ORIGINAL RESEARCH REPORT

Individual Differences in Sensitivity to Style During Literary Reading: Insights from Eye-Tracking

Emiel van den Hoven*, Franziska Hartung*, Michael Burke† and Roel M. Willems*,‡,§

Style is an important aspect of literature, and stylistic deviations are sometimes labeled foregrounded, since their manner of expression deviates from the stylistic default. Russian Formalists have claimed that foregrounding increases processing demands and therefore causes slower reading – an effect called retardation. We tested this claim experimentally by having participants read short literary stories while measuring their eye movements. Our results confirm that readers indeed read slower and make more regressions towards foregrounded passages as compared to passages that are not foregrounded. A closer look, however, reveals significant individual differences in sensitivity to foregrounding. Some readers in fact do not slow down at all when reading foregrounded passages. The slowing down effect for literariness was related to a slowing down effect for high perplexity (unexpected) words: those readers who slowed down more during literary passages also slowed down more during high perplexity words, even though no correlation between literariness and perplexity existed in the stories. We conclude that individual differences play a major role in processing of literary texts and argue for accounts of literary reading that focus on the interplay between reader and text.

Introduction

Literary reading can be distinguished from other types of reading in a number of aspects. Some differences can be attributed to characteristics of the text, such as the frequent and systematic use of rhetorical devices, whereas the reader and the reading context play an important role as well. How and to what extent those factors influence the literariness of the reading experience has been a matter of debate. According to the *text-oriented perspective* (e.g., [1]), text features are independent of readers and can be more or less literary. The *reader-oriented perspective* (e.g., [2]), however, claims that the (perceived) literariness depends on a reader's attention to certain aspects of the text. Interactional approaches emphasize that an author can manipulate text characteristics so that the text fulfills certain necessary conditions of being literary, but the reader also needs to react in a certain way to those manipulations for the literary experience to emerge (e.g., [3–8]).

An important characteristic of literary reading is *foregrounding.* The term *foregrounding* is a translation by Garvin [9] of the Czech term *aktualisace, actualization* in English. The term refers to words, expressions or structures that *stand out* from their textual context, because they deviate stylistically in one or more features from the text. It is assumed that foregrounding causes readers to shift their attention from the content to the style of a text [10]. There has been much speculation about the effects foregrounding may have on the reader and the reading process. Mukařovský argued that foregrounded structures cause *de-automatization* of reading, which means that the text structure is processed less automatically.

Shklovsky [11] referred, much earlier, to the same process as *defamiliarization.* He explicitly links defamiliarization to aesthetic appreciation: In order for aesthetic appreciation to emerge, the time it takes for the process of perception to be completed must be prolonged. The slowing down to foregrounded passage is sometimes called *retardation* [11]. It is important to note that Shklovsky does not claim that aesthetic appreciation itself causes the increase in processing time. Rather, the longer processing times that result from increased difficulty allow for aesthetic experience to arise. Relating this idea to characteristics of the reader, we expect that experience with reading

* Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

† University College Roosevelt, Utrecht University, The Netherlands

‡ Centre for Language Studies, Radboud University, Erasmusplein 1, Nijmegen, The Netherlands

§ Donders Institute for Brain, Cognition and Behaviour, Radboud University, Kapittelweg 29, 6525 EN Nijmegen, The Netherlands

Corresponding author: Emiel van den Hoven (emiel.vandenhoven@frequenz.uni-freiburg.de)

can influence the effects foregrounding has on reading. More experienced readers are expected to experience less problems with texts high on foregrounding because they are more used to language that deviates from the stylistic norm. Retardation therefore may depend on individual experience with reading.

Empirical research has shown that foregrounding generally influences aesthetic appreciation (e.g., [12]), as well as reading times. But the effect of foregrounding has been shown to be subject to individual differences as well. A landmark study by Miall and Kuiken [13–16] confirmed that foregrounded passages are read more slowly than passages that are not foregrounded. Importantly, they found that a reader's level of experience influences to which type of foregrounding (phonetic versus semantic) they are sensitive.

The main goal of the current study is to directly test the hypothesis posed by the Russian Formalists (e.g., [11]) who proposed that readers slow down during the reading of literary passages, compared to non-literary passages. While several reading times studies have investigated this issue before, we extend the empirical literature on this matter by measuring word by word reading times with eye-tracking, and by an explicit focus on individual differences in our analysis. Foregrounded passages are expected to decrease reading speed in the majority of readers, as compared to non-foregrounded passages.

Using eye-tracking, a decrease in reading speed can be measured in numerous ways, but because the chance of a Type I error increases with using multiple dependent measures, we will restrict the analysis to two representative measures, namely *gaze duration* and *chance of regression.* Gaze duration is the total fixation time on a word during the first time a word is fixated [17]. So when a word is consecutively fixated multiple times, or when it is fixated only once, gaze duration consists of the sum of the fixation times. Out of the available reading time metrics, we consider gaze duration the best candidate because (i) it takes into account all fixations on a word during the first pass, rather than just the first fixation; (ii) it takes all words into account rather than just the ones that have been fixated only once; and (iii) it allows for the distinction between progressive fixations and regressions. The chance of regression (henceforth simply *regressions*) is based on a simple binary measure that indicates whether or not a reader fixates on a word w_i after having fixated on any of the words $w_{i+1} \ldots w_n$. This measure represents cognitive difficulty experienced with a word only after the word has either been fixated or skipped.

Since foregrounding draws attention to the wording of the text rather than its content, it may not only cause readers to slow down, but it may also enhance readers' memory of the surface form of the text. Verbatim memory for a text has often been claimed to be short-lived (e.g., [18–20]), but many studies do find higher-than-chance scores on surprise surface form recognition tests [14, 21, 22]. Moreover, when an element that is in focus (through, e.g., syntactic devices like cleft structures, or through the use of italics) is changed in between two text readings, the change is more likely to be noticed than when an element that is not in focus is changed [23, 24]. Therefore, it seems likely that the increase in attention to surface form caused by foregrounding results in improved recognition of the text's surface form.

The above considerations lead us to the following hypotheses:

H1. *Gaze durations and regressions are increased for words that are foregrounded compared to words that are not foregrounded.*

Differences between individuals are expected to play a role in the effect of foregrounding on reading behavior. We expect that previous exposure to literary language is a crucial factor influencing reading behavior. Frequent readers have more experience with foregrounding than infrequent readers, and are less likely to slow down when reading foregrounded parts of a text, compared to infrequent readers:

H2. *Infrequent fiction readers show a larger retardation effect when exposed to foregrounding than frequent fiction readers.*

The increased focus on style caused by foregrounding may have additional effects, besides affecting reading behavior. If readers pay more attention to the surface structure during foregrounded passages, then this increased attention is likely to result in enhanced memory for the surface form. This leads to the following hypothesis:

H3. *The memory for the surface form is better for foregrounded passages than for non-foregrounded passages.*

Literary fiction, as compared to non-narrative texts, often causes the reader to get immersed into the story and construct multimodal situation models [25]. Being immersed or transported in(to) a story is linked to mental simulation (e.g. [7, 8, 26, 27], and defined as "the state of feeling cognitively, emotionally, and imaginally immersed in a narrative world" ([28], see also [29, 30], see also [6] on *disportation*). Immersion is associated with enjoyment [31, 32], meaning that the more we engage with a story, the more we enjoy it. In order to take effects of immersion into account, we added a self-report measure of transportation into a narrative and *story liking* as exploratory factors hypothesized to be related to foregrounding. It is possible that the degree of immersion and story liking affect readers' reactions to foregrounding. For instance, readers who are more immersed, or like the story more, may pay less attention to the style of the text, and thus be less affected by foregrounding. Alternatively, immersed readers, or those that like the story more, may attribute higher importance to the text, which could facilitate an intention for thorough understanding, and hence result in longer reading times for foregrounded passages. The hypotheses are therefore unspecified with regard to the direction of the effect:

H4a. *Readers who are more immersed in the story react differently to foregrounded words than readers who are less immersed in the story.*

H4b. *Readers who like the story more react differently to foregrounded words than readers who like the story less.*

To test our hypotheses, we had thirty participants read three short stories from Dutch literature, while measuring their eye-movements. After reading the stories, participants filled in a questionnaire measuring how strongly they were immersed in each story, scored how much they liked each story, and filled out questions about personal reading habits, and a multiple choice test that measured the recognition performance of the surface structure of sentences in the stories.

Methods
Participants
Thirty healthy participants (25 females) without language impairments were recruited from the participant database of the Max Planck Institute for Psycholinguistics. Age ranged from 18 to 28 years (M = 21.90, SD = 2.63). The participants' native language was Dutch. Participants had normal (N = 17), or corrected to normal visual acuity with glasses (N = 4) or contact lenses (N = 9). Near vision was tested with a near vision test. None of the participants made an error when character size was above 1mm at a distance of 40cm, or 0.14° of the visual field (cf. the characters in the experiment, which were 3.3mm at a distance of 97cm, or 0.18° of the visual field). None of the participants studied literature science.

