

# Reading Shakespeare Sonnets: Combining Quantitative Narrative Analysis and Predictive Modeling —an Eye Tracking Study

Shuwei Xue  
Freie Universität Berlin, Germany

Jana Lüdtkke  
Freie Universität Berlin, Germany

Teresa Sylvester  
Freie Universität Berlin, Germany

Arthur M. Jacobs  
Freie Universität Berlin, Germany  
Dahlem Institute for Neuroimaging of Emotion  
(D.I.N.E.), Berlin, Germany  
Center for Cognitive Neuroscience Berlin  
(CCNB), Berlin, Germany

As a part of a larger interdisciplinary project on Shakespeare sonnets' reception (Jacobs et al., 2017; Xue et al., 2017), the present study analyzed the eye movement behavior of participants reading three of the 154 sonnets as a function of seven lexical features extracted via Quantitative Narrative Analysis (QNA). Using a machine learning-based predictive modeling approach five 'surface' features (word length, orthographic neighborhood density, word frequency, orthographic dissimilarity and sonority score) were detected as important predictors of total reading time and fixation probability in poetry reading. The fact that one phonological feature, i.e., sonority score, also played a role is in line with current theorizing on poetry reading. Our approach opens new ways for future eye movement research on reading poetic texts and other complex literary materials (cf. Jacobs, 2015c).


Keywords: Literary reading, eye movements, eye tracking, QNA, predictive modeling

*Not marble, nor the gilded monuments  
Of princes, shall outlive this powerful rhyme;*  
William Shakespeare, Sonnets 55 (ll. 1-2)

## Introduction

When was the last time you read a poem, or a piece of literature? The answer of many people might well be 'today' or 'yesterday'. Even though reading literature may no longer count among the essential activities of people's leisure time, it still has a significant number of benefits in promoting, for example, general and cross-

cultural education, social cognition or cognitive development (e.g., Kidd & Castano, 2013; Koopman, 2016; Marr, 2018; Samur et al., 2018). However, within the fields of reading and eye tracking research, single words or single sentences from non-literary materials appear to be the most extensively investigated text materials (e.g., Clifton et al., 2007; Radach & Kennedy, 2013; Rayner, 2009). Although psycholinguistic features, e.g., *word length* or *word frequency*, work differently in a connected text context (Kuperman et al., 2010, 2013; Wallot et al., 2013), empirical research using natural materials like narrative texts or poems are quite rare and the majority of studies on literary works confine to text-based qualitative aspects (e.g., 'close reading'). Reading research seems to be experiencing difficulty to open itself for empirical studies focusing on more natural and ecologically valid reading acts, as recently admonished by several researchers (e.g., Jacobs, 2015a; Radach et al., 2008; Wallot et al., 2013).

Received November 15, 2018; Published March 27, 2019.  
Citation: Xue, S., Lüdtkke, J., Sylvester, T. & Jacobs, A. M. (2019)  
Reading Shakespeare sonnets: combining quantitative narrative  
analysis and predictive modeling—an eye tracking study. *Journal  
of Eye Movement Research*, 12(5):2.  
Digital Object Identifier: 10.16910/jemr.12.5.2  
ISSN: 1995-8692  
This article is licensed under a [Creative Commons Attribution 4.0  
International license](https://creativecommons.org/licenses/by/4.0/). 

With the present study, we aim to explore which and how psycholinguistic features influence literary reading (e.g., some famous poems) by analyzing participants' eye movement behavior which provides a valid measure of moment-to-moment comprehension processes (e.g., Rayner, 1998; Rayner et al., 2006). To achieve our objective, we faced two major challenges: dissecting the complex literary works into measurable and testable features and applying computational methods which can handle the intercorrelated psycholinguistic features and the non-linear relationship between them and reading behavior. In the following sections, we expound the two challenges separately, and at the end put forward our hypotheses.

### Quantitative Narrative Analysis (QNA)

As we all know, natural texts mostly show a high level of complexity. They are built of single words that can be characterized by more than 50 lexical and sublexical features influencing their processing in single-word recognition tasks (Graf et al., 2005). The actual amount of these (or other) lexical features influencing eye movement parameters in natural reading of literary texts is a wide-open empirical question. These complex units then are combined to larger units like phrases, sentences, stanzas or paragraphs which again are characterized by an overabundance of text features (Jacobs, 2015a, 2018b) including a great variety of rhetorical devices (cf. Lausberg, 1960). While it is far from easy to qualitatively describe all these features—as evidenced by extensive debates on e.g., the classification of metaphors and similes (Schrott & Jacobs, 2011)—, the challenge to quantify relevant text features properly is even greater and still in its beginnings. To start empirical investigations using (more) natural and complex materials, appropriate models and methods are necessary to handle the plethora of text and/or reader features and their multiple (nonlinear) interactions. On the modeling side, the Neurocognitive Poetics Model of literary reading (NCPM; Jacobs, 2011, 2015a, b; Nicklas & Jacobs, 2017; Willems & Jacobs, 2016) is a first theoretical account offering predictions about the relationship between different kinds of text features and reader responses, e.g., in eye tracking studies using natural text materials (Müller et al., 2017; van den Hoven et al., 2016). On the methods side, inspired by the NCPM, our group has been working for quite some time on different QNA approaches. In contrast to qualitative analysis, these try to quantitatively describe a maximum of the psycholinguistic features of complex natural verbal

