radach and colleagues argue that this is due to the fact that more postlexical processing happens in rich reading contexts, which makes effects of purely lexical reading processes less pronounced in such contexts. Similarly, Dirix et al. (2019) compared eye-tracking data from subjects reading an entire novel (GECO corpus) or newspaper articles (Dundee corpus) to data from lexical decision studies. They found that the word frequency effect was larger in the reaction times from lexical decision experiments than in the eye-tracking data from reading studies. They interpret this as lexical characteristics of words becoming redundant in the presence of rich contextual information.

In sum, studies thus far seem to suggest that increased reading skills and more natural reading contexts lead to a decreased effect of word characteristics on low-level word processing. We will refer to this relationship as a decrease in readers' *sensitivity* to low-level word characteristics. One of the natural reading contexts that is relatively understudied in this respect thus far is the reading of literary narratives. As we will argue, narrative contexts elicit a range of high-level processes that might interact with low-level reading processes such that the ways in which readers engage with a narrative might similarly be associated with variability in low-level reading behavior.

High-level aspects of reading in a literary context

Pleasurable reading experiences with *literary narratives* come in all forms and vary from one reader to the next. Empirical research on the topic of literary reading has shown that there is a range of psychological responses to the reading of literary texts that depend on the type of text that is read, the reading habits, and preferences of the reader, as well as their reading skill. Most of the research seems to suggest that there is a distinction between two main types of reading experiences, with engaging reading experiences such as flow (e.g., Thissen et al., 2018), transportation (e.g., Green & Brock, 2000), or narrative absorption (e.g., Hakemulder et al., 2017) on the one hand, and esthetic reading experiences, such as esthetic emotions (e.g., Schindler et al., 2017), artifact emotions (e.g., Tan, 1996), or being moved (e.g., Menninghaus et al., 2015) on the other. The Neuro-Cognitive Poetics Model (NCPM, Jacobs, 2015) outlines this divide and the assumptions underlying it in one model: The immersion route posits that the reading of simple, backgrounded texts will lead to fluent processing, which in turn might elicit engaging reading experiences, whereas the esthetic route is argued to come into play when readers read complex, foregrounded texts that will lead to disfluent processing and esthetic reading experiences. Even though Jacobs and his colleagues themselves stress that these types of reading experiences are not mutually exclusive (Jacobs & Lüdtke, 2017), the research field studying these two types of narrative experiences remains mostly divided.

Recently, however, research has shown that the two routes in the NCPM cannot always be clearly separated. For instance, it has been argued that foregrounding devices (i.e., text features used in literary texts to elicit defamiliarization) make it easier for readers to focus on particular parts of a text (Sanford & Emmott, 2012; see Kuijpers, 2014 for empirical evidence that foregrounded texts lead to absorption in high print exposure readers) and that they can evoke reflection and emotional responses facilitating absorption (Bálint et al., 2016). Thus it seems that a disruption in fluent reading could also lead to an absorbing reading experience. Still, most researchers of literary reading assume that engaged experiences arise when less attention is paid to the formal characteristics of the text due to fluent reading, leading the reader to focus on the story world instead.

In this article, we will focus on the experience of narrative absorption because it has been linked reliably to reading enjoyment, appreciation, and reading motivation and because we want to further explore the supposed relationship between reading behavior and narrative absorption. Narrative absorption is an experiential state that leaves the reader feeling "lost in a book" (see Kuijpers, 2014; Nell, 1988), which results in a lower bodily awareness, a lower awareness of surroundings, and an altered sense of time (Busselle & Bilandzic, 2009; Carleton et al., 2010; Gerrig, 1993; Green & Brock, 2000; Kuijpers et al., 2014). Encompassed in narrative absorption constructs are higher-order cognitive processes such as imaginative engagement, emotional engagement, deictic shift, and focused attention (see *Narrative Engagement*, Busselle & Bilandzic, 2009; *Transportation*, Green & Brock, 2000; *Story World Absorption*, Kuijpers et al., 2014). Some multidimensional constructs of narrative absorption even include a dimension of narrative understanding, such as the Narrative Engagement Scale by Busselle and Bilandzic (2009), which was developed for use with audiovisual media and uses theories of situation model building for its theoretical foundation. Whether or not narrative comprehension is a part of narrative absorption is a contested issue, though most seem to agree that at the very least narrative comprehension is a precondition of narrative absorption (Busselle & Bilandzic, 2009; Kuijpers, 2014).

Readers vary in their tendency to become absorbed in narratives. Previous studies have shown that print exposure, as measured with the Author Recognition Test (ART; Stanovich & West, 1989), influences narrative absorption scores differently depending on the text that is being read. High-print-exposure readers seem to prefer—and get absorbed more easily in—more literary narrative texts, whereas low-print-exposure readers feel absorbed more easily in nonliterary narrative texts (Kuijpers, 2014).

The current study

As we have seen, there is accumulating evidence that shows that high-level aspects of reading, e.g., reading skill and contextual richness, can influence low-level reading behavior, such as the degree to which readers are sensitive to lexical characteristics during word processing. Because of the relevance of high-level processes during literary, narrative reading, we will investigate the relationship between word processing on the one hand and forms of narrative engagement on the other hand. Specifically, we will focus on how sensitivity to word characteristics is related to narrative absorption and liking, while also controlling for differences in print exposure.

To do so, we reanalyzed eye-tracking data from three previously published studies. In each of these studies subjects read literary short stories while their eye movements were tracked. Several experiential responses were measured after reading, including subjects' narrative absorption and liking of the story.

As the data were originally collected for different purposes, our study is best characterized as exploratory. We did not have clearly delineated a priori hypotheses about the relationship between sensitivity to word characteristics and narrative absorption and liking. However, previous research on the effect of contextual richness on reading behavior (e.g., Dirix et al., 2019) suggests that readers who are more engaged in a story might rely more on the discourse and less on word characteristics. Moreover, in line with research on the role of fluency in esthetic experiences (e.g., Gao et al., 2019; Reber et al., 2004; Vaughn et al., 2010), Jacobs's (2015) Neurocognitive Poetics Model predicts that when readers enter an immersive state, reading becomes more automatic and fluent.

Note that due to the nature of the design of the study, we can only detect correlational relationships between sensitivity to word characteristics and narrative absorption. The causal mechanisms behind any of the detected relationships (i.e., whether decreased/increased sensitivity leads to more narrative absorption or whether more narrative absorption leads to decreased/increased sensitivity) will need to be explored in future research (see also Discussion). Nevertheless, this study leverages a data set of notable size, allowing us to map out the possible relationships between low- and high-level aspects of reading.

Methods

Data sets from three previous eye-tracking studies conducted at Radboud University and the Max Planck Institute for Psycholinguistics in Nijmegen were combined for this study. In each of these studies, subjects read literary stories under natural reading circumstances, i.e., without an additional task, while their eye movements were measured. The first study (Van den Hoven et al., 2016) was an inquiry about the effect of foregrounding (i.e., aspects of literary style) on reading behavior. The second study (Mak & Willems, 2018) studied the different types of simulation readers engage in during story reading. The third study (Eekhof et al., 2018) tested the effect of verb tense on mental simulation during literary reading.

Subjects

The combined data set analyzed in this study contains data from 171 subjects aged between 18 and 46 years (M = 22.95, SD = 4.46; 137 females, 32 males, 1 other) and does not contain data that were already rejected in the previous studies. In the first study, 29 subjects (24 females) with a mean age of 21.93 (SD = 2.67, range = 18–28) were included after one subject, and 10 individual story readings were rejected due to inaccurate eye-tracking data. In the second study, 102 subjects (81 females) with a mean age of 23.27 (SD = 4.59, range = 18–40) were included after seven subjects were rejected due to inaccurate eye-tracking data or insufficient performance on comprehension check questions, four individual story readings were rejected due to inaccurate eye-tracking data, and eight individual story readings were rejected due to insufficient performance on comprehension check questions. In the third study, 40 subjects (32 females) with a mean age of 22.88 (SD = 5.11, range = 18–46) were included after three subjects were rejected due to inaccurate eye-tracking data, and one individual story reading was rejected due to insufficient performance on comprehension check questions. Subjects were recruited from the subject pool of the Max Planck Institute for Psycholinguistics (Study 1) or the subject pool of Radboud University (Studies 2 and 3) in Nijmegen. All subjects reported normal or corrected-to-normal vision and were native speakers of Dutch.

Studies were approved by the local ethics committees and were conducted in accordance with the Declaration of Helsinki. Subjects signed informed consent forms at the start of the experiment and received €16 (Study 1) or €15 (Studies 2 and 3) or course credit for their participation.