Materials
Three short stories from Dutch literature were selected (see **Table 1**). The selection was made on the basis of the length of the stories, as well as their potential to be of interest to the target group. All participants read all three stories (in randomized order). Story 1 had been read before by two of the participants and Story 3 had been read before by three of the participants. None of the participants had read Story 2 before. Because there were so few second time story-readings, those data were not excluded from the analysis. (Separate analyses were conducted in which second-time readings were excluded, which led to qualitatively similar results for all models).

Pretest
16 participants from the same participant database who did not take part in any other parts of the study read each of the stories twice. The first time, they were instructed to read the story as they normally would. For the second reading, they were instructed to underline all the words, sentences and passages that they considered to be "literary". We will use the terms *foregrounding* and *literariness* interchangeably here. The two terms are not wholly synonymous, but our empirical operationalization of literariness ensures that only those literary passages that capture attention are included in the measure, whereas passages that might be considered literary exactly because they do not draw attention to themselves (i.e., they are *backgrounded*) are not included, since they are by definition unlikely to be noticed by our participants. In line with the idea that secondary processing leads to better appreciation of the qualities of literary texts (e.g., [12]), the instructions should lead to an adequate measure of the intersubjective perception of literariness. This pretest resulted in a literariness-score between 0 and 16 for every word in each of the three stories: a score of 0 if none of the participants had underlined it and a score of 16 if every participant had underlined it. On average, participants underlined 1106.19 words (SD = 573.36, ranging from 37 to 1989 words) in all three stories combined. That amounts to an average of 12.36% (SD = 6.64%) of the total amount of words. A repeated measures ANOVA showed that there were no statistically significant differences between stories in the percentage of words that was underlined, $F(2,30)$ = 1.18, p = 0.32. Note that this operationalization of literariness does not to allow to distinguish between different types of foregrounding/literariness (e.g. phonological or semantic foregrounding).

As shown in **Figure 1**, there was a high autocorrelation between the scoring of a word w_i and the scoring of the word that immediately followed it w_{i+1}. The size of the correlation gradually decreased as the distance between the words increased. This reflects the fact that participants mainly underscored sentences and passages rather than single words.

Participants did not show perfect consensus on what they considered the beginning and end of a literary passage. For instance, almost all participants agreed that the passage "Een zonnige metalen waterdruppel" ("A sunny metal water droplet") was literary. However, some participants also included the preceding parts of the text, whereas others did not. **Figure 2** visualizes this gradual change from low to high literariness scores. The size of the characters shows literariness scoring, the bigger characters being underlined more often.

To test whether there was consensus among the participants regarding their literariness judgments (i.e., their judgments of what counts as literary language were not completely idiosyncratic), we estimated a chance

	Title		Author	Year of publication	Word count
Story 1	De Straf	(The Punishment)	Judicus Verstegen [33]	1973	2982
Story 2	Kogeltjes	(Bullets)	Willem Melchior [34]	1992	3526
Story 3	De Vijand	(The Enemy)	Jacques Hamelink [35]	1966	2436

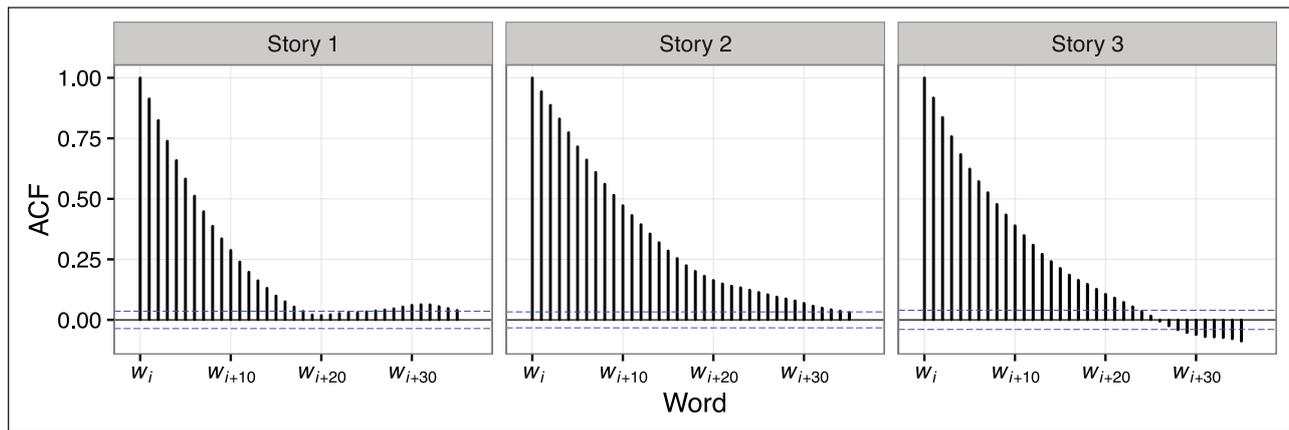**Table 1:** Descriptive information of the experimental stories.

Figure 1: The autocorrelation function (ACF) between underlining scores for word w_i and the words following it, for all three stories. The correlations gradually decrease as the lag increases, reflecting the fact that participants mainly underscored sentences and passages rather than single words. Blue dashed lines indicate 95% CIs of white noise.

distribution by simulating 1000 distributions of underlining scores with the same (stationary) transition probabilities as the data using MCMC sampling. The probability of underlining a word given the underlining of the previous word, $P(\text{underlining}_w | \text{underlining}_{w-1})$, was calculated separately for each participant. A two-sample Kolmogorov-Smirnov test indicated that the underlining scores differed significantly from the estimated chance distribution, $D = 0.94$, $p < .0001$. **Figure 3** shows the density plot for underlining in the three stories combined, compared to the mean percentages from the estimated chance distribution. All participants agreed on the non-literary status of 33.36% of the words, cf. 11.05% in the chance distribution. The percentage of words decreased as agreement upon literariness increased, but not as rapidly as might be expected purely based on chance. Therefore, the perception of literariness in these three stories was partially idiosyncratic, but there was also consensus.

Expert's ratings
As a validation of the non-expert ratings by the participants in the pretest, we compared them to an expert's (one of the authors, MB) rating of foregrounding in the stories. The expert performed the same task as the participants in the pretest, while being blind to the participants' ratings. The point-biserial correlation between the expert's scorings and the mean of the participants' scorings was significant, $r = .39$, $p < .0001$, thus corroborating our operationalization of foregrounding.

Additional measures
The participants filled in an immersion questionnaire after reading each story, and an additional test battery at the end of the experiment. The immersion questionnaire was based on the story world absorption scale (SWAS, [36]), and selected items from the 30-item version of the narrative engagement questionnaire (NEQ) developed by Buselle and Bilandzic [32]. Both questionnaires measure 4 dimensions of story engagement and show considerable overlap. SWAS measures attention, transportation, emotional engagement and mental imagery. NEQ measures narrative understanding, attentional focus, emotional engagement, and narrative presence. Narrative understanding is the only dimension not covered by SWAS, and relevant items from this subscale of the NEQ were included.

Participants were also asked to give a general score for Story Liking on a 10-point response scale (1 = extraordinarily bad, 10 = extraordinarily good). In addition, participants were asked to indicate whether they had read the story before and to answer four simple multiple choice questions regarding story content, specific to each story. The goal of these content questions was merely to ensure

De bezoeker was een vijand. En hij dacht aan een gouden kogel, klein als een erwt en heel rond, en glinsterend. Een zonnige metalen waterdruppel.

'Laat hem los,' zei de vijand.

The visitor was an enemy. And he thought of a golden bullet, small like a pea and very round, and shiny. A sunny metal water droplet.

'Let him go,' said the enemy.