materials, as impressively demonstrated using the example of the 154 Shakespeare sonnets (Jacobs et al., 2017). Additionally, this approach proposes advanced tools for computing both cognitive and affective-aesthetic features potentially influencing reader responses at all three levels of observation, i.e., the experiential (e.g., questionnaires and ratings; Jacobs, 2017; Jacobs et al., 2015a, 2016a, b, 2017; Jacobs & Kinder, 2017, 2018; Jacobs & Lüdtke, 2017), the behavioral (e.g., eye movements; Xue et al., 2017), and the neuronal (Hsu et al., 2015).

Shakespeare's sonnets indeed are a particularly challenging and fascinating stimulus material for QNA and count among the most aesthetically successful or popular pieces of verbal art in the world. Facilitating QNA, most of them have the same structure and rhythmic pattern, typically decasyllabic 14-liners in iambic pentameter with three quatrains and a concluding couplet, making them perfect research materials. They have been the object of countless essays by literary critics and of theoretical scientific studies (e.g., Jakobson & Jones, 1970; Simonton, 1989; Vendler, 1997). Furthermore, all 154 sonnets have been extensively 'QNA-ed' in our previous work yielding precise predictions concerning e.g., eye movement data (Jacobs et al., 2017). Furthermore, to our knowledge, none of the previous studies on reading literary texts or poems (e.g., Carrol & Conklin, 2014; Dixon & Bortolussi, 2016; Jacobs et al., 2016b; van den Hoven et al., 2016; Lauwereyns & d'Ydewalle, 1996; Müller et al., 2017; Sun et al., 1985) examined the eye movement behavior of Shakespeare sonnets.

Since it is not possible to identify all relevant features characterizing a natural text [e.g., over 50 features mentioned for single word recognition (Graf et al., 2005) or over 100 features computed for the corpus of Shakespeare sonnets (Jacobs et al., 2017)], nearly all empirical studies we know of tested only a few selected features while ignoring the others without giving explicit reasons for this neglect, e.g., by using eye tracking (Rayner et al., 2001; Reichle, 2003; Rayner & Pollatsek, 2006; Engbert et al., 2005; Reilly & Radach, 2006; Rayner, 2009). Thus, for the present study about the influence of basic psycholinguistic features we decided to start—relatively—simple by concentrating on a set of seven easily computable (sub)lexical surface features combining well established and less tested ones. We excluded complex inter- and supralexical features (e.g., surprisal, syntactic simplicity), as well as any features that cannot be computed via QNA

(e.g., age-of-acquisition, metaphoricity). The resulting set of surface features consists of two standard features (*word length*, *word frequency*) used in many eye movement studies and three standard features from word recognition studies much less used in the eye movement field (*orthographic neighborhood density*, *higher frequent neighbors*, and *orthographic dissimilarity*), and two phonological features theoretically playing a role in poetry reading (*consonant vowel quotient*, *sonority score*). In the following paragraphs, we further explain these features and summarize their effects, if available, observed in eye tracking studies using single sentences or short nonliterary texts:

In eye tracking studies of reading non-literary texts it is widely acknowledged that longer and low frequency words attract longer total reading time (sum of all fixations on the target word) and more fixations (e.g., Just & Carpenter, 1980; Inhoff & Rayner, 1986; Raney & Rayner, 1995; Pynte et al., 2008). Apart from these two basic surface features, a wealth of research also found effects of *orthographic neighborhood density* (number of words that can be created by changing a single letter of a target word, e.g., bat, fat, and cab are neighbors of cat, Coltheart et al., 1977) in word recognition and reading tasks (see Andrews, 1997, for a review). While effects of *orthographic neighborhood density* are usually facilitative, the presence of *higher frequent neighbors* in the hypothetical mental lexicon inhibits processing of a target word (Grainger et al., 1989; Grainger & Jacobs, 1996; Perea & Pollatsek, 1998). However, there are no clear conclusions as to the effects of both features on eye movements in reading (Williams et al., 2006). Furthermore, using the Levenshtein distance metric, we can also compute an additional *orthographic dissimilarity index* for all words, going beyond the standard operationalization based on words of the same length. As far as we know, systematic effects of the above features on eye movements in the reading of poetry have not been reported so far.