Materials

All studies used existing and published Dutch literary short stories (see Table 1), which is considered a genre on its own (see, e.g., Pratt, 1981). Literary short stories usually revolve around a limited number of protagonists and a single, central event (Abrams & Harpham, 2009). The stories in this data set had various topics, ranging from, for example, a fatal accident during a break-in (*De vijand*, Jacques Hamelink), to a wedding (*De Chinese bruiloft*, Sanneke van Hassel), to an unexpected reunion with a former teacher (*De invaller*, René Appel). In Studies 1 and 2, the stories were presented in their original form. In Study 3, in addition to the original stories, adapted versions were created in which the tense of each story was changed from present to past tense or vice versa (with no difference in readability reported between the original and adapted versions, see Eekhof et al., 2018).[2] All stories were originally written in Dutch, with the exception of *Symbols and Signs* (Study 2), for which a professional and published Dutch translation was used.

The stories were annotated on a word-by-word basis for several word characteristics: word length (number of letters), lemma frequency (the absolute sum of occurrences of all forms of one lemma in the 44 million word SUBTLEX-NL corpus; Keuleers et al., 2010), age of acquisition (the average age at which a word is first acquired; Brysbaert et al., 2014), concreteness (the extent to which the referent of a word is a perceptible object on a scale from 1 to 5; Brysbaert et al., 2014), orthographic neighborhood

Table 1 Descriptive Information and Statistics for the Stimulus Stories in the Combined Data Set.

Study	Story	Author	Year of Publication	Word count	Mean SWAS Score (*SD*)	Mean Liking Score (*SD*)
1	*De straf* 'The Punishment'	J. Verstegen	1973	2,982	4.47 (0.75)	4.61 (0.86)
	Kogeltjes 'Bullets'	W. Melchior	1992	3,526	4.57 (0.73)	4.59 (1.06)
	De vijand 'The Enemy'	J. Hamelink	1966	2,436	4.66 (0.78)	4.70 (1.03)
2	*De mensen die alles lieten bezorgen* 'The People That Had Everything Delivered'	R. van Essen	2014	2,988	4.89 (0.79)	5.17 (0.90)
	De Chinese bruiloft 'The Chinese Wedding'	S. van Hassel	2012	2,659	4.06 (1.01)	4.01 (1.20)
	Signalen en symbolen 'Symbols and Signs'	V. Nabokov	1948/2003	2,143	3.56 (1.02)	3.68 (1.32)
3	*Het is muis* 'It's Mouse'	S. van Hassel	2012	2,016	3.96 (1.14)	4.73 (1.45)
	Hoe de wolven dansen 'How the Wolves Dance'	J. Lammers	2017	1,176	4.28 (1.09)	4.85 (1.14)
	De invaller 'The Substitute'	R. Appel	2003	743	4.66 (1.00)	5.15 (0.95)
	Ze is overal 'She Is Everywhere'	E. van Eeden	2015	1,074	4.29 (1.06)	5.05 (1.09)

Note. Study 1 = Van den Hoven et al. 2016. Study 2 = Mak and Willems 2018. Study 3 = Eekhof et al. 2018.

size (the number of words that differ from the target word by just one letter; Marian et al., 2012), and the absolute position of the word in the sentence.

In all studies, a questionnaire with 7-point scales from 1 (*totally disagree*) to 7 (*totally agree*) was used to measure narrative absorption after each story. In the first study, this questionnaire included 11 items from the Story World Absorption Scale (SWAS; Kuijpers et al., 2014). In Studies 2 and 3, this questionnaire included the complete set of 18 SWAS items. As both questionnaires were designed to measure the same construct and contained mostly overlapping items, we decided to use the questionnaires in their original forms in our analyses, despite the difference in the number of items. In addition, subjects were asked to indicate how much they liked the story on either a 10-point scale—Study 1 from 1 (*very bad*) to 10 (*very good*)—or a 7-point scale—Studies 2 and 3 from 1 (*very bad*) to 7 (*very good*). Finally, a Dutch translation of the ART (Koopman, 2015) was used in all studies, which assesses readers' print exposure by having them underline names of authors they recognize from a list of real and fake names. To check whether subjects actually read the stories and paid a sufficient amount of attention, three (Studies 2 and 3, four response options) or four (Study 1, three response options) comprehension check questions were created. Subjects who answered more than one (Studies 2 and 3) or two (Study 1) of these questions incorrectly were excluded on a story-by-story basis (see Subjects section).

Data recording and stimulus presentation

Eye movements were measured with an EyeLink 1000 Plus eye tracker that recorded at a sampling rate of 1,000 Hz (Studies 1 and 3) or 500 Hz (Study 2). In all studies, a chin rest was used to reduce head movements. Eye movements were measured for the dominant eye when possible. The stories were presented using SR Research Experiment Builder (SR Research, Ottawa, Canada) on an Acer AL2023 20" LCD screen (Study 1) or on a BenQ XL 2420 T 24" LED screen, with a 1,024 × 768 (32 bits per pixel) resolution (Studies 2 and 3) and a refresh rate of 60 Hz. Every story was preceded by a 9-point calibration and validation session. Stories were divided into sections that resembled the original paragraphs of the stories as much as possible and were presented in easily readable fonts (see Supplementary Table S1). Every section started with a 1,000 ms presentation of a fixation cross marking the position of the first word. A drift correction took place every five to 10 sections (Study 1) or every five (Study 2) or four (Study 3) sections. In all studies, each word of each story was automatically defined as an interest area by SR Research Data Viewer (SR Research, Ottawa, Canada) such that there was no space between any two horizontally or vertically adjacent interest areas.

Procedure

The experiments took place in soundproof booths. In each study, subjects were instructed to read as naturally as possible, while trying to minimize any movements. They were informed that they would be asked about their reading experiences after each story. The design of each study was such that every subject read all three or four stories in a counterbalanced order. After each story, the narrative absorption and liking questionnaires were filled out on paper outside the booth. After the final story, subjects completed the ART, the comprehension check questions, and other questionnaires not relevant for the purposes of the current study.

Data analysis

Preprocessing of eye-tracking data

The raw data of all three previous studies had been preprocessed prior to the current study, using SR Research Data Viewer (SR Research, Ottawa, Canada), as well as RStudio (R Core Team, 2017) in the context of the research for which the data were originally collected. All studies used a similar approach.

First of all, fixations were checked visually in SR Research Data Viewer (SR Research, Ottawa, Canada) for each section of each story on a subject-by-subject basis. When fixations diverged too much from the lines of the text, entering the interest areas of another line, they were manually aligned on the vertical axis. If manual alignment was not possible, individual sections were removed. Secondly, Studies 2 and 3 removed all fixations that occurred before subjects made a fixation on any area of interest on the first line of a section. No further automatic cleaning procedures (e.g., merging small adjacent fixations) were applied to the fixations, as both the previous studies and the current study were interested in aggregate measures such as gaze duration and regression probability. From the manually preprocessed fixations, interest area reports that included the metric of interest for this study, gaze duration (i.e., the summation of all fixations on an interest area during the first run of reading), were generated for each subject in SR Research Data Viewer (SR Research, Ottawa, Canada). These reports were then further preprocessed in RStudio (R Core Team, 2017). Study 1 removed gaze durations that deviated 3.5 or more standard deviations from the mean. In Study 2, gaze durations shorter than 50 ms or longer than 3,600 ms were removed. In Study 3, gaze durations shorter than 50 ms or longer than 1,200 ms were removed.

After combining the three original data sets, we unified data preprocessing as much as possible by removing any remaining gaze durations < 50 ms and > 1,200 ms from the combined data set, as well as interest areas for which no gaze duration data were available because readers skipped them (i.e., NAs). This led to a removal of 38.16% of the data, i.e., 26.43% of all content words and 52.57% of all function words. To avoid a spillover effect of the fixation cross, data for the first word of each section were removed. The preprocessing steps are also visualized in Figure 1.

We decided to focus on gaze duration (i.e., the summation of all fixations on an interest area during the first run of reading) for our analyses because this metric reflects sensitivity to lexical features (see, e.g., Rayner & Duffy, 1986; Schilling et al., 1998) and is thought to be an encompassing measure of early lexical processes (Kliegl & Laubrock, 2017). Moreover, this metric has been argued to reflect the "upper bound of early processing" (Kliegl & Laubrock, 2017, p. 77) and thus seems to be most relevant for studying the interactions between low- and high-level processes of reading. By taking this targeted, theory-driven approach, we avoid issues pertaining to multiple but dependent comparisons that occur when analyzing all possible metrics that can be derived from eye-tracking data (Kliegl & Laubrock, 2017; Orquin & Holmqvist, 2018; Von der Malsburg & Angele, 2017).

The final, combined data set that was used in this study contained 729,403 observations (478,336 content words, i.e., 65.58%, and 251,067 function words, i.e., 34.42%) or 4,479 unique words (4,290 content words, i.e., 95.78%, and 189 function words, i.e., 4.22%) for which gaze duration data were available, as well as information regarding the reader and their SWAS, liking, and ART scores. Descriptive statistics for the eye-tracking data are given in Table 2.

Lemma frequency information was available for 93.50% of the unique words in the data set, age of acquisition information was available for 51.44% of the unique words in the data set, concreteness information was available 51.26% of the unique words in the data set, orthographic neighborhood size was available for 74.15% of the unique words in the data set, and naturally, information about the length and sentence position of the word was available for all words. There were 278 unique words (6.21%) or 17,338 observations in the whole data set (2.38%) for which no word characteristic data were available other than word length and word position.[3] More descriptive statistics for the word characteristics are given in Table 3.