Figure 2: Illustration of literariness scores for an excerpt from Story 3. Larger font sizes represent more underlinings in the literariness pretest. Note that in the main (eye-tracking) experiment, all words were presented in the same font size.
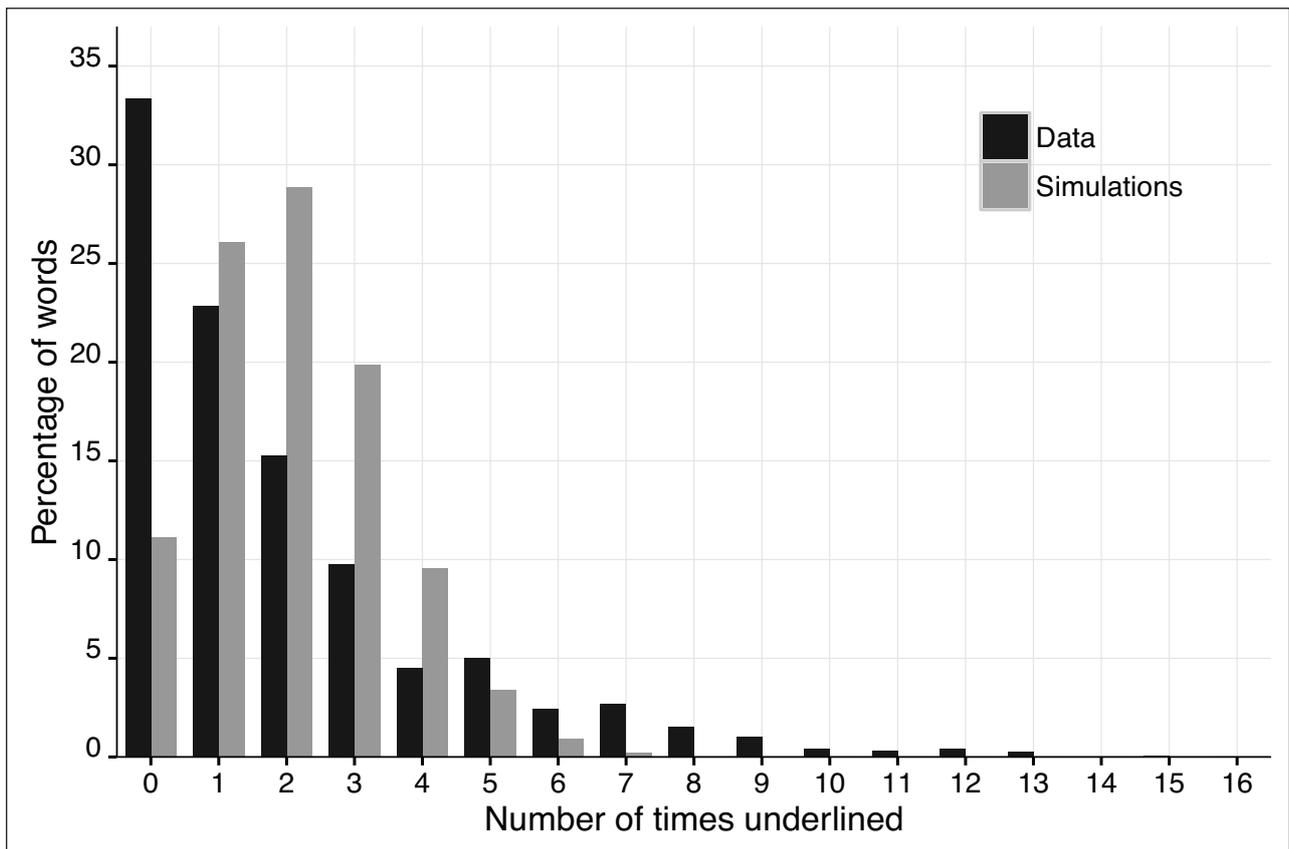
Figure 3: Density of literariness underlining in the pretest. Black bars represent the data; grey bars represent the MCMC simulations.

that the participants processed the stories to a sufficient degree to understand what they were about. Participants who failed to answer minimally two out of the four questions for a story correctly would be excluded from the analysis.

A principal components analysis based on eigenvalues greater than 1 was conducted to extract the underlying factors from the immersion questionnaire. Promax rotation was used (an oblique rotation method), with a kappa of 4. Appendix A shows the final five components (Empathy, Self-loss, Imagery, Compassion and Understanding) and the items that loaded on them. In cases where an item loaded highly on multiple factors, the item was included in the factor with the closest match on a conceptual level. None of the items failed to load highly on any factor. Reliability was sufficient for all components (Cronbach's $\alpha \geq .82$ for all components). The mean score of all items within a factor was taken to represent each participant's score on that factor.

An additional test was conducted to measure recognition of the exact wording of a selection of foregrounded and non-foregrounded passages for each story (the *sentence recognition test*). From each story, three foregrounded and three non-foregrounded passages were selected. Foregrounded passages included words that were both underlined at least 6 times in the pretest and scored as literary by the expert, whereas non-foregrounded passages were underlined neither by the participants nor by the expert. For the sentence recognition test, we generated alternative items for each sentence, which either

diverged from the original formulation in terms of their semantics, their syntax or both their semantics and their syntax. The participants' task was to recognize the original formulation in a multiple-choice test. (See Appendix B for the set of items.)

Finally we measured three aspects of differences in reading behavior. First, participants indicated how much they liked fiction on a response scale ranging from 1 (not at all) to 7 (very much). Second, they indicated how many fiction books they read per year (0; 1–3; 4–6; 7–9; 10–12; more than 12). As a final measure we used a Dutch version [37] of the *Author Recognition Test* (ART). This is an indirect measure of print exposure ([38] updated by [39]). The test assesses the participant's ability to recognize popular authors from a list. The test consists of 30 real author names and 12 foils. Every existing author that was recognized increased the participants' score by one, and every foil that was falsely recognized decreased their score by 1, so that in the end the potential total score on the ART was between -12 and 30.

Apparatus

A monocular tower-mounted EyeLink1000 eye-tracking system with a 25mm lens was used to collect eye-movement data. A head stabilizer minimized head movements. Eye position was recorded with a sampling rate of 1000Hz. Two separate DELL Precision 390 workstations were used for the presentation of the stimuli and data acquisition. Stimuli were presented on a 20″ Acer AL2023 LCD monitor with a refresh rate of 60Hz.

Stimulus presentation

SR Research's Experiment Builder software was used for the presentation of the stimuli. The stories were presented in 28–39 sections of on average 90.4 words (SD = 25.0). The division of the story into sections was kept as closely as possible to the author's original division of the story into paragraphs. The text was presented in the font Calisto MT, 15pt, in black color on a light grey background. The margins were 120 pixels on all sides. Interest areas for the eye-movement data were automatically defined by the Experiment Builder software. Each word corresponded to an interest area, and the limits for the interest areas were centered between adjacent words, leaving no space in between. Interest area margins on all sides of the text were 10 pixels. The monitor was 40.7cm by 30.5cm and the participants were seated at a distance of 97cm from the monitor.

Procedure

Participants were paid €16 as compensation for taking part in the study. Prior to the experiment, participants were informed about the procedure, and about possible contents of the story. The study was approved by the Ethics Committee Social Sciences of Radboud University Nijmegen (Ethics Approval Number ECG2013-1308-120). Participation was voluntary and participants could withdraw at any time without having to state their reasons. All participants gave written informed consent in accordance with the Declaration of Helsinki.

Participants performed an eye dominance test, so that the dominant eye could be tracked. For reasons not reported in the current article, skin conductance response electrodes were attached to the index and middle finger of their non-dominant hand.

The experiment took place in a sound proof cabin. Participants first read a practice story of 428 words, so that they could get used to the experimental setting and the task. They were informed that following each of the stories, they would have to answer questions regarding the content of the story and regarding their experience of reading the story. It was made clear that the content questions were not difficult and could be answered with ease without the need to remember trivial details of the story. Participants were instructed to move as little as possible. There was no time restriction and participants were encouraged to read the stories the way they would read them outside the laboratory.

The stories were presented in random order. A 9-point calibration preceded the beginning of each story. Every five to ten slides, a drift check was performed to make sure that the calibration was still valid. If this was not the case (four times in total), calibration was repeated. Prior to every slide, participants fixated on a fixation cross at the top left of the screen (where the first character of the text would appear) for 1000ms. After every slide, they pressed a button to continue to the next slide.

Eye-movement data preprocessing

Several variables that are known to influence reading times were controlled for in the analysis. We controlled for lexical frequency [40], word length [41], position on the screen, perplexity, orthographic and phonological neighborhood size, age of acquisition, word prevalence, and semantic relation.

The log-transformed lexical frequency per word was taken from the SUBTLEX-NL database [42]. Word length was measured in number of characters and position on the screen was measured as the horizontal distance from the left side of the screen measured in pixels, divided by 100 to make the scales of the measures more homogeneous.

Perplexity is a measure closely related to word surprisal, which indicates how unpredictable an incoming word is [43, 44]. A trigram model was trained to assign probabilities to words given their context in a large corpus of Dutch. Perplexity values for the words in our stories were then calculated by taking 2 to the power of the negative base-2 logarithm of the probability the model assigned to the current word given the preceding context. This means that in the case of high perplexity, the model was very surprised to encounter the word that was just encountered.

Orthographic and phonological neighborhood size information was obtained from the CLEARPOND database ([45], http://clearpond.northwestern.edu/). Age of acquisition norms were obtained from http://crr.ugent.be/archives/1602 [46].

We refer to word prevalence as the log odds of correctly identifying a letter string as a word rather than a non-word. These were obtained from Keuleers, Stevens, Mandera and Brysbaert [47]; http://crr.ugent.be/archives/1494.

All words' semantic relations to the previous content words in the sentence, a measure of semantic priming, were calculated as in Frank and Willems [48]. Semantic vectors were obtained from http://zipf.ugent.be/snaut/ [49].

We controlled for all of these variables by including them as predictors in the mixed effects model, if they were significant.

All fixations were checked to make sure they did not diverge so far from the line being read that they entered a different interest area, and were manually aligned using SR Research's EyeLink Data Viewer. Data for 10 entire story-readings (including all three from one participant) were excluded from the analysis due to inaccuracy of the eye-movement data. Another 74 individual slides were excluded for the same reason, amounting to the exclusion of 11.9% of all slides in total. Fixations on the first word of each slide were also excluded, because they were disproportionately long, reflecting the aftereffects of fixating on the fixation cross prior to each slide. This led to the exclusion of 1.1% of the data. Reading times that deviated more than 3.5 times the standard deviation from the mean were removed from the dataset, as were reading times of 0ms (0.6% in total).