Most people will agree with the statement that poetry is an artful combination of sound and meaning (Schrott & Jacobs, 2011). While the above features are basically ‘orthographic’, the effects of sublexical and lexical phonological features that have been found in a variety of silent reading studies (e.g., Aryani et al., 2013, 2016, 2018a, 2018b; Braun et al., 2009; Schmidtke et al., 2014b; Jacobs, 2015b, c; Ullrich et al., 2017; Ziegler & Jacobs,

1995) and the wide use of phonetic rhetorical devices in poetic language lead us to include also two phonological features: the *consonant vowel quotient* and the *sonority score*. *Consonant vowel quotient* is a simple proxy for the pronounceability of a word—which hypothetically is related to its ease of automatic phonological recoding (Lee et al., 2001). To quantify the acoustic energy or loudness of a sound, called sonority (Ladefoged, 1993), we used the *sonority score*, a simplified index based on the sonority hierarchy of English phonemes, which allows to estimate the degree of distance from the optimal syllable structure (e.g., Clements, 1990). It was previously applied in the study of aphasia (Stenneken et al., 2005) and has recently been proposed as an important feature influencing the subjective beauty of words (Jacobs, 2017). There is evidence that consonant status and sonority play a role in silent reading (Maionchi-Pino et al., 2008; Barent, 2013), especially of poetic texts (Kraxenberger, 2017). Both features have not been examined in literary reading studies using eye tracking.

### **Non-linear Interactive Models and Predictive Modeling**

With the help of QNA, we can quantify psycholinguistic features and predict reader responses successfully (e.g., Jacobs & Kinder, 2018). However, we still need to tackle the second challenge: within and between the disciplines involved in reading research there is an unspoken consent that all these psycholinguistic features influence the reading and interpretation of literary texts in a highly interactive and nonlinear way (Jacobs, 2015a, 2018b; Leech, 1969; Schrott & Jacobs, 2011). Kliegl et al. (1982) already pointed out that using standard accounts like hierarchical regressions is not a solution for handling intercorrelated predictors and the nonlinear relationship between predictors and reading behavior. Consequently, we must look for appropriate tools to tackle these problems. One option is offered by recent developments e.g., in the fields of bioinformatics (Strobl et al., 2009), ecology (e.g., Manel et al., 1999; Were et al., 2015), geology and risk analysis (Nefeslioglu et al., 2008; Saltelli, 2002), quantitative sociolinguistics (Tagliamonte & Baayen, 2012; Van Halteren et al., 2005), epidemiology (e.g., Tu, 1996), neurocognitive poetics (Jacobs, 2017, 2018b; Jacobs & Kinder, 2017, 2018; Jacobs et al., 2017), fMRI data analysis (e.g., Cichy et al., 2017) or applied reading research (Lou et al., 2017; Matsuki et al., 2016) highlighting the application of machine learning tools like

neural nets or bootstrap forests to predictive modeling accounts of big data sets with complex interactions and intercorrelations. Moreover, as an alternative and complement to the traditional ‘explanation approach’ of experimental psychology, machine learning principles and techniques can also help psychology become a more predictive and explorative science (Yarkoni & Westfall, 2017; Cichy & Kaiser, 2019). Thanks to such computational methods, tackling the challenge of analyzing human cognition, emotion or eye movement behavior in rich naturalistic settings (Lappi, 2015) has become a viable option especially as concerns literary reading (e.g., Jacobs & Willems, 2018; Willems, 2015; Willems & Jacobs, 2016).

For present study, two non-linear interactive models, i.e., neural nets and bootstrap forests, were compared with one general linear model (standard least squares regression), to find out which approach optimally predicted relevant eye movement parameters during the reading and experiencing of poetry. The neural net model is a multi-layer perceptron which can predict one or more response variables using a flexible function of the input variables. It has the ability to implicitly detect all possible (nonlinear) interactions between predictor variables and a number of other advantages over regression models when dealing with complex stimulus-response environments (e.g., Tu, 1996). Bootstrap forests predict a response by averaging the predicted response values across many decision trees. Each tree is grown on a bootstrap sample of the training data (Hastie et al., 2009). Both the non-linear interactive models and the general linear model were evaluated in a predictive modeling approach comparing a goodness of fit index ( $R^2$ ) for training and validation sets.

Taken together, in the context of our QNA-based predictive modeling approach, here we considered a minimalistic first attempt at introducing an already considerably more complex way of analyzing eye movements in reading poetic texts. We focused on potential effects of seven simple ‘surface’ features: *word length*, *word frequency*, *orthographic neighborhood density*, *higher frequency neighbors*, *orthographic dissimilarity index*, *consonant vowel quotient*, and *sonority score* on three eye movement parameters (first fixation duration, total reading time and fixation probability).

## Hypotheses

Since non-linear interactive models can deal with complex interactions and detect hidden structures in complex data sets (LeCun et al., 2015), we proposed that they would outperform the general linear model and produce satisfactory model fits for both the training and validation sets.

Based on previous eye tracking studies and existent models of eye movement control (e.g., Engbert et al., 2005; Klitz et al., 2000; Legge et al., 1997; Reichle et al., 2003; Reilly & Radach, 2006), we assumed that *word length* and *word frequency* play a key role in accounting for variance in total reading time and fixation probability, i.e., longer and low frequency words should attract longer total reading time and higher fixation probability also in poetry reading.