Table 2. Descriptive Statistics for the Eye-Tracking Data Set.

	All Words	Content Words	Function Words
Mean gaze duration (*SD*)	243.57 ms (136.72 ms)	254.18 ms (144.64 ms)	223.35 ms (117.59 ms)
Mean skipping probability (*SD*)	19.83% (39.87%)	16.11% (36.76%)	27.04% (44.41%)
Mean incoming regression probability (*SD*)	23.93% (42.67%)	22.04% (41.46%)	27.59% (44.70%)
Mean outgoing regression probability (*SD*)	18.91% (39.16%)	19.84% (39.88%)	17.11% (37.66%)

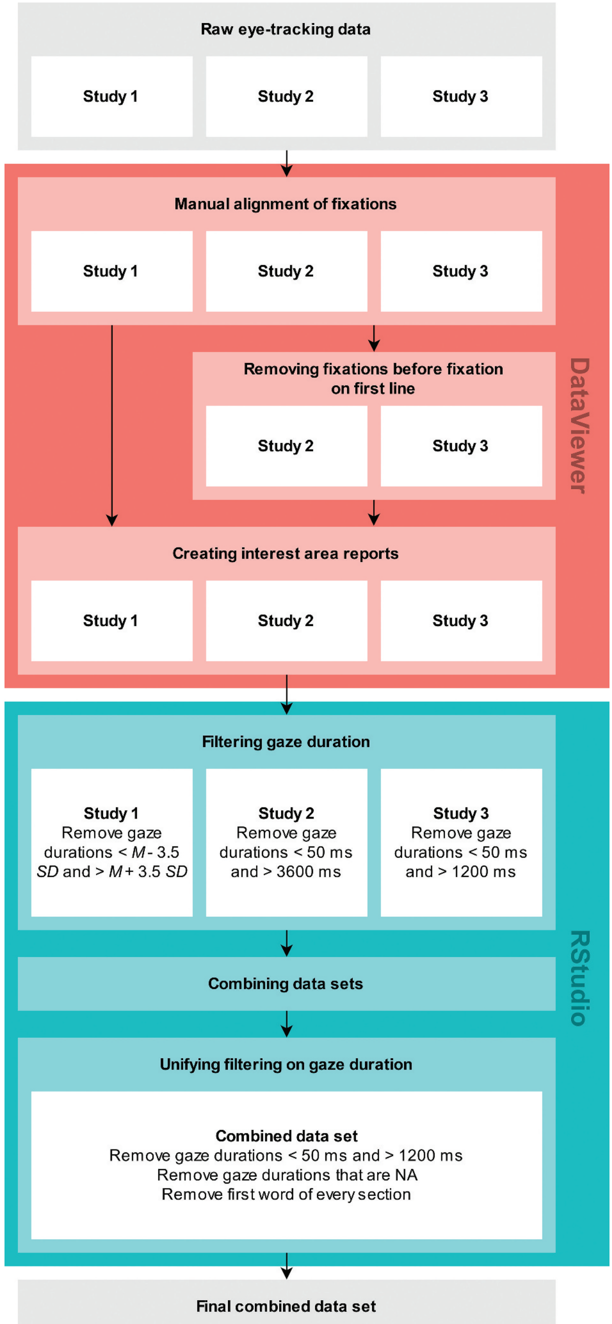

Figure 1. Preprocessing steps for the eye-tracking data.
Note. Study 1 = Van den Hoven et al. 2016. Study 2 = Mak and Willems 2018. Study 3 = Eekhof et al. 2018.

Statistical analyses

We used RStudio (RStudio version 1.3.959, R version 4.0.0; R Core Team, 2017) for all statistical analyses. Linear mixed models were fitted using the *lme4* package (Bates et al., 2015). All predictors of interest were added simultaneously. If a model did not converge, the random effect structure was simplified until the model did converge. To estimate degrees of freedom and *p* values, we used the

Table 3. Descriptive Statistics for the Word Characteristics of the Whole Data Set (Left) and Unique Words (Right).

	Whole Data Set			Unique Words		
	All Words	Content Words	Function Words	All Words	Content Words	Function Words
N (%)	729,403 (100)	478,336 (65.58)	251,067 (34.42)	4,479 (100)	4290 (95.78)	189 (4.22)
Mean word length (*SD*)	5.23 (2.77)	6.32 (2.75)	3.17 (1.21)	7.27 (2.87)	7.39 (2.85)	4.58 (1.83)
Mean lemma frequency (*SD*)	298,996.51 (521,513.33)	146,262.06 (422,667.61)	579,232.91 (567,345.64)	22,430.42 (128,553.15)	16,580.24 (105,104.08)	147,599.31 (343,447.27)
Mean age of acquisition (*SD*)	5.87 (1.73)	6.58 (1.92)	5.09 (1.04)	7.43 (1.99)	7.53 (1.99)	6.13 (1.47)
Mean concreteness (*SD*)	2.43 (1.21)	3.02 (1.23)	1.79 (0.81)	3.08 (1.15)	3.17 (1.12)	1.95 (0.78)
Mean orthographic neighborhood size (*SD*)	10.39 (8.56)	6.62 (6.44)	16.29 (8.14)	4.50 (5.41)	4.27 (5.17)	8.25 (7.62)
Mean position in sentence (*SD*)	10.36 (10.23)	10.55 (10.41)	10.00 (9.88)	11.36 (10.84)	11.41 (10.85)	10.37 (10.51)

Note. The total size of the SUBTLEX-NL corpus, from which lemma frequency measures were taken, is 43,729,424. All word characteristics were scaled and centered for analysis but are presented in their original scale here for interpretability. In addition, lemma frequency was log-transformed.

lmerTest package (Kuznetsova et al., 2017). A function reported online (R-hack/mer-utils.R, 2014) was used to calculate Variance Inflation Factors (VIFs), a quantifier of multicollinearity.

As a measure of narrative absorption, we calculated the mean score per subject per story for the SWAS items used in the different studies. The overall mean SWAS score was 4.27 (SD = 1.05). Since story liking was measured on a 10-point scale in Study 1 but on a 7-point scale in Studies 2 and 3, the scores for Study 1 were converted to a 7-point scale with a linear transformation. As a result of this transformation, the liking score does not only contain the discrete values 1 through 7 but also values in between for the data obtained from Study 1. The overall mean liking score was 4.54 (SD = 1.26). The mean SWAS and liking scores per story are displayed in Table 1.

We also calculated subjects' mean SWAS and liking scores and correlated these with the story-dependent SWAS and liking scores. Both the correlation between mean overall SWAS scores and story-dependent SWAS scores, $r(549)$ =.69, 95% CI [.64, .73], p < .001, and between mean overall liking scores and story-dependent liking scores, $r(543)$ = .63, 95% CI [.57, .68], p < .001, were significant but not perfect. This suggests that although subjects are quite consistent in how easily they are absorbed in a story, and how much they like a story, there is also some potentially story-dependent variation in narrative absorption and liking.

ART scores were calculated by adding up the correctly identified author names and subtracting the incorrectly identified names for each subject. The mean ART score was 7.48 (SD = 4.48) out of a possible maximum score of 30 and a minimum score of −12. SWAS and liking scores were highly correlated, $r(543)$ = .64, 95% CI [.59, .69], p < .001. On the other hand, there was no significant correlation between ART and SWAS scores, $r(549)$ = .04, 95% CI [−.05, .12], p = .36, or ART and liking scores, $r(543)$ = .03, 95% CI [−.06, .11], p = .51.

For all analyses, the word characteristics and questionnaire scores were centered and scaled to improve model fit. In addition, we log-transformed lemma frequency. Statistical analyses were performed both on the entire data set, including both content and function words, and on a subset of the data containing only content words.

Results

We were interested in how sensitivity to word characteristics is related to readers' self-reported measures of narrative absorption, liking, and print exposure. To test these relationships statistically, we employed a three-step analysis. Conceptually, these steps were as follows: First we established which word characteristics influenced gaze duration at the group level by testing for the main effects of word characteristics on gaze duration. Then we inspected individual differences in readers' sensitivity to word characteristics by using the coefficients of the by-subject random slopes for the effect of word characteristics on gaze duration. These coefficients represent the strength of the relationship between the word characteristics and gaze duration for each subject for each story separately. Finally, we related these individual differences in sensitivity to word characteristics, as measured by the coefficients, to the high-level processes measured with questionnaires: the SWAS scores, as a measure of narrative absorption, the liking scores, as a measure of story liking, and ART scores, as a measure of print exposure.[4]

Step 1: The effect of word characteristics on gaze duration

For both the analysis on content and function words, and the analysis on content words only, we started out with a linear mixed model with fixed effects of the six word characteristics, as well as random intercepts and slopes for every word characteristic per subject per story (i.e., story nested in subject). We then simplified the random effect structure until the model converged without singularity issues. In both cases, the final model still contained fixed effects for all six word characteristics but only included random by-subject, by-story (i.e., story nested in subject) intercepts and slopes for word length, lemma frequency, and age of acquisition. The variance inflation factors (VIFs) were < 2 for all fixed effects.