Although the word-related datasets we used to retrieve our predictors from (age of acquisition, orthographic and phonological neighborhood size, prevalence and semantic relation) are cover a substantial part of the Dutch lexicon, not all words in our stories were in all of these datasets. Words for which one of the predictors was lacking were not included in the analysis (22581 data points; 9.6% of the data). For this reason, valence and arousal, two potentially interesting predictors, were not included in the

analysis – these norms existed for only 35.8% of the word tokens in the data.

Two dependent variables were analyzed per word: gaze durations and regressions. Words that were skipped were treated as missing data.

Data analysis

All data were analyzed using the statistical software package R v3.0.2 [50]. A linear mixed model was created to analyze gaze durations, using the *lmer* function from the *lme4* library [51]. First, a model with all fixed effects terms, random intercepts per participant and per story, and random slopes per participant for literariness and perplexity was constructed to predict gaze duration. Models with more elaborate random effects structures did not converge. Subsequently, fixed effects were deleted one by one. If the model fit did not deteriorate after their exclusion (i.e., the likelihood ratio test was not significant), the simpler model was chosen. The *p*-values for individual predictors were likewise determined on the basis of the change in model fit (in χ^2) when the individual predictors were excluded from the model.

In a separate model, the scores for the predictors log frequency, log perplexity, orthographic and phonological neighborhood size, age of acquisition, word prevalence, semantic relation and literariness were taken from the previous word (word w_{i-1}) rather than from word w_i to account for spillover effects [52]. This analysis yielded similar results, but with stronger effects of literariness and semantic relation and less strong effects of frequency and

age of acquisition. It is likely that this was because literariness and semantic relation were highly autocorrelated, so word w_i and word w_{i-1} would have similar values for these predictors, whereas this was not the case for the other predictors. The only reason literariness and semantic relation did better in the spill-over model was probably that less variance was explained by the factors we wanted to control for. We consider the model with values for word w_i to be more valid and we only report this model here (but see the supplementary materials for details of the spill-over model).

Regressions were analyzed using generalized mixed effects logistic regression (from the *lme4* library, [48]), following the same procedure as for the gaze durations. All predictors were *z*-transformed to overcome problems with convergence. The *p*-values for individual predictors resulted from asymptotic Wald tests.

The recognition test data were likewise analyzed using generalized mixed effects logistic regression including random intercepts per participant. Random slopes for the predictors gaze duration and foregrounding were not included due to convergence problems.

Results
Gaze duration
Appendix C shows the step-by-step results of the model selection process for the gaze duration model. Log word frequency, position on the screen, phonological neighborhood size, age of acquisition and word prevalence were all significant predictors of gaze duration. **Figure 4** shows
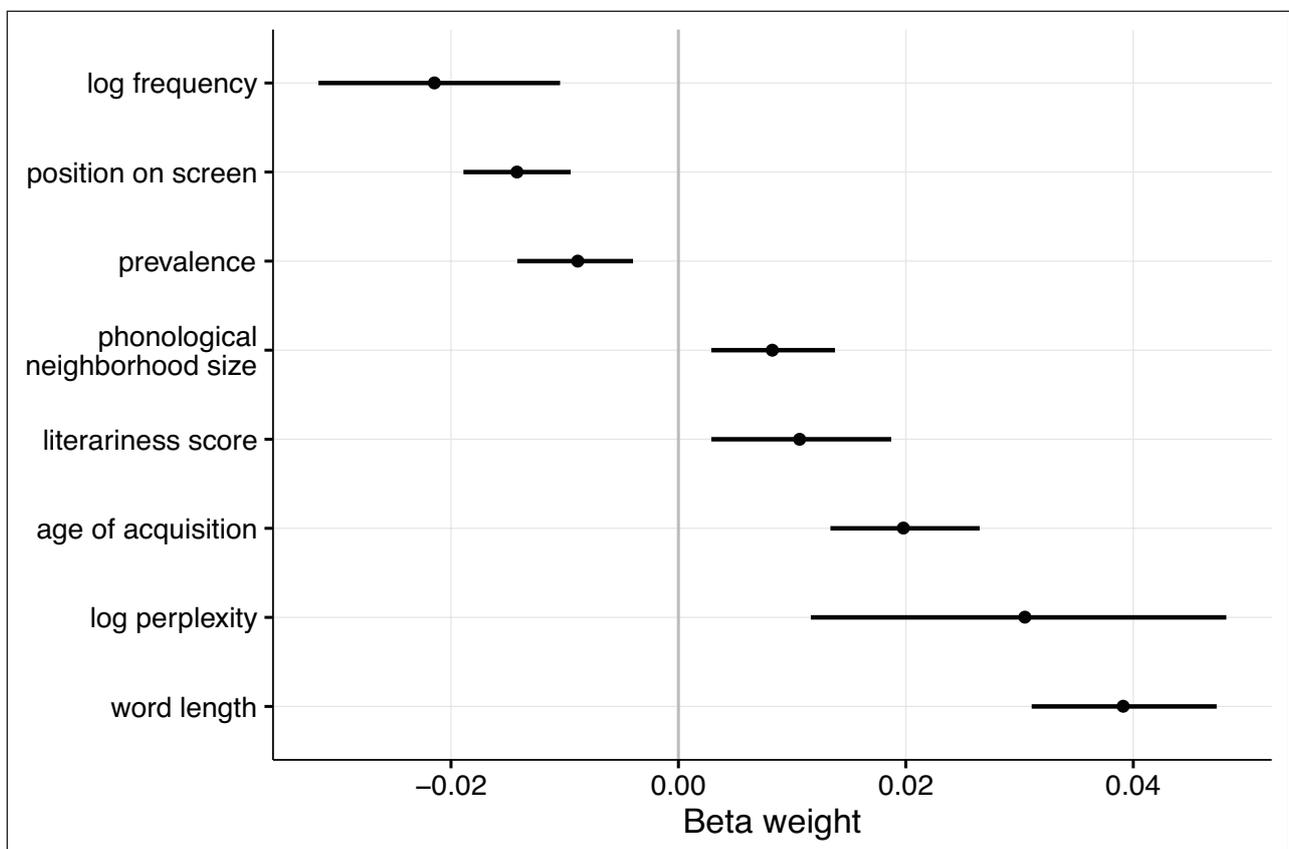

Figure 4: Standardized beta weights of all fixed effects included in the final mixed-effects model. 95% confidence intervals were obtained through bootstrapping.
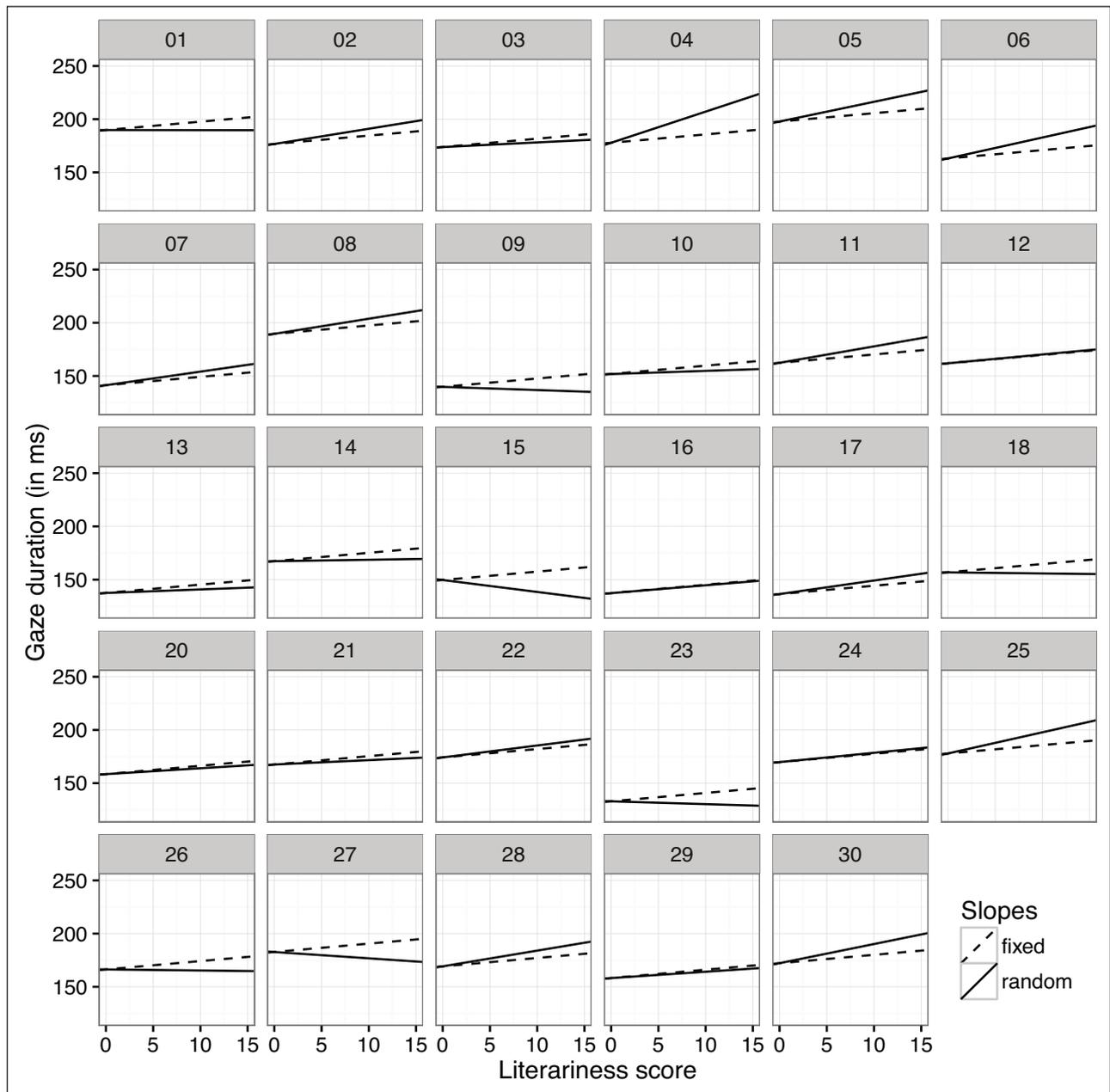
Figure 5: Effect of literariness on gaze durations. Dashed lines are the results with fixed slopes, solid lines when slopes were allowed to vary between participants. The figure illustrates the sizeable individual differences between participants in the effect of foregrounding on gaze durations. Some participants slow down on more foregrounded passages (positive slopes), whereas others speed up (negative slopes).