On account of the facilitative effect of *orthographic neighborhood density* and the inhibitory effect of *higher frequency neighbors* in the above mentioned word recognition studies, we also expected words with many (lower frequency) orthographic neighbors to produce shorter total reading time and lower fixation probability than low *orthographic neighborhood density* words and words with *higher frequency neighbors*. Similarly, we hypothesized that higher *orthographic dissimilarity* of a word (as a proxy for its orthographic salience) would increase its total reading time and fixation probability.

As concerns the two phonological features, *consonant vowel quotient* and *sonority score*, our hypothesis was that words with a high *consonant vowel quotient* (as a proxy for hindered phonological processing) and *sonority score* (as a proxy for increased aesthetic potential) require a more exigent processing (e.g., Jacobs et al., 1998; Mäionchi-Pino et al., 2008, 2012) and thus would attract longer reading time and higher fixation probability. All effects were assumed to be smaller or non-significant for first fixation durations which usually reflect fast and automatic reading behavior less influenced by lexical parameters (Hyönä & Hujanen, 1997; Clifton et al., 2007).

## Methods

### Participants

Fifteen native English participants (five female;  $M_{\text{age}} = 31.5$  years,  $SD_{\text{age}} = 14.1$ , age range: 18–68 years) were recruited from an announcement released at Freie Uni-

versität Berlin. All participants had normal or corrected-to-normal vision. They were naive to the purposes of the experiment and were not trained literature scholars of poetry. Participants gave their informed, written consent before commencing the experiment and received either course credit or volunteered freely. This study was conducted in line with the standards of the ethics committee of the Department of Education and Psychology at Freie Universität Berlin.

### Apparatus

Participants' eye movements were recorded with a sampling rate of 1000 Hz, using a remote SR Research EyeLink 1000 desktop-mount eye tracker (SR Research Ltd., Mississauga, Ontario, Canada). Stimulus presentation was controlled by Eyelink Experiment Builder software (version 1.10.1630, <https://www.sr-research.com/experiment-builder>). Stimuli were presented on a 19-inch LCD monitor with a refresh rate of 60 Hz and a resolution of 1,024 × 768 pixels. A chin-and-head rest was used to minimize head movements. The distance from the participant's eyes to the stimulus monitor was approximately 50 cm. We only tracked the right eye. Each tracking session was initialized by a standard 9-point calibration and validation procedure to ensure a spatial resolution error of less than 0.5° of visual angle.

### Design and Stimuli

The three Sonnets chosen from the Shakespeare Corpus of 154 sonnets were: Sonnets 27 ('Weary with toil...'), 60 ('Like as the waves...') and 66 ('Tired with all these...'). The choice was made by an interdisciplinary team of experts taking into account the considerable poetic quality and representativeness of the motifs not only within the Shakespeare Sonnet's corpus but also within European poetry. The motifs are: love as tension between body and soul (sonnet 27), death as related to time and soul (sonnet 60) and social evils during the period Shakespeare lived (sonnet 66). All three have the same metrical and rhythmical structure as most Shakespeare sonnets (see Introduction). Inspired by our previous QNA study on Shakespeare sonnets (Jacobs et al., 2017), we conducted a fine-grained lexical analysis of all words used in the present three sonnets, summarized in Table 1. The Pearson Chi-square test indicated no significant differences in the distribution of four main word classes between the three sonnets ( $\chi^2 = 6.31$ ,  $df = 6$ ,  $p$

= .39). We therefore collapsed the data across all sonnets to increase statistical power for predictive modeling.

**Table 1.** Number of Words per Category within Each Sonnet and within all Three Sonnets

Sonnet	<i>Closed-class</i>	<i>Adj./ Adv.</i>	<i>N.</i>	<i>V.</i>	Total
	count [ % ]	count [ % ]	count [ % ]	count [ % ]	
27	49 [44.14]	20 [18.02]	28 [25.23]	14 [12.61]	111
60	48 [44.44]	12 [11.11]	30 [27.78]	18 [16.67]	108
66	33 [36.26]	20 [21.98]	21 [23.08]	17 [18.68]	91
Total	130 [41.94]	52 [16.77]	79 [25.48]	49 [15.81]	310

*Note.* *Closed-class* refers to the category of function words; *Adj./ Adv.* refers to adjective or adverb; *N.* refers to noun; *V.* refers to verb; % is the percentage of each word category within each sonnet or within all three sonnets.

### Procedure

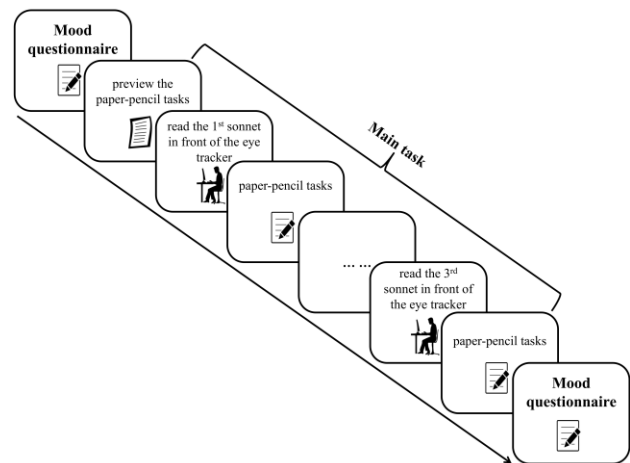
The experiment was conducted in a dimly lit and sound-attenuated room. The data acquisition for each sonnet was split in two parts: a first initial reading of the sonnet with eye tracking and a following paper-pencil memory test accompanied by several rating questions and marking tasks.