Table 4. Coefficients for the Final Model Predicting Gaze Duration by Word Characteristics Using Content and Function Words (Left) or Content Words Only (Right).

	Content and Function Words					Content Words Only				
Effect	Estimate	*SE*	*df*	*t*	*p*	Estimate	*SE*	*df*	*t*	*p*
(Intercept)	237.60	1.57	536.50	151.48	<.001***	238.75	1.57	552.56	151.66	<.001***
Word length	12.51	0.50	591.61	24.80	<.001***	14.45	0.59	792.09	24.48	<.001***
Lemma frequency	−6.71	0.47	1144.75	−14.16	<.001***	−10.22	0.58	1241.08	−17.72	<.001***
Position in sentence	−0.59	0.18	507,285.62	−3.31	<.001***	−0.78	0.26	257,994.99	−2.98	.003**
Concreteness	−0.35	0.22	504,646.25	−1.61	.11	−2.80	0.31	256,701.18	−9.15	<.001***
Age of acquisition	3.31	0.30	670.03	10.98	<.001***	2.28	0.34	722.97	6.65	<.001***
Orthographic neighborhood size	1.27	0.25	507,625.00	5.10	<.001***	5.15	0.47	257,623.87	10.89	<.001***

Note. All word characteristics were scaled and centered for analysis.
$p < .05$, $^{**}p < .01$, $^{***}p < .001$.

For the analysis on content and function words, five word characteristics turned out to be significant predictors of gaze duration: word length, lemma frequency, position in sentence, age of acquisition, and orthographic neighborhood size (see Table 4). In line with previous findings, lemma frequency was negatively associated with gaze duration (i.e., frequent words were read faster), whereas word length, age of acquisition, and orthographic neighborhood size were positively associated with gaze duration (i.e., longer words, words learned later in life, and words with more neighbors were read slower). Position in the sentence was also negatively associated with gaze duration (i.e., words later in the sentence were read faster).

In addition to these five word characteristics, concreteness was also significantly associated with gaze duration in the analysis performed on content words only (see Table 4). Words with more concrete meanings were read faster.

Step 2: Individual differences in readers' sensitivity to word characteristics

Because the final models from the first step of the analysis contained random slopes per subject per story (i.e., story nested in subject) for three word characteristics (word length, lemma frequency, and age of acquisition), we were able to inspect individual differences between readers and between stories in the degree to which readers' gaze durations were affected by these three word characteristics. We extracted the coefficients of these random slopes from both the model based on content and function words and the model based on content words only as a measure of the strength of the relationships

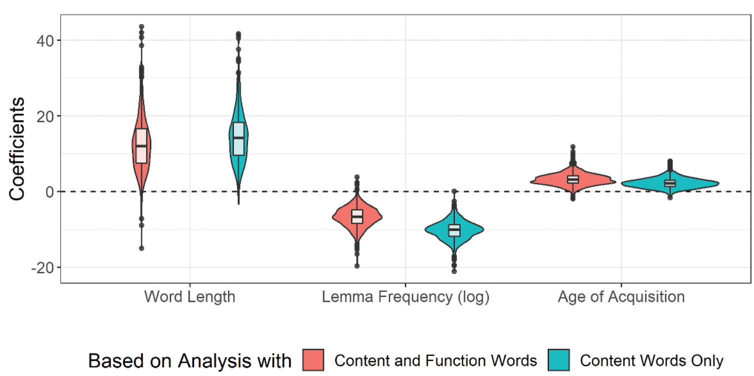

Figure 2. Violin plots and box plots for the coefficients of the random slopes of relationships of word characteristics with gaze duration. *Note*. Word characteristics were scaled and centered for analyses. Dots denote values more than 1.5 times the interquartile range away from the upper or lower quartile.

between the word characteristics and gaze duration for each subject for each story. The coefficients of these relationships as extracted from the two models are visualized in Figure 2. Positive coefficients indicate relatively slower reading (gaze duration increases), whereas negative slopes indicate relatively faster reading (gaze duration decreases) as a result of an increase in the word characteristics. The closer the coefficients are to zero (marked by the dotted line), the weaker the relationship between gaze duration and word characteristics, which we interpreted as decreased sensitivity to these word characteristics. As can be seen from the spread of the coefficients, there are individual differences between subjects with respect to the degree their gaze durations were affected by word length, lemma frequency, and age of acquisition, i.e., with respect to their sensitivity to these word characteristics.[5] In addition, there seems to be more spread in the coefficients extracted from the model based on content words and function words, compared to the coefficients extracted from the content words only model.

Step 3: The relationship between sensitivity to word characteristics and high-level aspects

The coefficients extracted from the two models were used in the third and final step of the analysis, in which we explored whether a relationship exists between subjects' sensitivity to word characteristics, as measured by the coefficients, and the high-level processes of narrative absorption, as measured with SWAS scores, story liking, as measured with liking scores, and print exposure, as measured with ART scores. For each of the three word characteristics for which we had extracted the coefficients of the random slopes, we fitted separate linear mixed models with fixed effects of narrative absorption, liking, and print exposure, and random intercepts for subjects and stories, to test for a relationship between sensitivity to these word characteristics, as measured with the coefficients on the one hand and narrative absorption and liking on the other hand, while controlling for individual differences in print exposure.

Table 5. Coefficients for the Models Predicting Sensitivity to Word Characteristics Based on Content and Function Words (Left) or Content Words Only (Right) by SWAS and ART Scores and Liking and ART Scores.

	Content and Function Words					Content Words Only				
Effect	Estimate	*SE*	*df*	*t*	*p*	Estimate	*SE*	*df*	*t*	*p*
Sensitivity to word length										
Intercept[a]	11.61	0.85	12.58	13.61	<.001***	13.69	0.74	11.71	18.42	<.001***
SWAS scores	−0.64	0.26	469.85	−2.48	.01*	−0.59	0.26	488.84	−2.25	.03
ART scores[a]	−2.07	0.45	164.18	−4.57	<.001***	−1.86	0.42	161.50	−4.47	<.001***
Intercept[b]	11.60	0.88	12.47	13.26	<.001***	13.68	0.78	11.66	17.61	<.001***
Liking scores	−0.48	0.24	432.33	−2.01	.05	−0.39	0.25	450.26	−1.56	.12
ART scores[b]	−2.13	0.45	165.90	−4.72	<.001***	−1.92	0.42	163.99	−4.62	<.001***
Sensitivity to lemma frequency										
Intercept[a]	−6.73	0.19	38.29	−34.81	<.001***	−10.17	0.20	18.78	−49.89	<.001***
SWAS scores	0.18	0.12	151.23	1.55	.12	0.15	0.11	439.56	1.30	.20
ART scores[a]	0.39	0.18	167.86	2.17	.03	0.37	0.15	165.29	2.52	.01*
Intercept[b]	−6.74	0.19	44.12	−36.28	<.001***	−10.18	0.21	18.60	−48.87	<.001***
Liking scores	0.31	0.11	121.54	2.78	.006**	0.14	0.11	452.43	1.31	.19
ART scores[b]	0.39	0.18	168.37	2.18	.03	0.36	0.15	164.24	2.48	.01*
Sensitivity to age of acquisition										
Intercept[a]	3.16	0.16	14.01	19.30	<.001***	2.10	0.16	10.79	13.49	<.001***
SWAS scores	−0.09	0.07	505.41	−1.27	.21	−0.10	0.05	507.69	−1.76	.08
ART scores[a]	−0.41	0.10	161.95	−4.12	<.001***	−0.36	0.08	161.62	−4.51	<.001***
Intercept[b]	3.16	0.17	13.72	18.25	<.001***	2.10	0.16	10.83	13.01	<.001***
Liking scores	−0.02	0.07	479.36	−0.28	.78	−0.06	0.05	464.57	−1.11	.27
ART scores[b]	−0.42	0.10	163.84	−4.24	<.001***	−0.37	0.08	163.61	−4.65	<.001***

Note. SWAS, liking, and ART scores were scaled and centered for analysis. Due to collinearity, the effect of SWAS scores and liking scores were tested in separate models.
[a]Coefficients based on the model that included SWAS and ART scores as predictors.
[b]Coefficients based on the model that included liking and ART scores as predictors.
**p* < .025, ***p* < .01, ****p* < .001.