the standardized beta weights of the fixed effects terms. Literariness as perceived by the participants of the pretest (henceforth simply *literariness*) and log perplexity were also significant predictors of gaze duration. The slopes of these latter two predictors were allowed to vary per participant. **Figure 5** shows the differences between participants in the effect of literariness on gaze duration without random slopes (dashed lines) and when allowing the slopes to vary across participants (solid lines). The results are plotted for each participant separately. The figure shows that there are differences between participants in how they reacted to literary passages: Some slowed down

whereas others sped up their reading when encountering foregrounded passages.

Regressions
Table 2 shows the coefficients for the mixed effects logistic regression model fit to the regression data. Log word frequency, position on the screen, phonological neighborhood size, orthographic neighborhood size, age of acquisition, word prevalence and semantic relation were all significant predictors of the chance of regressing to the word, as were literariness and log perplexity, also when the slopes of these latter predictors were allowed to vary

	ß	SE ß	t	P
Constant	–1.26	0.11	–11.37	
log frequency	0.26	0.17	15.21	<.0001
word length	–0.24	0.13	–18.13	<.0001
Prevalence	–0.19	0.75	–2.48	<.05
screen position	–0.48	0.73	–66.49	<.0001
age of acquisition	0.36	0.93	3.84	<.0001
log perplexity	0.23	0.15	14.93	<.0001
orthographic neighborhood size	–0.27	0.10	–2.62	<.01
phonological neighborhood size	0.23	0.91	2.55	<.05
semantic relation	–0.29	0.88	–3.25	<.01
literariness score	0.43	0.80	5.38	<.0001

Table 2: Coefficients for the mixed effects logistic regression model fit to regressions.
Note: ß indicates the standardized beta.

per participant. The model did not improve after the exclusion of any of the predictors.

The relation between Immersion, Story Liking and the Retardation Effect

The analysis of the gaze duration data resulted in a score for each participant that indicated to which degree literariness affected reading times. This score equaled the slope of the regression line. It could be positive or negative, indicating slowing down or speeding up during reading of foregrounded passages. We will call these scores *Retardation Effects* (though for some participants there was no slowing down but speeding up for the more literary words). In this section we discuss how Retardation Effects related to Immersion and Story Liking. Mean scores and standard deviations for the five factors of the immersion questionnaire that resulted from the factor analysis – Empathy, Self-loss, Imagery, Compassion and Understanding – and Story Liking are presented in **Table 3**.

The scores from the immersion questionnaire were included in a correlation matrix, together with the general score for Story Liking and Retardation Effects. For this correlation analysis Retardation Effects were calculated per participant-story pair instead of per participant, in order to avoid correlating each single Retardation Effect score with three Immersion or Story liking scores. **Table 4**

shows the correlation matrix. None of the factors from the immersion questionnaire correlated significantly with the Retardation Effect (nor did the mean of the subscales, *Overall Immersion*). The score on Story Liking did not correlate significantly with the Retardation Effect either.

The relation between reading experience and the Retardation Effect

Scores on the ART ranged from 0 to 14 (*M* = 8.14, *SD* = 3.59), indicating that most participants were able to recognize several authors from the list. Fiction reading scores ranged from 0 fiction books per year to more than 12. Most participants indicated they read 1–3 books per year. Fiction liking scores ranged from 3 to 7 on the 7-point scale (*M* = 5.59, *SD* = 1.23), indicating that most participants enjoyed reading fiction.

Scores relating to fiction reading were included in a correlation matrix (see **Table 5**), together with the Retardation Effect and random slopes for perplexity. Scores on the ART and amount of fiction books read per year showed a significant positive relationship, corroborating the reliability of the ART as an index of print exposure. The negative correlations between scores on the ART and reading experience on the one hand and the Retardation Effect on the other did not reach significance after bonferroni correction. Fiction liking did not correlate significantly

	Story 1		Story 2		Story 3		Total	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Empathy	4.57	1.21	3.67	1.21	4.46	1.13	4.23	1.24
Self-loss	4.41	1.29	4.78	1.28	4.94	1.11	4.71	1.23
Imagery	5.01	1.30	5.11	1.33	5.33	1.13	5.15	1.25
Compassion	4.97	1.23	4.75	1.26	5.08	1.09	4.93	1.19
Understanding	4.89	0.93	4.30	1.05	4.40	1.22	4.53	1.09
Story Liking	6.41	1.30	6.38	1.59	6.55	1.55	6.45	1.47

Table 3: Descriptive results for the factors from the immersion questionnaire (scale: 1–7) and Story Liking (scale: 1–10).

	Factors immersion questionnaire						
	Empathy	**Self-loss**	**Imagery**	**Compassion**	**Understanding**	**Immersion**	**Story liking**
Self-loss	.60***						
Imagery	.47***	.59***					
Compassion	.66***	.66***	.58***				
Understanding	.58***	.44**	.19	.31			
Immersion	.85***	.85***	.72***	.83***	.64***		
Story Liking	.53***	.54***	.34	.43**	.49***	.60***	
Retardation	.09	.16	.04	.03	–.09	.06	.21

Table 4: Correlation matrix for the factors from the immersion questionnaire, Story Liking and Retardation.
Note: *$p < .05$, **$p < .01$, ***$p < .001$. *p*-values are bonferroni-corrected.

with the Retardation Effect either. There was a significant correlation, however, between the Retardation Effect and random slopes for perplexity. This correlation is illustrated in **Figure 6**.

Recognition
On average, participants recognized 8.3 out of 18 passages correctly ($SD = 2.6$; range = 4–15), significantly higher than chance, $t(29) = 8.23$, $p < .0001$. Passages that deviated both semantically and syntactically from the original were less often falsely recognized than those that deviated from the original in only one dimension, but the latter two did not differ from one another (see **Table 6**).

Inspection of the data in **Table 6** shows a slight preference for correct answers in the foregrounded condition compared to the non-foregrounded condition, but a generalized logistic regression analysis showed that literariness was not a significant predictor of correct recognition, $b = –0.282$, $SD = 0.188$, $p = .134$. Neither literariness, nor the total amount of time spent reading a passage significantly improved the chance of recognizing the correct surface structure of the passage.

Discussion
This study investigated the effects of foregrounding on reading behavior as measured by gaze durations, regressions and recognition of surface structure. Foregrounding was operationalized by having laypeople underscore liter-

	Perplexity	**Fiction reading**	**ART**	**Fiction liking**
Fiction reading	–.41			
ART	–.38	.63**		
Fiction liking	–.07	.63**	.25	
Retardation	.59**	–.37	–.35	.04

Table 5: Correlation matrix for the measures relating to reading experience, ART, fiction reading, fiction liking, random slopes for perplexity and the Retardation Effect.
Note: *$p < .05$, **$p < .01$, ***$p < .001$. *p*-values are bonferroni-corrected.
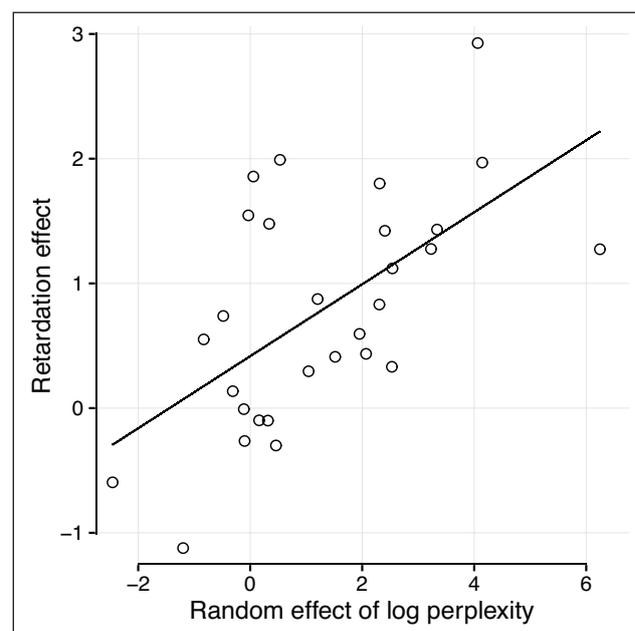
ary passages. Thirty participants read three short stories from Dutch literature, while their eye-movements were recorded. In addition, we measured story immersion, reading behavior and recognition of the exact wording of sentences from the stories that were read.