For the initial reading participants were instructed to "read each sonnet attentively and naturally" for their own understanding. Prior to the onset of the sonnet on a given trial, participants were presented with a black dot fixation marker (0.6° of visual angle), to the left of (the left-side boundary of) the first word in line 1; the distance between the cross and first word was 4.6°. The sonnets were presented to the participants automatically, when they fixated on a fixation marker presented left to the first line. Participants read the sonnets following their own reading speed. They could go back and forth as often as they wanted within a maximum time window of two minutes. Thirteen participants stopped reading before this deadline. To achieve a certain level of ecological validity, all sonnets were presented left-aligned in the center of the monitor (distance: 8.0° from the left margin of the screen) by using a variable-width font (Arial) with a letter size of 22-point size (approximately 4.5 × 6.5 mm, 0.5 × 0.7 degrees of visual). In order to facilitate accurate eye tracking 1.5-line spacing was used.

For the second part of data acquisition, participants went to another desk to work on the paper-pencil tasks self-developed in close cooperation with literature scholars. Our questionnaire had altogether 18 close- and open-ended questions concerning memory, topic identification, attention, understanding and emotional reactions. It also included three marking tasks where participants had to indicate unknown words, key words and the most beautiful line of the poem (the rating results will be reported elsewhere by the ‘humanities’ section of our interdisciplinary team; Papp-Zipernovszky, Mangen, Lüdtke & Jacobs, in preparation). After answering the questionnaire for the first sonnet, participants continued with reading the second sonnet in front of the eye tracker and so on. The order of the three sonnets was counterbalanced across participants. In order to make the reading of the first sonnet comparable to the reading of the latter two, participants became acquainted with the questionnaire before the initial reading of the first sonnet.

At the beginning and end of the experiment, we used an English translation of the German multidimensional mood questionnaire (MDBF; Steyer et al., 1997) to evaluate the participants’ mood state. This questionnaire assesses three bipolar dimensions of subjective feeling (depressed vs. elevated, calmness vs. restlessness, sleepiness vs. wakefulness) on a 7-point rating scale. The results showed that our participants were in a neutral mood of calmness and slight sleepiness. Simple *t*-tests comparing the mood ratings at the beginning and the end of the experiments indicated no significant mood changes (all *t* (14)s < 1). Thus, reading sonnets did not induce longer-lasting changes in the global dimensions assessed by the MDBF.

Altogether, the experiment took about 40 minutes (see Figure 1 for an illustration of the procedure).



**Figure 1. The Procedure of the Experiment.** An English translation of the German multidimensional mood questionnaire (MDBF; Steyer et al., 1997) was presented to the participants before and after the main tasks to evaluate whether sonnets reading induced longer-lasting changes in participants’ mood state. The data acquisition for each sonnet was split in two parts: first initial reading of the sonnet with eye tracking and the following paper-pencil tasks. After answering the questionnaire for the first sonnet, participants continued with reading the second sonnet in front of the eye tracker and so on. The order of the three sonnets was counterbalanced across participants. In order to make the reading of the first sonnet comparable to the reading of the latter two, participants became acquainted with a questionnaire example before the initial reading of the first sonnet.

## Data Analysis

**Psycholinguistic features.** All seven psycholinguistic features were computed for all unique words (word-type, 205 words, data for words appearing several times in the texts were the same) in the three sonnets based on the Gutenberg Literary English Corpus as reference (GLEC; Jacobs, 2018a): *word length* (*wl*) is the number of letters per word; *word frequency* (*logf*) is the log transformed number of occurrences of word; *orthographic neighborhood density* (*on*) is the number of words of the same length as the target word differing by one letter; *higher frequent neighbors* (*hfn*) is the number of orthographic neighbors with higher word frequency than the target word; *orthographic dissimilarity density* (*odc*) is the target word’s mean Levenshtein distance from all other words in the corpus, a metric that generalizes on to words of different lengths; *consonant vowel quotient* (*cvq*) is the quotient of consonants and vowels in one word; *sonority score* (*sonscore*) is the sum of phonemes’ sonority hierarchy with a division by the square root of *wl* (the sonority

hierarchy of English phonemes yields 10 ranks: [a] > [e o] > [i u j w] > [r] > [l] > [m n ŋ] > [z v] > [f θ s] > [b d g] > [p t k]; Clements, 1990; Jacobs & Kinder, 2018), e.g., in our three sonnets, ART got the *sonscore* of  $10 \times 1 [a] + 7 \times 1 [r] + 1 \times 1 [t] = 18 / \text{SQRT}(3) = 10.39$ .