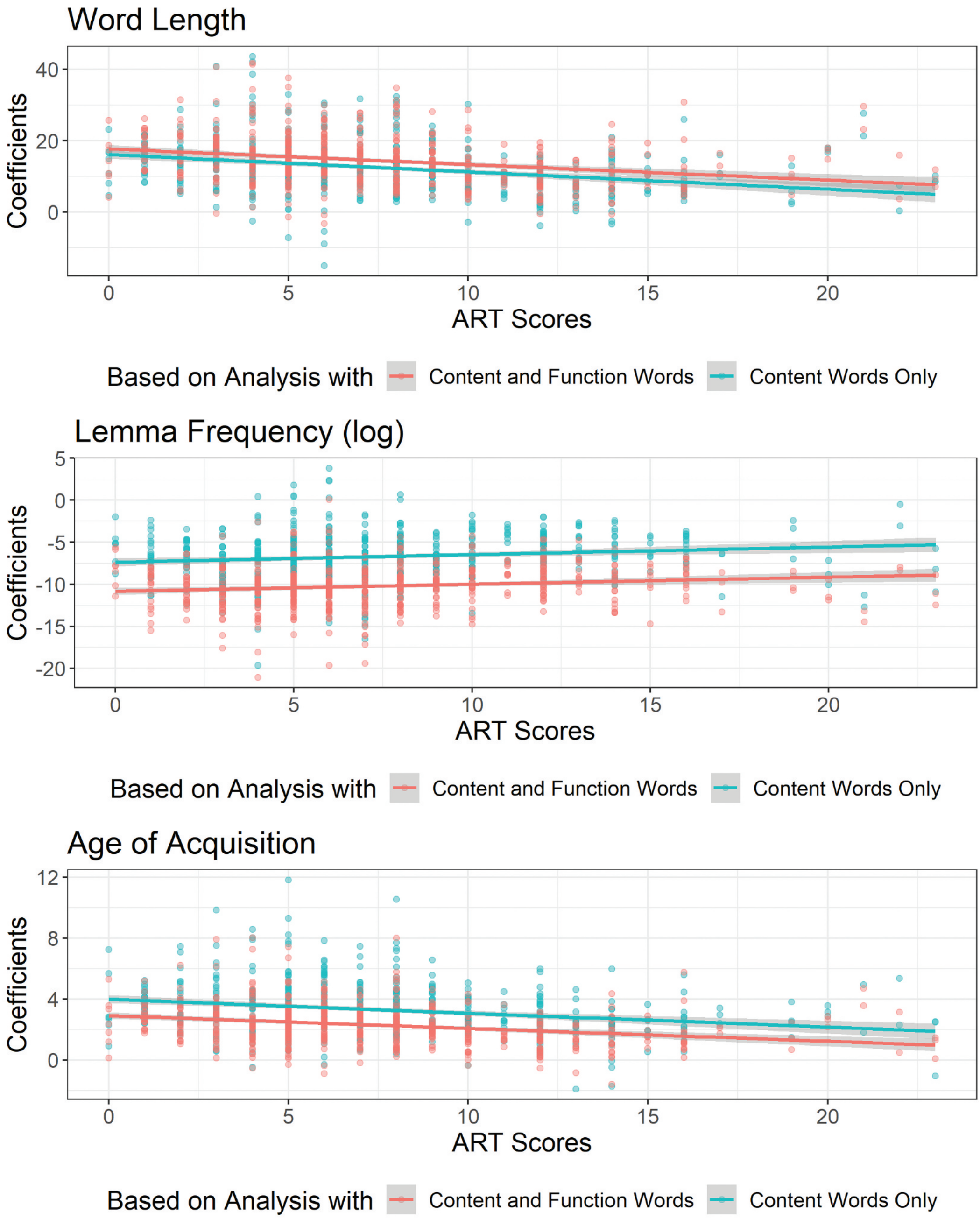

Figure 3. Scatterplots of the relationships between the coefficients of the random slopes of the relationships between gaze duration and word length, gaze duration and lemma frequency, gaze duration and age of acquisition, and ART scores. *Note*. Word characteristics were scaled and centered for analyses. ART scores were scaled and centered for analyses but are presented on their original scale here for interpretability. Gray shading denotes 95% confidence intervals.

Despite VIFs being low, we observed multicollinearity of narrative absorption and liking during initial stages of data exploration, which was also visible in the high correlation between the two measures (see Data Analysis). We therefore fitted separate models to test the effects of narrative

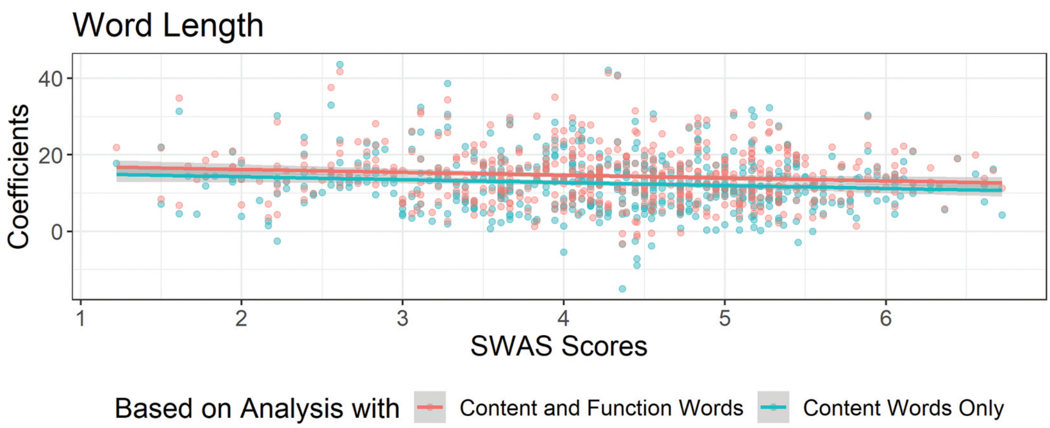

Figure 4. Scatterplot of the relationship between the coefficients of the random slopes of the relationships between gaze duration and word length and SWAS scores. *Note*. Word characteristics were scaled and centered for analyses. SWAS scores were scaled and centered for analyses but are presented on their original scale here for interpretability. Gray shading denotes 95% confidence intervals.

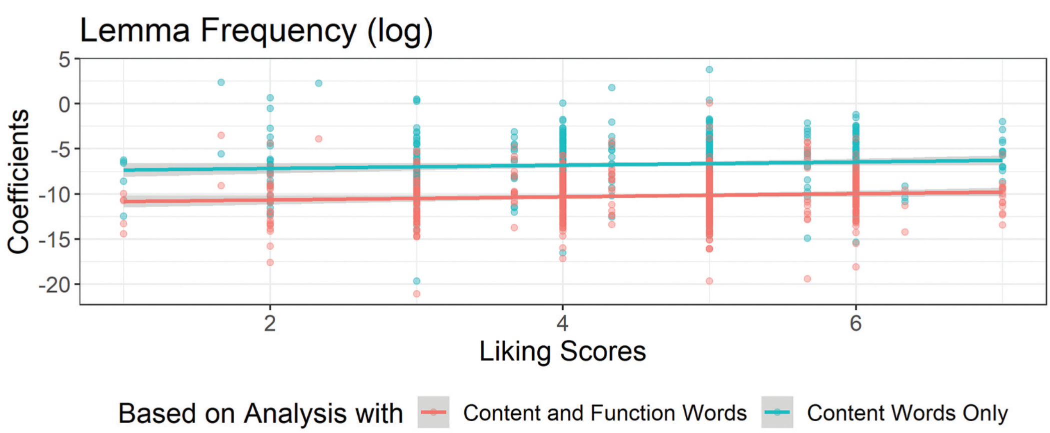

Figure 5. Scatterplot of the relationship between the coefficients of the random slopes of the relationships between gaze duration and lemma frequency and liking scores. *Note*. Word characteristics were scaled and centered for analyses. Liking scores were scaled and centered for analyses but are presented on their original scale here for interpretability. Gray shading denotes 95% confidence intervals.

absorption and liking and set the critical alpha to .025 for these analyses to correct for multiple comparisons.

The results are displayed in Table 5. In both the analysis based on content and function words, as well as the analysis based on content words only, ART scores were significantly associated with the coefficients of the relationship of word length and age of acquisition with gaze duration, whereas the relationship between ART scores and the coefficients of the relationship of lemma frequency with gaze duration was only significant in the analysis based on content words only. In the case of word length and age of acquisition, this was a negative association. That is, word length and age of acquisition were in general associated with slower reading (i.e., more positive coefficients; see Step 1), but ART scores attenuated these associations (coefficients became relatively less positive), meaning that readers with higher print-exposure scores showed a weaker relationship between these word characteristics and gaze duration. The association between ART scores and the coefficients of the relationship of lemma frequency with gaze duration was positive: In general, lemma frequency was associated with faster reading (i.e., more negative coefficients; see Step 1), but ART scores attenuated this association (coefficients became relatively less negative), meaning that readers with higher print exposure showed a weaker relationship between lemma frequency and gaze duration. So, overall, print exposure attenuated the relationship between word characteristics and gaze duration such that high print exposure was associated with decreased sensitivity to word characteristics (see Figure 3).

Like ART scores, SWAS scores were also significantly related to the coefficients of the relationship between word length and gaze duration in the analysis based on content and function words. The same relationship bordered on significance in the analysis with content words only. Again, this was a negative relationship, meaning that narrative absorption attenuated the relationship between word length and gaze duration (coefficients for this association became relatively less positive with an increase in narrative absorption; see Figure 4). That is, readers who were more absorbed in a story were less sensitive to word length.