Reading behavior
We investigated the effect of foregrounding on gaze durations and regressions, while controlling for several other variables. Lexical frequency, age of acquisition, word prevalence, word length, phonological neighborhood size, position on the screen and perplexity all had a significant effect on gaze duration and the chance of regression. Orthographic neighborhood size and semantic priming did have a significant effect on the chance of regression, but not on gaze duration. In accordance with the Russian Formalists' notion of retardation, it was found that in general, foregrounded words were indeed read slower than words that were not foregrounded, confirming H1 (see also [13, 14]). However, when we zoom in on the level of individual participants, the picture becomes more

Figure 6: Correlation between the Retardation Effect and the random slopes of perplexity per participant.

Response option	Condition	
	Foregrounded	**Non-foregrounded**
Correct	136	114
Semantic deviation	55	60
Syntactic deviation	56	59
Semantic and syntactic deviation	23	37

Table 6: Total number of responses per condition for each response option in the recognition test.

complex [53]. Clearly, there is significant variability between readers; some indeed slow down when reading foregrounded words, but others show the reverse effect: They speed up (**Figure 5**). The chance of regressing to a word was higher for words that are foregrounded as compared to words that are not foregrounded. Individual differences in sensitivity to text features has been observed before in behavioral (e.g. [54]), and neuroimaging studies [50, 55], and our current findings highlight the importance of taking individual differences seriously in the study of literary reading (e.g. [6, 56]).

Foregrounded passages can be more difficult than non-foregrounded passages in a number of ways. For example, some passages include ellipsis, or other syntactic deviations that raise linguistic expectations that are not met. A pronoun at the beginning of a sentence creates the expectation that a verb will follow. If this expectation is not met, as in "Zij niet, zij zei dat nú niet" ("She not, she said that now not"), readers potentially regress to the part where a verb was expected in order to exclude the possibility that they simply missed a word.

It seems likely, in line with Shklovsky's [11] position, that both regressions and slowing down are due to difficulty in processing, but it may equally be due to aesthetic feelings the text provokes. **Figure 7** shows a working model of the cognitive processes underlying the observed effect. In this model, the aesthetic response is triggered by the slowing down, and vice versa. Foregrounding de-automatizes perception in this framework not only by appealing to aesthetic preferences directly, but also by increasing processing demands. During foregrounded passages, readers can no longer rely on their usual expectations about the text as it unfolds, as they usually do, and need to pay more attention. This comes down to increased awareness of the surface form of the text, and it may lead to increased appreciation (as long as the linguistic input is not too difficult). Conversely, however, aesthetic appreciation, in the form of being interested in and concerned and fascinated by the text (as in [8]) can also influence reading times directly, as fascinated readers will be motivated to read the text more carefully. We want to point out that this is a hypothesized scenario; the directionality of the effect cannot be determined on the basis of the present data.

Our findings and our interpretation of them are largely in line with Jacobs' [7, 8, 57] recently developed Neurocognitive Poetics Model (NCPM) of literary reading. The NCPM is a dual-route model that predicts that foregrounded text elements are processed more slowly than backgrounded elements. Whereas backgrounded passages allow the reader to become immersed into the story because they consist of familiar words and do not draw attention to the surface structure, foregrounded passages evoke, among other things, attention ("explicit processing"), as well as aesthetic feelings. The aesthetic feelings are in turn predicted to cause slower reading.

With regard to our study, the NCPM correctly predicts not only that readers should be sensitive to foregrounding, but also that there should be no strong relation between immersion and sensitivity to foregrounding, since the two depend on different modes of processing. Of course we should be careful in interpreting a null result, but we have found no evidence that readers who are more immersed in a story also slow down more during foregrounding.

What the NCPM does not explicitly include, however, is a direct link between the modulation of attention and reading pace. We believe that increased attention, which may be a result of the violation of expectations that is brought about by deviations in the style of the text, can also directly affect the readers' pace, without the need for them to experience aesthetic feelings. This interpretation is supported by the strong correlation between response to foregrounding and response to perplexity. The slowing down response to high perplexity, unexpected words, cannot easily be explained with reference to increased aesthetic feelings, but is more likely due to general difficulty with reading passages that deviate from the norm in the language (in the *parole* sense), regardless of their literary status. Yet there is a correlation between sensitivity to foregrounding and sensitivity to perplexity (and both slowing down effects are stronger for those participants who read less often, but these latter correlations were not statistically significant after bonferroni-correction, so we
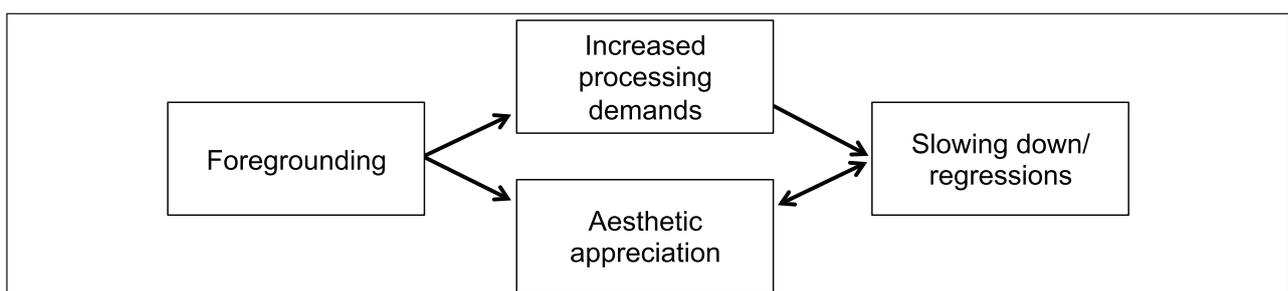

Figure 7: Hypothesized direction of the relation between slowing down and aesthetic response.

should interpret them with caution). This suggests that part of the retardation effect is also simply due to literary language being more difficult than non-literary language, in ways that are not captured very well by any of our control variables.

More supporting evidence for our interpretation comes from a study by Song and Schwarz [58]. In their study, participants read the question "*How many animals of each kind did Moses take on the Ark?*" and either gave an answer, indicated they did not know the answer, or indicated that they could not say because the question was ill-formed (it was Noah who took animals on the Ark, not Moses). There were two conditions: one in which typeface was easy to read and one in which it was difficult to read. The difficult typeface led to significantly more discoveries of the anomaly than the easy typeface. Song and Schwarz conclude that this simple increase of low-level processing demands resulted in deeper processing of the text.

In an fMRI study by Bohrn, Altmann, Lubrich, Menninghaus, and Jacobs [59], it was found that reading unfamiliar proverbs, which are assumed to be more difficult to understand, increased both cognitive and affective processing compared to familiar proverbs. These results are in line with the theoretical considerations of Mukařovský [10] and Shklovsky [11]. If processing is too fluent, or automatic, there is little chance for appreciating the poetic dimension of the text (see [60, 61]). Clearly, there is a limit to this: Very idiosyncratic language use can hinder the flow of information so much that it impairs comprehension. Such a scenario was not the case for the texts that we selected in the present study.

The NCPM makes another interesting prediction with regard to eye-movements during literary reading that we did not have hypotheses about. The model predicts readers to exhibit smaller saccades during foregrounded passages than during backgrounded passages. We here report the results from a mixed effects regression analysis of all rightward saccades in our dataset. We included in the model the same fixed effect predictors that were used for the gaze duration and regression analyses. The effect of literariness was allowed to vary per participant and story, and the effect of perplexity was allowed to vary

per participant as well. Predictor values were based on the word that formed the launch site of the saccade. Predictors were z-transformed to overcome convergence problems. The results of the final model are shown in **Table 7**.