The correlations between our seven features are given in Table 2. There were several significant correlations (e.g., *wl* & *on*,  $r = .81$ ,  $p < .0001$ ) indicating the usefulness of machine learning tools in literary text reading studies.

**Table 2.** Correlations between Seven QNA Features

Variables	1	2	3	4	5	6	7
1. <i>wl</i>	–						
2. <i>logf</i>	-.75	–					
3. <i>on</i>	-.81	.68	–				
4. <i>hfn</i>	-.31	.00	.36	–			
5. <i>odc</i>	.74	-.48	-.39	-.18	–		
6. <i>cvq</i>	.19	-.10	-.24	-.05	.10	–	
7. <i>sonscore</i>	.72	-.55	-.57	-.28	.62	.00	–

**Eye tracking parameters.** Raw data were pre-processed using the EyeLink Data Viewer (<https://www.sr-research.com/data-viewer/>)<sup>1</sup>. Rectangular areas of interest (AOI) were defined automatically for each word; their centers were coincident with the center of each word. For the upcoming analysis we first calculated for each word, participant and sonnet the first fixation duration (duration of first fixation on the target word) as a measure of word identification, gaze duration (the sum of all fixations on the target word during first pass), re-reading time (sum of fixations on the target word after first pass), and the total reading time (sum of all fixations on the target word) as a measure of general comprehension difficulty (Boston et al., 2008). In a next step we aggregated the data over all participants to obtain the mean values for each word within each sonnet. For this aggregation skipped words were treated as missing values (skipping rate:  $M = .13$ ,  $SD = .04$ ). The amount of skipping was taken into account by calculating the fixation probability for each word. Words fixated by all participants, like ‘captain’ (sonnet 66), ‘cruel’ (sonnet 60) or

<sup>1</sup> Firstly, if fixations of a line drifted from the whole line, we corrected them into the right position. Secondly, fixation durations less than 80 ms were merged with nearby fixations (if the distance between them was less than one degree) or removed from further analysis.

‘quiet’ (sonnet 27) had a probability of 100%. Words fixated by only one or two participants like ‘to’ (sonnet 27), ‘in’ (sonnet 60), or ‘I’ (sonnet 27) had fixation probabilities below 20%. In total, over 40% of the words had a fixation probability of 100% leading to a highly asymmetric distribution. Due to the fact that our psycholinguistic features do not differ for the same word occurring at different positions within a poem all eye tracking measures were aggregated again across sonnets. For all words appearing twice or more often within all three sonnets data were collapsed into a general mean.

Before running the three different models we calculated the correlations between the five aggregated eye tracking parameters. Because gaze duration had a high correlation with first fixation duration ( $r = .56$ ,  $p < .0001$ ) and total reading time ( $r = .73$ ,  $p < .0001$ ), and regression time had a high correlation with total reading time ( $r = .97$ ,  $p < .0001$ ), we only chose first fixation duration, total reading time and fixation probability as response parameters in the predictive modeling analyses (see Table 3).

**Table 3.** Correlations between Five Common Eye-movement Parameters used in Reading Research

Variables	1	2	3	4	5
1. First fixation duration	–				
2. Gaze duration	.56	–			
3. Total reading time	.30	.73	–		
4. Fixation probability	.13	.31	.48	–	
5. Regression time	.16	.53	.97	.47	–

**Predictive modeling.** JMP 14 Pro ([https://www.jmp.com/en\\_us/software/predictive-analytics-software.html](https://www.jmp.com/en_us/software/predictive-analytics-software.html)) was used to run all statistical analyses<sup>2</sup>. The values of all variables (seven predictors and three eye movement parameters) were standardized

<sup>2</sup> Based on the results of pilot and related work (e.g., Jacobs & Lüdtke, 2017), for the neural nets model we used the following parameter set: one hidden layer with 3 nodes, hyperbolic tan (TanH) activation function; number of boosting models = 10, learning rate = 0.1; number of tours = 10. For the bootstrap forests model, we used the default set: number of trees in the forest = 100, number of terms sampled per split = 1, minimum/maximum splits per tree = 10/ 2000, minimum size split = 5, except that we defined the max number of terms = 3. For standard least squares regression analysis, we only specified the seven fixed effects (*wl*, *logf*, *on*, *hfn*, *odc*, *cvq*, and *sonscore*) and predicted each eye tracking parameter using the same seven predictors (emphasis option: effect leverage).

before modeling. To counter possible overfitting, for all three models we used a cross-validation procedure using 90% of the data as training set and the remaining 10% as validation set<sup>3</sup>. Given the intrinsic probabilistic nature of two of the models and the limited sample size ( $N = 205$  words, i.e., about 20 in the validation sets), predictive modeling results varied across repeated runs, depending on which words were selected as training or validation subset. Therefore, the procedure was repeated 1000 times and the model fit scores were averaged (e.g., Were et al., 2015).