Story liking was positively related to the coefficients of the relationship between lemma frequency and gaze duration in the analysis based on content and function words. That is, liking attenuated the association between this word characteristic and gaze duration (coefficients for this association became relatively less negative with an increase in liking) such that readers who liked the stories more showed a weaker relationship between lemma frequency and gaze duration (see Figure 5). This relationship was numerically similar but not significant in the analysis based on content words only.

In sum, at least in the case of word length (for narrative absorption) and lemma frequency (for story liking), narrative absorption and liking also seem to attenuate the relationship between word characteristics and gaze duration. Sensitivity to age of acquisition was not related to narrative absorption or liking.

Discussion

In this article, we investigated how low- and high-level processes of reading interact by studying how sensitivity to word characteristics—measured as the association between these characteristics and gaze duration—is related to narrative absorption and story liking. First, both in analyses on content and function words as well as on content words only, we replicated the previous findings that an increase in word frequency (operationalized as lemma frequency) is associated with faster reading, while increases in word length, age of acquisition, and orthographic neighborhood size are associated with slower reading, meaning that more frequent words are read faster, while long words, words learned later in life, and words with more orthographic neighbors are read more slowly (Juhasz & Rayner, 2003, 2006; Pollatsek et al., 1999; Rayner & Duffy, 1986). Moreover, we found that words positioned toward the end of sentences are read faster, perhaps because word prediction and recognition are facilitated by the increased amount of contextual constraints near the end of a sentence. Although this finding seemingly contradicts previous research on sentence wrap-up effects (i.e., sentence-final words being fixated longer; e.g., Rayner et al., 2000, 1989), these results are in line with work by Magliano et al. (1993), who found that gaze durations on clause-final words in a short narrative were shorter than gaze durations on midclause words. Note, however, that our measure of word position does not directly reflect sentence wrap-up effects as word position was operationalized as the absolute numerical position of the word in the sentence. A negative effect of concreteness (i.e., more concrete words being read faster) was only found in the analysis on content words. This is probably due to the fact that concreteness ratings were not available for more than 99% of the function words. Nonetheless, our study demonstrates the robustness of the relationship between word characteristics and reading behavior in the natural context of literary reading.

Based on previous literature on reader and context-induced variability in reading behavior, we then set out to investigate how individual differences in sensitivity to these word characteristics are related to narrative absorption and liking. Our results revealed that first and foremost, individual differences in sensitivity to word characteristics are related to differences in print exposure (as measured by ART scores), which we controlled for in our analyses. Readers with higher print exposure show weaker relationships of word length, age of acquisition, and, to a lesser degree, word frequency with gaze duration (i.e., their reading times are not as affected by word characteristics compared to readers with low print exposure). The finding that print exposure attenuates sensitivity to word characteristics during natural reading was robust across analyses on content and function words, and content words only, and is reminiscent of the literature discussed in the Introduction showing that generally speaking, more skilled readers are less influenced by word characteristics (e.g., Mainz et al., 2017).

However, beyond the effect of print exposure, our results were also suggestive of a negative relationship between sensitivity to word characteristics and the more subjective experiences that arise during reading. Narrative absorption is significantly associated with a decrease in sensitivity to word length, and story liking is significantly associated with a decrease in sensitivity to word frequency. These relationships are less pronounced compared to the associations between sensitivity to word characteristics and print exposure: Although results of the analyses on content words and function words and on contents words only were numerically similar, the relationships between narrative absorption and liking and sensitivity to word characteristics were only significant in the analyses on content and function words. This is most probably due to the fact that there was a larger spread in the coefficients of the model based on content and function words. Nevertheless, our results suggest that readers who experience a more absorbing and pleasant reading experience are less affected by word characteristics.

A possible explanation for these latter findings could be that the readers who were exposed to more print also experienced more narrative absorption and liking as a result of their reading skill. However, note that we controlled for ART scores in all our analyses. Thus, the relationship between sensitivity to word characteristics and narrative absorption and liking in our data goes beyond the relationship between sensitivity to word characteristics and ART scores, which was also found. Moreover, the correlations between ART scores and both SWAS scores and liking scores in our data set show that there is no direct relationship between print exposure and narrative absorption. Additionally, as pointed out in the Introduction, the relationship between print exposure and absorption is not straightforward and seems to be fluctuating based on the text that is being read: Readers with high print exposure do not automatically experience more absorption when reading; they are, however, more likely to experience absorption in highly foregrounded texts (Kuijpers, 2014). Our findings thus seem to suggest that there might be a unique relationship between decreased sensitivity to word characteristics on the one hand and increased narrative absorption and liking on the other hand that cannot solely be explained by print exposure. Alternatively, other underlying third variables related to reading skill or other individual differences such as transportability might play a role.

Our findings seem to be in line with part of Jacobs's (2015) Neurocognitive Poetics Model, which predicts that immersive states are characterized by fluent reading: The more-absorbed readers in our sample were less affected by word characteristics, which could be argued to be an indicator of increased reading fluency. Furthermore, our results both support and extend Dirix et al.'s (2019) studies by showing that not only contextual richness but also the high-level processes that usually arise when reading contextually rich texts, e.g., absorption and liking, are related to variability in reading behavior.

However, even though our results suggest that an absorbed reading experience is accompanied by a shift away from word processing, one of the limitations of our study is that the causal direction of this relationship is still unclear. On the one hand, it could be that an increase in narrative absorption leads to decreased sensitivity to word characteristics. In line with Dirix et al.'s (2019) reasoning, it could be argued that readers who are more engaged with the story can make better use of the context to predict upcoming words and as such rely less on word characteristics, resulting in more fluent reading. The reversed causal direction is also conceivable: In that case, fluent reading, as signaled by a decreased sensitivity to word characteristics, would lead to more narrative absorption. This would align with the idea that a shift away from the immediate environment releases a reader's psychological constraints, such as self-awareness, and can facilitate a transcendent experience, potentially involving meaningful contemplation and reflection (see Oliver et al., 2017; Oliver & Raney, 2011; Slater et al., 2014). These properties of an absorbing experience add to its enjoyable character and ultimately lead to reading pleasure. In more cognitive terms, it could be argued that readers who can afford to spend fewer cognitive resources on word processing can spend more cognitive resources on the construction of a situation model (Kintsch & van Dijk, 1978; Zwaan, 2004) of the narrative. Building a rich situation model is an important aspect of narrative absorption and enjoyment (Busselle & Bilandzic, 2009; Kuijpers et al., 2019).

Moreover, with the current setup it is not possible to discern whether the found relationships between subjective experiences that arise during story reading and sensitivity to word characteristics are mainly driven by differences within readers or between readers. That is, on the one hand, it could be that readers experience different degrees of narrative absorption and liking depending on the story they read and that this is related to different degrees of sensitivity to word characteristics. On the other hand, it could be that readers are relatively consistent in their tendency to become absorbed in and enjoy stories and that this is related to a relatively stable sensitivity to word characteristics.

Both limitations could be addressed in future studies by including both state- (i.e., story-dependent) and trait-based measures of narrative absorption and enjoyment, such as the Tellegen Absorption Scale (Tellegen & Atkinson, 1974), which measures individual differences in the tendency to become absorbed. Such studies could distinguish between the story-dependent "online" effect of narrative absorption on reading behavior and general differences between readers in the degree to which they experience absorption.[6]

Interestingly, and perhaps counterintuitively, our findings resonate with accounts from the literature on mindless reading and mind wandering. Mindless reading is a form of reading during which the eyes keep moving across the text while the mind is wandering (Rayner & Fischer, 1996; Reichle et al., 2010; Schad et al., 2012). Mind wandering during reading has also been found to lead to a decreased sensitivity to word characteristics (Reichle et al., 2010; Schad et al., 2012; see also Foulsham et al., 2013). These findings have been interpreted as visual and linguistic processing becoming deprioritized to facilitate internally generated spontaneous cognition, such as mental time travel or scene construction (Faber et al., 2018, 2020; Smallwood, 2011). Although these features might be shared between the concepts of narrative absorption and mind wandering, an open question is to what extent they overlap, as mind wandering is usually linked to disengagement rather than engagement with the text (Smallwood & Schooler, 2015). Future research could elucidate to what degree these phenomena rely on similar cognitive and neural processes, possibly by combining eye-tracking measures, self-report measures of narrative absorption, and an online measure that captures a subjective continuous measurement of levels of narrative absorption, such as an "absorption dial." Research into online physiological measures of narrative absorption is still scarce, as it is an experience that is difficult to capture without making subjects aware of or interrupting the experience that should be measured. However, our results put us one step closer to the development of such online measures and to testing their usability, as we now know that decreased sensitivity to word characteristics could be an additional objective indicator of absorbed reading.[7]

In conclusion, our study found that narrative absorption and liking are associated with a decreased sensitivity to word characteristics, suggesting that low- and high-level aspects of reading interact and compete for cognitive resources. As Radach et al. (2008) note, research on reading has long been characterized by a divide in studies looking at either low-level processes such as word processing or high-level processes such as comprehension, engagement, or enjoyment. Here we have shown that not only reading skill but also differences in the degree to which readers engage with a text can lead to differences in reading behavior in lower levels of linguistic processing. Given the importance of such high-level processes as narrative absorption, persuasion, and mental imagery in natural reading contexts, we believe our findings stress the need for these two "research cultures" to join forces to map out the interplay between the technical and subjective experiences of reading. After all, if we want to understand reading behavior as it occurs in daily life, we need to acknowledge that a successful and pleasurable reading experience goes beyond the efficient processing of individual words.