Only prevalence did not have a significant on saccade size. Literariness did have a significant effect, but in the opposite direction from what the NCPM predicts: saccades launched from more literary words are generally longer than those launched from less literary words. A closer look at the individual participants' effect of literariness on saccade size tells us something about how we might interpret this effect: The slope of the effect of literariness on saccade length shows a strong negative correlation with the Retardation Effect, $r = -.66$, $p < .001$, and a moderate negative correlation with the individuals' slope of perplexity on gaze duration, $r = -.37$, $p < .05$. Participants who slow down more during literary passages also display smaller saccades during literary passages. This reading behavior is in line with earlier findings that readers show distinct reading profiles, or "strategies" [62–64]. According to the "Risky Reader Hypothesis" [59, 60], more proactive, "risky readers", display long saccades and many regressions. They rely relatively strongly on guessing which words are in the parafovea, but often need to regress to an earlier word when this strategy fails. Conservative readers on the other hand, display shorter saccades and fewer regressions. It seems that the readers who slow down more during literary passages are the rather conservative readers, whereas those who slow down less are more proactive. This leaves open the possibility that, as a reviewer suggests, some of these more proactive participants may have simply skipped over the text during foregrounded passages because it was too difficult for them, leading to both shorter gaze durations and longer saccades.

Reading experience
We cannot confirm H2 – *Infrequent fiction readers show a larger retardation effect when exposed to foregrounding than frequent fiction readers.* After bonferroni-correction, neither the correlation between retardation and ART/fiction reading nor the correlation between perplexity and ART/fiction reading was significant. The sample size for these correlations was rather small (N = 29), so a follow-up

	ß	SE ß	t	p
constant	0.52	0.72	0.73	<.01
log frequency	0.23	0.68	−3.41	<.0001
word length	0.34	0.59	5.81	<.0001
screen position	0.12	0.37	−32.53	<.05
age of acquisition	0.92	0.46	−1.99	<.0001
log perplexity	0.80	0.59	−13.53	<.01
orthographic neighborhood size	0.16	0.51	−3.63	<.05
phonological neighborhood size	0.11	0.46	2.48	<.0001
semantic relation	0.19	0.44	−4.43	<.05
literariness score	0.17	0.63	2.63	<.01

Table 7: Coefficients for the mixed effects model fit to saccade size. *Note.* ß indicates the standardized beta.

study with a larger number of participants is needed to say anything conclusive about this issue.

Recognition

According to H3, *the memory for the surface form is better for foregrounded passages than for non-foregrounded passages.* This hypothesis cannot be confirmed – we found no effect of foregrounding on the chance of recognition after controlling for total reading times. Correct phrases were recognized above chance-level (46% of the time), but there was no difference between foregrounded and non-foregrounded passages (cf. [14]).

Story liking and immersion

We cannot confirm H4a – *Readers who are more immersed in the story react differently to foregrounded words than readers who are less immersed in the story.* None of the factors of the immersion questionnaire significantly correlated with the effect of foregrounding on reading times. Because this is a null effect, we have to interpret it with caution, but immersion does not seem to play a role in slowing down during reading of foregrounded passages, as predicted by the NCPM [8, 54].

Story Liking did not show a significant positive correlation with the Retardation Effect either: The participants who liked the story better were not more likely to slow down during reading of foregrounded passages. Therefore H4b – *Readers who like the story more react differently to foregrounded words than readers who like the story less* – cannot be confirmed either.

It should be noted that the immersion questionnaire does not necessarily capture immersion as it is experienced during reading. Participants need to recall their feeling of immersion after the story is already finished, and their memories need not be accurate. In future research, this issue can be partially overcome by, for instance, splitting the story into parts and collecting immersion ratings per story part (see [8]).

Conclusions

This study partially confirmed the Russian Formalist's idea of retardation – the idea that foregrounding makes readers slow down. By using direct measurements of eye movements combined with advanced statistical modeling, our study allows a more differentiated understanding of Miall and Kuiken's [13] initial findings. That is, readers do not always slow down during foregrounded passages, this depends on the reader.

We cannot say with certainty what it is that determines whether readers slow down or speed up. Slowing down is not related to how much readers appreciate the story, nor does it correlate with any aspect of immersion, be it empathy with the characters, self-loss, imagery, compassion or understanding, even though all of these factors contribute to how much readers appreciate the story. We can also not conclude from this study that the degree of retardation depends on experience with reading fiction, although the correlation between retardation and slowing down during high-perplexity words suggests that reading proficiency may play a role. What exactly the relevant variables

are that cause the differences in effects between readers is therefore still an open question. Relating back to the introduction, our results provide evidence in favor of an interactional account of literary reading, an account that focuses on the reader as well as the text, and the interaction between the two.

Additional Files

The additional files for this article can be found as follows:

· **Additional File 1: Appendix A.** http://dx.doi.org/10.1525/collabra.39.s1
· **Additional File 2: Appendix B.** http://dx.doi.org/10.1525/collabra.39.s2
· **Additional File 3: Appendix C.** http://dx.doi.org/10.1525/collabra.39.s3

Competing Interests

The authors have no competing interests to declare.

References

1. Hauptmeier, H., Meutsch, D., and Viehoff, R. 1989. Empirical research on understanding literature. Poetics Today, 563–604. DOI: http://dx.doi.org/10.2307/1772905
2. Fish, S. E. 1980. Is there a text in this class?: The authority of interpretive communities. Cambridge: Harvard University Press. DOI: http://dx.doi.org/10.1017/s0047404500009878
3. Iser, W. 1976. The act of reading: A theory of aesthetic response. London: Routledge. (Original work published 1974).
4. Hobbs, J. R. 1990. Literature and Cognition. Menlo Park: CSLI.
5. Dixon, P., Bortolussi, M., Twilley, L. C., and Leung, A. 1993. Literary processing and interpretation: Towards empirical foundations. Poetics, 22(1): 5–33. DOI: http://dx.doi.org/10.1016/0304-422x(93)90018-c
6. Burke, M. 2011. Literary Reading, Cognition and Emotion: An Exploration of the Oceanic Mind. London: Routledge. DOI: http://dx.doi.org/10.4324/9780203840306
7. Jacobs, A. M. 2015a. Neurocognitive Poetics: methods and models for investigating the neuronal and cognitive- affective bases of literature reception. Frontiers in Human Neuroscience, 9: 186. DOI: http://dx.doi.org/10.3389/fnhum.2015.00186
8. Jacobs, A. M. 2015b. Towards a Neurocognitive Poetics Model of Literary Reading. In: Willems, R. M. (Ed.), Cognitive Neuroscience of Natural Language Use. Cambridge, UK: Cambridge University Press. DOI: http://dx.doi.org/10.1017/cbo9781107323667.007

9. Garvin. (ed.). 1964. A Prague school reader on esthetics, literary structure, and style. Washington: Georgetown University Press.

10. Mukařovský, J. 1964. Standard language and poetic language. Reprinted in part In: Garvin, P. (Ed.), A Prague school reader on esthetics, literary structure, and style. Washington: Georgetown University Press. (Original work published 1932).

11. Shklovsky, V. 1965. Art as technique. In: Lemon, L. T., and Reis, M. J. (Eds. & Trans.), Russian formalist criticism: Four essays. Lincoln: University of Nebraska Press (Original work published 1917), pp. 3–24.

12. Hakemulder, J. 2004. Foregrounding and its effect on readers' perception. Discourse Processes, 38(2): 193–218. DOI: http://dx.doi.org/10.1207/s15326950dp3802_3

13. Miall, D. S., and Kuiken, D. 1994. Foregrounding, defamiliarization, and affect: Response to literary stories. Poetics, 22(5): 389–407. DOI: http://dx.doi.org/10.1016/0304-422x(94)00011-5

14. Zwaan, R. A. 1994. Effect of genre expectations on text comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(4): 920. DOI: http://dx.doi.org/10.1037/0278-7393.20.4.920

15. Miall, D. S., and Kuiken, D. 1999. What is literariness? Three components of literary reading. Discourse Processes, 28(2): 121–138. DOI: http://dx.doi.org/10.1080/01638539909545076

16. Miall, D. S., and Kuiken, D. 2001. Shifting perspectives: Readers' feelings and literary response. In: van Peer, W., and Chatman, S. (Eds.), New perpectives on narrative perspective. Albany, New York: State University of New York Press, pp. 289–301

17. Just, M. A., and Carpenter, P. A. 1980. A theory of reading: from eye fixations to comprehension. Psychological Review, 87(4): 329. DOI: http://dx.doi.org/10.1037/0033-295x.87.4.329

18. Bransford, J. D., Barclay, J. R., and Franks, J. J. 1972. Sentence memory: A constructive versus interpretive approach. Cognitive psychology, 3(2): 193–209. DOI: http://dx.doi.org/10.1016/0010-0285(72)90003-5

19. Ratcliff, R., and McKoon, G. 1978. Priming in item recognition: Evidence for the propositional structure of sentences. Journal of Verbal Learning and Verbal Behavior, 17(4): 403–417. DOI: http://dx.doi.org/10.1016/s0022-5371(78)90238-4

20. Loebell, H., and Bock, K. 2003. Structural priming across languages. Linguistics, 41(5): 791–824. DOI: http://dx.doi.org/10.1515/ling.2003.026

21. Fletcher, C. R., and Chrysler, S. T. 1990. Surface forms, textbases, and situation models: Recognition memory for three types of textual information. Discourse Processes, 13(2): 175–190. DOI: http://dx.doi.org/10.1080/01638539009544752

22. Gurevich, O., Johnson, M. A., and Goldberg, A. E. 2010. Incidental verbatim memory for language. Language and Cognition, 2(1): 45–78. DOI: http://dx.doi.org/10.1515/langcog.2010.003

23. Sturt, P., Sanford, A. J., Stewart, A., and Dawydiak, E. 2004. Linguistic focus and good-enough representations: An application of the change-detection paradigm. Psychonomic Bulletin and Review, 11: 882–888. DOI: http://dx.doi.org/10.3758/bf03196716

24. Sanford, A. J. S., Sanford, A. J., Molle, J., and Emmott, C. 2006. Shallow processing and attention capture in written and spoken discourse. Discourse Processes, 42: 109–130. DOI: http://dx.doi.org/10.1207/s15326950dp4202_2

25. Zwaan, R. A., and Van Oostendorp, H. 1993. Do readers construct spatial representations in naturalistic story comprehension? Discourse Processes, 16(1–2): 125–143. DOI: http://dx.doi.org/10.1080/01638539309544832

26. Schrott, R., and Jacobs, A. 2011. Gehirn und Gedicht. Wie wir unsere Wirklichkeiten konstruieren. München: Carl Hanser.