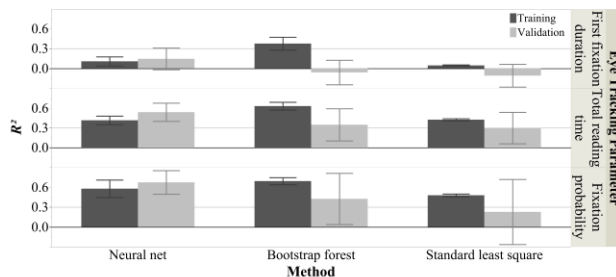
When the model fits of non-linear interactive tools (i.e., neural nets, bootstrap forests) were acceptable ( $R^2 > .30$ ; low  $SD$ ), feature importances ( $FIs$ ) were calculated.  $FI$  is a term used in machine learning ([https://scikit-learn.org/stable/modules/feature\\_selection.html](https://scikit-learn.org/stable/modules/feature_selection.html)). They were computed as the total effect of each predictor assessed by the *dependent resampled inputs* option of the JMP14 Pro software. The total effect is an index quantified by sensitivity analysis reflecting the relative contribution of a feature both alone and in combination with other features (for details, see also Saltelli, 2002). This measure is interpreted as an ordinal value on a scale of 0 to 1 with  $FI$  values  $> .1$  considered ‘important’ (Strobl et al., 2009). To make our results better comparable with previous work, we also tested the effects of ‘important predictors’ ( $FIs > .10$ ) in simple linear regressions using again the cross-validation procedure (90%/ 10% split) for 1000 times, although the intercorrelations between the predictors were not eliminated. If general linear model, i.e., standard least squares regression, got acceptable model fit as described above, instead of reporting  $FIs$  and simple regression results, we would report the mean of 1000 iterations’ parameter estimates.

We repeated the described analytical procedure for all three eye tracking parameters separately.

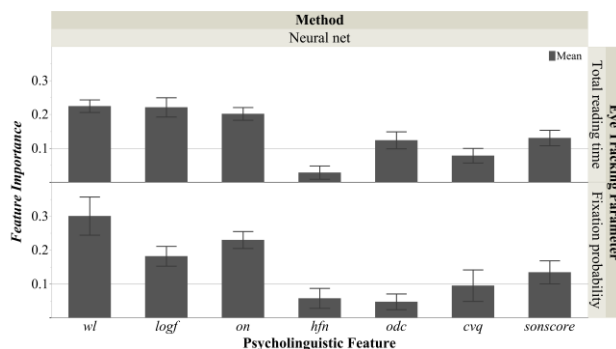
## Results

Figure 2 shows the overall mean  $R^2$ s (averaged across 1000 iterations) for the three eye tracking parameters for both the training and validation sets using all three modeling approaches. Figure 3 shows the seven  $FIs$  for the

optimal non-linear interactive approach. Below we illustrate our results for the three eye tracking parameters respectively. At the end of the results section we also reported the effects of ‘important predictors’ ( $FI > .10$ ) in simple linear regressions.



**Figure 2. Model Fits of Different Measure Groups via Different Modeling Methods.** This figure shows the mean  $R^2$ s from 1000 iterations for three eye tracking parameters for both the training and validation sets using all three modeling approaches. Each error bar is constructed using 1 standard deviation from the mean.



**Figure 3. Feature Importances for Total Reading Time and Fixation Probability.** Figure 3 shows the feature importances ( $FIs$ ) for the neural net model. The  $FIs$  were calculated by using the *dependent resampled inputs* option and mean total effects of 1000 iterations. The total effect is an index quantified by sensitivity analysis, which reflects the relative contribution of that feature both alone and in combination with other features (for details, see Saltelli, 2002). All seven psycholinguistic features were computed for all unique words (word-type, 205 words, data for words appearing several times in the texts were the same) in the three sonnets based on the Gutenberg Literary English Corpus as reference (GLEC; Jacobs, 2018a): *wl* was the number of letters per word; *logf* was log transformed word, *on* was the number of words of the same length as the target differing by one letter, *hfn* was the number of orthographic neighbors with higher word frequency than the target word; *odc* was the target word’s mean Levenshtein distance from all other words in the corpus; *cvq* was the quotient of consonant and vowels in one word; *sonscore* was a simplified index based on the sonority hierarchy of English phonemes which yields 10 ranks (Clements, 1990; Jacobs & Kinder, 2018). Each error bar is construct-

<sup>3</sup> Using a 70/30% training/test cross-validation decreased model fits, probably due to the limited sample size.



ed using 1 standard deviation from the mean. (Note that, because of the bad model fits (see Figure 2), the *FIs* in explaining first fixation duration were excluded from this figure).