Notes

1. For variability in reading behavior in the context of reading development, see, e.g., Davies et al. (2017); Kliegl et al. (2004); Leinenger and Rayner (2017).
2. Excluding the data from this study from the final data set did not alter the pattern of results presented here.
3. These were mostly complex compounds such as *chipszakgeritsel* ‘the rustling sound made by a bag of crisps’ or diminutives such as *koelkastmagneetje* ‘small fridge magnet’.
4. Note that our research question could in principle also be answered in a single model by including interaction terms for the interactions between the word characteristics and SWAS, liking, and ART scores. However, to avoid issues of nonconvergence and interpretability related to the inclusion of the required high number of complex interaction terms, we decided to use the current approach.
5. Note that due to the partial pooling of data (the incorporation of many subjects’ data into one hierarchical model that maintains the nested structure of gaze duration within subjects), by-subject estimates of slope coefficients are shrunk toward the mean, and their magnitudes are therefore less extreme than they would be with no pooling (e.g., fitting a separate multiple regression model to each subject’s data). Such shrinkage prevents overfitting, making the coefficients estimated from the hierarchical model more appropriate for our purposes than coefficients from separate regression models (see Gelman & Hill, 2006).
6. We kindly thank Tobias Richter and an anonymous reviewer for this suggestion.
7. We kindly thank Kiel Christianson for this suggestion.

Acknowledgments

We would like to thank Kiel Christianson and one anonymous reviewer for their feedback on an earlier version of this manuscript. We would also like to thank Tobias Richter, Johanna Kaakinen, and two anonymous reviewers for their comments on this article.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) under Vidi Grant [276-89-007] awarded to Roel Willems and Veni Grant [VI.Veni.191G.001] awarded to Myrthe Faber.

ORCID

Lynn S. Eekhof http://orcid.org/0000-0001-6779-2370

Data availability

Analysis scripts and data are available on the OSF https://osf.io/9zkx3/.

References

Abrams, M. H., & Harpham, G. G. (2009). *A glossary of literary terms* (9th ed.). Wadsworth Cengage Learning.

Appel, R. (2003, June 16). *De Invaller. NRC Handelsblad*. https://www.nrc.nl/nieuws/2003/06/16/de-invaller-7642950-a1353672

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology*, *58*(6), 1065–1086. https://doi.org/10.1080/02724980443000476

Bálint, K., Hakemulder, F., Kuijpers, M. M., Doicaru, M., & Tan, E. S. (2016). Reconceptualizing foregrounding: Identifying response strategies to deviation in absorbing narratives. *Scientific Study of Literature*, *6*(2), 176–207. https://doi.org/10.1075/ssol.6.2.02bal

Barton, J. J. S., Hanif, H. M., Björnström, L. E., & Hills, C. (2014). The word-length effect in reading: A review. *Cognitive Neuropsychology*, *31*(5–6), 378–412. https://doi.org/10.1080/02643294.2014.895314

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica*, *150*, 80–84. https://doi.org/10.1016/j.actpsy.2014.04.010

Busselle, R., & Bilandzic, H. (2009). Measuring narrative engagement. *Media Psychology*, *12*(4), 321–347. https://doi.org/10.1080/15213260903287259

Butler, B., & Hains, S. (1979). Individual differences in word recognition latency. *Memory & Cognition*, *7*(2), 68–76. https://doi.org/10.3758/BF03197587

Carleton, R. N., Abrams, M. P., & Asmundson, G. J. G. (2010). The attentional resource allocation scale (ARAS): psychometric properties of a composite measure for dissociation and absorption: Research article: Attentional resource allocation scale (ARAS). *Depression and Anxiety*, *27*(8), 775–786. https://doi.org/10.1002/da.20656

Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition*, *28*(1), 143–153. https://doi.org/10.3758/BF03211582

Davies, R. A. I., Arnell, R., Birchenough, J. M. H., Grimmond, D., & Houlson, S. (2017). Reading through the life span: Individual differences in psycholinguistic effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(8), 1298–1338. https://doi.org/10.1037/xlm0000366

Dirix, N., Brysbaert, M., & Duyck, W. (2019). How well do word recognition measures correlate? Effects of language context and repeated presentations. *Behavior Research Methods*, *51*(6), 2800–2816. https://doi.org/10.3758/s13428-018-1158-9

Eekhof, L. S., Eerland, A., & Willems, R. M. (2018). Readers' insensitivity to tense revealed: No differences in mental simulation during reading of present and past tense stories. *Collabra: Psychology*, *4*(1), 16. https://doi.org/10.1525/collabra.121

Faber, M., Bixler, R., & D'Mello, S. K. (2018). An automated behavioral measure of mind wandering during computerized reading. *Behavior Research Methods*, *50*(1), 134–150. https://doi.org/10.3758/s13428-017-0857-y

Faber, M., Krasich, K., Bixler, R. E., Brockmole, J. R., & D'Mello, S. K. (2020). The eye–mind wandering link: Identifying gaze indices of mind wandering across tasks. *Journal of Experimental Psychology: Human Perception and Performance, 46*(10), 1201–1221. https://doi.org/10.1037/xhp0000743

Foulsham, T., Farley, J., & Kingstone, A. (2013). Mind wandering in sentence reading: Decoupling the link between mind and eye. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale*, *67*(1), 51–59. https://doi.org/10.1037/a0030217

Gao, X., Dera, J., Nijhof, A. D., & Willems, R. M. (2019). Is less readable liked better? The case of font readability in poetry appreciation. *PLoS ONE*, *14*(12), e0225757. https://doi.org/10.1371/journal.pone.0225757

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790942

Gerrig, R. J. (1993). *Experiencing narrative worlds: On the psychological activities of reading*. Routledge. https://doi.org/10.4324/9780429500633

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*(5), 701–721. https://doi.org/10.1037/0022-3514.79.5.701

Hakemulder, F., Kuijpers, M. M., Tan, E. S., Bálint, K., & Doicaru, M. M. (2017). *Narrative absorption*. John Benjamins Publishing Company.

Hamelink, J. (1966). *Horror vacui*. Polak & Van Gennep.

Jacobs, A. M. (2015). Neurocognitive poetics: Methods and models for investigating the neuronal and cognitive-affective bases of literature reception. *Frontiers in Human Neuroscience*, *9*, 186. https://doi.org/10.3389/fnhum.2015.00186

Jacobs, A. M., & Lüdtke, J. (2017). Immersion into narrative and poetic worlds: A neurocognitive poetics perspective. In F. Hakemulder, M. M. Kuijpers, E. S. Tan, K. Bálint, & M. M. Doicaru (Eds.), *Narrative absorption* (pp. 69–96). John Benjamins Publishing. https://doi.org/10.1075/lal.27.05jac

Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1312–1318. https://doi.org/10.1037/0278-7393.29.6.1312

Juhasz, B. J., & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition*, *13*(7–8), 846–863. https://doi.org/10.1080/13506280544000075

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–650. https://doi.org/10.3758/BRM.42.3.643

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. https://doi.org/10.1037/0033-295X.85.5.363

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1–2), 262–284. https://doi.org/10.1080/09541440340000213

Kliegl, R., & Laubrock, J. (2017). Eye-movement tracking during reading. In A. M. B. De Groot & P. Hagoort (Eds.), *Research methods in psycholinguistics and the neurobiology of language: A practical guide* (pp. 68–88). Wiley.