27. Zwaan, R. A. 2004. The Immersed Experiencer: Toward an Embodied Theory of Language Comprehension. In: Ross, B. H. (Ed.), The psychology of learning and motivation: Advances in research and theory, 44, 35–62. DOI: http://dx.doi.org/10.1016/s0079-7421(03)44002-4

28. Sestir, M., and Green, M. C. 2010. You are who you watch: Identification and transportation effects on temporary self-concept. Social Influence, 5(4): 272–288. DOI: http://dx.doi.org/10.1080/15534510.2010.490672

29. Gerrig, R. J. 1993. Experiencing narrative worlds: On the psychological activities of reading. New Haven: Yale University Press.

30. Green, M. C., and Brock, T. C. 2000. The role of transportation in the persuasiveness of public narratives. Journal of personality and social psychology, 79(5): 701. DOI: http://dx.doi.org/10.1037/0022-3514.79.5.701

31. Green, M. C., Brock, T. C., and Kaufman, G. F. 2004. Understanding media enjoyment: The role of transportation into narrative worlds. Communication Theory, 14(4): 311–327. DOI: http://dx.doi.org/10.1111/j.1468-2885.2004.tb00317.x

32. Busselle, R., and Bilandzic, H. 2009. Measuring narrative engagement. Media Psychology, 12(4): 321–347. DOI: http://dx.doi.org/10.1080/15213260903287259

33. Verstegen, J. 1973. De nieuwe vrijheid. Utrecht: AWB.

34. Melchior, W. 1992. De roeping van het vlees. Amsterdam: Contact.

35. Hamelink, J. 1966. Horror vacui. Amsterdam: Polak & Van Gennep.

36. Kuijpers, M. M., Hakemulder, F., Tan, E. S., and Doicaru, M. M. 2014. Exploring absorbing reading experiences. Scientific Study of Literature, 4(1): 89–122. DOI: http://dx.doi.org/10.1075/ssol.4.1.05kui

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

38. Stanovich, K. E., and West, R. F. 1989. Exposure to Print and Orthographic Processing. Reading Research Quarterly, 24(4): 402–33. DOI: http://dx.doi.org/10.2307/747605

39. Acheson, D. J., Wells, J. B., and MacDonald, M. C. 2008. New and updated tests of print exposure and reading abilities in college students. Behavior Research Methods, 40(1): 278–289. DOI: http://dx.doi.org/10.3758/brm.40.1.278

40. Inhoff, A. W., and Rayner, K. 1986. Parafoveal word processing during eye fixations in reading: Effects of word frequency. Perception & Psychophysics, 40, 431–439. DOI: http://dx.doi.org/10.3758/bf03208203

41. Rayner, K., and Fischer, M. H. 1996. Mindless reading revisited: Eye movements during reading and scanning are different. Perception & Psychophysics, 58: 734–747. DOI: http://dx.doi.org/10.3758/bf03213106

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

43. Monsalve, I. F., Frank, S. L., and Vigliocco, G. 2012. Lexical surprisal as a general predictor of reading time. Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics. Avignon, France: Association for Computational Linguistics, pp. 398–408.

44. Frank, S. L., Otten, L. J., Galli, G., and Vigliocco, G. 2015. The ERP response to the amount of information conveyed by words in sentences. Brain and Language, 140, 1–11. DOI: http://dx.doi.org/10.1016/j.bandl.2015.05.010

45. Marian, V., Bartolotti, J., Chabal, S., and Shook, A. 2012. CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. PLoS ONE, 7(8): e43230. DOI: http://dx.doi.org/10.1371/journal.pone.0043230

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

47. Keuleers, E., Stevens, M., Mandera, P., and Brysbaert, M. 2015. Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. Quarterly Journal of Experimental Psychology, 68(8): 1665–1692. DOI: http://dx.doi.org/10.1080/17470218.2015.1022560

48. Frank, S. L., and Willems, R. M. submitted. Anticipation in language comprehension: prediction and priming show distinct patterns of brain activity.

49. Mandera, P., Keuleers, E., and Brysbaert, M. 2017. Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. Journal of Memory and Language, 92: 57–78. DOI: http://dx.doi.org/10.1016/j.jml.2016.04.001

50. R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org/.

51. Bates, D., Maechler, M., Bolker, B., and Walker, S. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-6. Available at: http://CRAN.R-project.org/package=lme4.

52. Rayner, K., and Duffy, S. A. 1986. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. Memory & Cognition, 14: 191–201. DOI: http://dx.doi.org/10.3758/bf03197692

53. Nijhof, A. D., and Willems, R. M. 2015. Simulating Fiction: Individual Differences in Literature Comprehension Revealed with fMRI. PLoS ONE, 10(2): e0116492. DOI: http://dx.doi.org/10.1371/journal.pone.0116492

54. Braun, K., and Cupchik, I. G. C. 2001. Phenomenological and Quantitative Analyses of Absorption in Literary Passages. Empirical Studies of the Arts 19(1): 85–109. DOI: http://dx.doi.org/10.2190/W6TJ-4KKB-856F-03VU

55. Altmann, U., Bohrn, I. C., Lubrich, O., Menninghaus, W., and Jacobs, A. M. 2012. The power of emotional valence – from cognitive to affective processes in reading. Frontiers in Human Neuroscience, 6(192). DOI: http://dx.doi.org/10.3389/fnhum.2012.00192

56. Gibbs. 2011. The individual in the scientific study of literature. Scientific Study of Literature, 1(1): 95–103. DOI: http://dx.doi.org/10.1075/ssol.1.1.10gib

57. Jacobs, A. M. 2011. Neurokognitive poetik: elemente eines modells des literarischen lesens (Neurocognitive poetics: elements of a model of literary reading). In: Schrott, R., and Jacobs, A. M. (Eds.), Gehirn und Gedicht: Wie Wir Unsere Wirklichkeiten Konstruieren. Munich: Carl Hanser Verlag, pp. 492–520.

58. Song, H., and Schwarz, N. 2008. Fluency and the detection of misleading questions: Low processing fluency attenuates the Mozes illusion. Social Cognition, 26: 791–799. DOI: http://dx.doi.org/10.1521/soco.2008.26.6.791

59. Bohrn, I. C., Altmann, U., Lubrich, O., Menninghaus, W., and Jacobs, A. M. 2012. Old proverbs in new skins– an fMRI study on defamiliarization. Frontiers in psychology, 3: 1–18. DOI: http://dx.doi.org/10.3389/fpsyg.2012.00204

60. Menninghaus, W., Bohrn, I. C., Knoop, C. A., Kotz, S. A., Schlotz, W., and Jacobs, A. M. 2015. Rhetorical features facilitate prosodic processing while handicapping ease of semantic comprehension. Cognition, 143, 48–60. DOI: http://dx.doi.org/10.1016/j.cognition.2015.05.026

61. Reber, R., Schwarz, N., and Winkielman, P. 2004. Processing fluency and aesthetic pleasure: is beauty in the perceiver's processing experience? Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc, 8(4): 364–382. DOI: http://dx.doi.org/10.1207/s15327957pspr0804_3

62. Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., and Pollatsek, A. 2006. The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. Psychology and Aging, 21(3): 448–465. DOI: http://dx.doi.org/10.1037/0882-7974.21.3.448

63. Rayner, K., Castelhano, M. S., and Yang, J. 2009. Eye movements and the perceptual span in older and younger readers. Psychology and Aging, 24(3): 755–760. DOI: http://dx.doi.org/10.1037/a0014300

64. Koornneef, A., and Mulders, I. 2016. Can we 'read' the eye-movement patterns of readers? Unraveling the relationship between reading profiles and processing strategies. Journal of Psycholinguistic Research, 1–18. DOI: http://dx.doi.org/10.1007/s10936-016-9418-2