Koopman, E. M. (2015). Empathic reactions after reading: The role of genre, personal factors and affective responses. *Poetics*, *50*, 62–79. https://doi.org/10.1016/j.poetic.2015.02.008

Kuijpers, M. M. (2014). *Absorbing stories: The effects of textual devices on absorption and evaluative responses*. Utrecht University. https://dspace.library.uu.nl/handle/1874/304849

Kuijpers, M. M., Douglas, S., & Kuiken, D. (2019). Personality traits and reading habits that predict absorbed narrative fiction reading. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(1), 74–88. https://doi.org/10.1037/aca0000168

Kuijpers, M. M., Hakemulder, F., Tan, E. S., & Doicaru, M. M. (2014). Exploring absorbing reading experiences: Developing and validating a self-report scale to measure story world absorption. *Scientific Study of Literature*, *4*(1), 89–122. https://doi.org/10.1075/ssol.4.1.05kui

Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, *65*(1), 42–73. https://doi.org/10.1016/j.jml.2011.03.002

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lammers, J. (2017, January 10). *500 à 1000: Hoe de wolven dansen*. *De Revisor*. http://www.revisor.nl/entry/2363/hoe-de-wolven-dansen

Leinenger, M., & Rayner, K. (2017). What we know about skilled, beginning, and older readers from monitoring their eye movements: Implications for teaching reading. In J. A. León & I. Escudero (Eds.), *Reading Comprehension In Educational Settings* (pp. 1–27). John Benjamins Publishing. https://doi.org/10.1075/swll.16.01lei

Magliano, J. P., Graesser, A. C., Eymard, L. A., Haberlandt, K., & Gholson, B. (1993). Locus of interpretive and inference processes during text comprehension: A comparison of gaze durations and word reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 704–709. https://doi.org/10.1037/0278-7393.19.3.704

Mainz, N., Shao, Z., Brysbaert, M., & Meyer, A. S. (2017). Vocabulary knowledge predicts lexical processing: Evidence from a group of participants with diverse educational backgrounds. *Frontiers in Psychology*, *8*, 1164. https://doi.org/10.3389/fpsyg.2017.01164

Mak, M., & Willems, R. M. (2018). Mental simulation during literary reading: Individual differences revealed with eye-tracking. *Language, Cognition and Neuroscience*, *34*(4), 511–535. https://doi.org/10.1080/23273798.2018.1552007

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS ONE*, *7*(8), e43230. https://doi.org/10.1371/journal.pone.0043230

Melchior, W. (1992). *De roeping van het vlees*. Atlas Contact.

Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Kuehnast, M., & Jacobsen, T. (2015). Towards a psychological construct of being moved. *PLoS ONE*, *10*(6), e0128451. https://doi.org/10.1371/journal.pone.0128451

Nabokov, V. (2003). *Een Russische Schoonheid 1*. De Bezige Bij.

Nell, V. (1988). *Lost in a book: The psychology of reading for pleasure*. Yale University Press.

Oliver, M. B., Ferchaud, A., Yang, C., Huang, Y., & Bailey, E. (2017). Absorption and meaningfulness: Examining the relationship between eudaimonic media use and engagement. In F. Hakemulder, M. M. Kuijpers, E. S. Tan, K. Bálint, & M. M. Doicaru (Eds.), *Linguistic approaches to literature* (pp. 253–269). John Benjamins Publishing Company. https://doi.org/10.1075/lal.27.13oli

Oliver, M. B., & Raney, A. A. (2011). Entertainment as pleasurable and meaningful: Identifying hedonic and eudaimonic motivations for entertainment consumption. *Journal of Communication*, *61*(5), 984–1004. https://doi.org/10.1111/j.1460-2466.2011.01585.x

Orquin, J. L., & Holmqvist, K. (2018). Threats to the validity of eye-movement research in psychology. *Behavior Research Methods*, *50*(4), 1645–1656. https://doi.org/10.3758/s13428-017-0998-z

Perea, M., & Rosa, E. (2000). The effects of orthographic neighborhood in reading and laboratory word identification tasks: A review. *Psicologica*, *21*(2), 327–340.

Perfetti, C., Goldman, S. R., & Hogaboam, T. W. (1979). Reading skill and the identification of words in discourse context. *Memory & Cognition*, *7*(4), 273–282. https://doi.org/10.3758/BF03197600

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687

Pollatsek, A., Perea, M., & Binder, K. S. (1999). The effects of "neighborhood size" in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(4), 1142–1158. https://doi.org/10.1037/0096-1523.25.4.1142

Pratt, M. L. (1981). The short story: The long and the short of it. *Poetics*, *10*(2), 175–194. https://doi.org/10.1016/0304-422X(81)90033-4

R Core Team. (2017) . *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological Research*, *72*(6), 675–688. https://doi.org/10.1007/s00426-008-0173-3

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201. https://doi.org/10.3758/BF03197692

Rayner, K., & Fischer, M. H. (1996). Mindless reading revisited: Eye movements during reading and scanning are different. *Perception & Psychophysics*, *58*(5), 734–747. https://doi.org/10.3758/BF03213106

Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology Section A*, *53*(4), 1061–1080. https://doi.org/10.1080/713755934

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C., Jr. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, *4*(3–4), SI21–SI49. https://doi.org/10.1080/01690968908406362

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, *8*(4), 364–382. https://doi.org/10.1207/s15327957pspr0804_3

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological Science*, *21*(9), 1300–1310. https://doi.org/10.1177/0956797610378686

R-hack/mer-utils.R. (2014, May 20). https://github.com/aufrank/R-hacks/blob/master/mer-utils.R

Sanford, A. J., & Emmott, C. (2012). *Mind, brain and narrative*. Cambridge University Press.

Schad, D. J., Nuthmann, A., & Engbert, R. (2012). Your mind wanders weakly, your mind wanders deeply: Objective measures reveal mindless reading at different levels. *Cognition*, *125*(2), 179–194. https://doi.org/10.1016/j.cognition.2012.07.004

Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition*, *26*(6), 1270–1281. https://doi.org/10.3758/BF03201199

Schindler, I., Hosoya, G., Menninghaus, W., Beermann, U., Wagner, V., Eid, M., & Scherer, K. R. (2017). Measuring aesthetic emotions: A review of the literature and a new assessment tool. *PLoS ONE*, *12*(6), e0178899. https://doi.org/10.1371/journal.pone.0178899

Sears, C. R., Siakaluk, P. D., Chow, V. C., & Buchanan, L. (2008). Is there an effect of print exposure on the word frequency effect and the neighborhood size effect? *Journal of Psycholinguistic Research*, *37*(4), 269–291. https://doi.org/10.1007/s10936-008-9071-5

Slater, M. D., Johnson, B. K., Cohen, J., Comello, M. L. G., & Ewoldsen, D. R. (2014). Temporarily expanding the boundaries of the self: Motivations for entering the story world and implications for narrative effects: Temporarily expanded boundaries of the self. *Journal of Communication*, *64*(3), 439–455. https://doi.org/10.1111/jcom.12100

Smallwood, J. (2011). Mind-wandering while reading: Attentional decoupling, mindless reading and the cascade model of inattention: mind-wandering while reading. *Language and Linguistics Compass*, *5*(2), 63–77. https://doi.org/10.1111/j.1749-818X.2010.00263.x

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology*, *66*(1), 487–518. https://doi.org/10.1146/annurev-psych-010814-015331

Spinelli, D., De Luca, M., Di Filippo, G., Mancini, M., Martelli, M., & Zoccolotti, P. (2005). Length effect in word naming in reading: Role of reading experience and reading deficit in Italian readers. *Developmental Neuropsychology*, *27*(2), 217–235. https://doi.org/10.1207/s15326942dn2702_2

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, *24*(4), 402–433. https://doi.org/10.2307/747605

Tainturier, M. J., Tremblay, M., & Lecours, A. (1992). Educational level and the word frequency effect: A lexical decision investigation. *Brain and Language*, *43*(3), 460–474. https://doi.org/10.1016/0093-934X(92)90112-R

Tan, E. S. (1996). *Emotion and the structure of narrative film: Film as an emotion machine*. Routledge. https://doi.org/10.4324/9780203812761

Tellegen, A., & Atkinson, G. (1974). Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology*, *83*(3), 268–277. https://doi.org/10.1037/h0036681

Thissen, B. A. K., Menninghaus, W., & Schlotz, W. (2018). Measuring optimal reading experiences: The reading flow short scale. *Frontiers in Psychology*, *9*, 2542. https://doi.org/10.3389/fpsyg.2018.02542

Van den Hoven, E., Hartung, F., Burke, M., & Willems, R. (2016). Individual differences in sensitivity to style during literary reading: Insights from eye-tracking. *Collabra: Psychology*, *2*(1), 25. https://doi.org/10.1525/collabra.39

Van Eeden, E. (2015). Ze is overal. In A. L. Snijders (Ed.), *Nederland leest de mooiste korte verhalen (Utrecht editie)*. Stichting Collectieve Propaganda van het Nederlandse Boek

Van Essen, R. (2014). *Hier wonen ook mensen*. Atlas Contact.

Van Hassel, S. (2012). *Ezels*. De Bezige Bij.

Vaughn, L. A., Childs, K. E., Maschinski, C., Paul Niño, N., & Ellsworth, R. (2010). Regulatory fit, processing fluency, and narrative persuasion: Regulatory fit, fluency, and narrative persuasion. *Social and Personality Psychology Compass*, *4*(12), 1181–1192. https://doi.org/10.1111/j.1751-9004.2010.00325.x

Verstegen, J. (1973). *De nieuwe vrijheid*. AWB.

von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language*, *94*, 119–133. https://doi.org/10.1016/j.jml.2016.10.003

Yap, M. J., Tse, C.-S., & Balota, D. A. (2009). Individual differences in the joint effects of semantic priming and word frequency revealed by RT distributional analyses: The role of lexical integrity. *Journal of Memory and Language*, *61*(3), 303–325. https://doi.org/10.1016/j.jml.2009.07.001

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *Psychology of learning and motivation* (pp. 35–62). Elsevierhttps://doi.org/10.1016/S0079-7421(03)44002-4