### Mean First Fixation Duration

Figure 2 shows that while in the training set (*train*) the bootstrap forests model's fit was satisfactory (mean  $R^2_{train} = .38$ ,  $SD = .10$ ), it did not generalize to the validation set (*val*) at all (mean  $R^2_{val} = -.10$ ,  $SD = .19$ ). The neural nets model and standard least squares regression also showed poor fits for both training (neural nets: mean  $R^2_{train} = .11$ ,  $SD = .07$ ; standard least squares: mean  $R^2_{train} = .05$ ,  $SD = .01$ ) and validation set (neural nets: mean  $R^2_{val} = .15$ ,  $SD = .16$ ; mean  $R^2_{val} = -.10$ ,  $SD = .17$ ). Thus, none of the three models seemed appropriate for predicting first fixation durations during poetry reading (at least not in the present text-reader context). Given the poor model fits, *FIs* were not calculated.

### Mean Total Reading Time

As illustrated in Figure 2, all three model fits in the training set were good (neural nets: mean  $R^2_{train} = .42$ ,  $SD = .07$ ; bootstrap forests: mean  $R^2_{train} = .63$ ,  $SD = .06$ ; standard least squares: mean  $R^2_{train} = .43$ ,  $SD = .02$ ). However, only the neural net model performed well for both the training and validation sets (mean  $R^2_{val} = .54$ ,  $SD = .14$ ), while bootstrap forests' and standard least squares regression's fits in the validation set were smaller and had higher standard deviations (bootstrap forests: mean  $R^2_{val} = .35$ ,  $SD = .25$ ; standard least squares: mean  $R^2_{val} = .30$ ,  $SD = .24$ ).

The *FI* analysis of the optimal neural nets model, shown in Figure 3, suggests that two of the seven features were of minor importance (*FIs* for *hfn* and *cvq* were  $< .10$ ), the rest being important: *wl* (.23), *logf* (.22), and *on* (.20) turned out to be vital predictors, followed by two other less important ones: *sonscore* (.13) and *odc* (.12).

### Fixation Probability

Similar to total reading time, for fixation probability Figure 2 also shows that the fits for the training set of all three models were good (neural nets: mean  $R^2_{train} = .58$ ,  $SD = .13$ ; bootstrap forests: mean  $R^2_{train} = .70$ ,  $SD = .05$ ; standard least squares: mean  $R^2_{train} = .48$ ,  $SD = .02$ ). Again, only the neural nets performed well for both the training and validation sets (mean  $R^2_{val} = .68$ ,  $SD = .18$ ), while the model fits in the validation sets of bootstrap forests and standard least squares regression were insuffi-

cient (bootstrap forests: mean  $R^2_{val} = .43$ ,  $SD = .39$ ; standard least squares: mean  $R^2_{val} = .23$ ,  $SD = .49$ ).

For the *FIs* of neural net model shown in Figure 3, only four predictors were of importance: *wl* (.30)  $>$  *on* (.23)  $>$  *logf* (.18)  $>$  *sonscore* (.14) (*FIs* for *odc*, *hfn* and *cvq* were  $< .10$ ).

### Simple linear regressions

Simple linear regression results indicate that: Words with longer *wl* (total reading time: mean  $R^2_{train} = .37$ ,  $SD = .02$ ; mean  $R^2_{val} = .29$ ,  $SD = .27$ ; fixation probability: mean  $R^2_{train} = .33$ ,  $SD = .01$ ; mean  $R^2_{val} = .14$ ,  $SD = .75$ ), lower *logf* (total reading time: mean  $R^2_{train} = .36$ ,  $SD = .02$ ; mean  $R^2_{val} = .25$ ,  $SD = .26$ ; fixation probability: mean  $R^2_{train} = .27$ ,  $SD = .02$ ; mean  $R^2_{val} = .06$ ,  $SD = .66$ ) and smaller *on* (total reading time: mean  $R^2_{train} = .26$ ,  $SD = .01$ ; mean  $R^2_{val} = .18$ ,  $SD = .23$ ; fixation probability: mean  $R^2_{train} = .33$ ,  $SD = .02$ ; mean  $R^2_{val} = .09$ ,  $SD = .73$ ) had longer total reading time and a higher fixation probability. Words with higher *odc* (total reading time: mean  $R^2_{train} = .17$ ,  $SD = .02$ ; mean  $R^2_{val} = .07$ ,  $SD = .26$ ) attracted longer total reading time. The linear relationship between *sonscore* and the two eye movement parameters was positive: total reading time: mean  $R^2_{train} = .19$ ,  $SD = .01$ ; mean  $R^2_{val} = .11$ ,  $SD = .20$ ; fixation probability: mean  $R^2_{train} = .15$ ,  $SD = .001$ ; mean  $R^2_{val} = .02$ ,  $SD = .41$ .

## Discussion

Following up on earlier proposals (Jacobs et al., 2017), this study aimed to identify psycholinguistic surface features that shape eye movement behavior while reading Shakespeare sonnets by using a combination of QNA and predictive modeling techniques. Since understanding what happens while readers read poetry is a very complex task, a major challenge of Neurocognitive Poetics is to develop appropriate tools facilitating this task (Jacobs, 2015b), in particular new combined computational QNA and machine learning tools (e.g., Jacobs, 2017; Jacobs & Kinder, 2017, 2018). A wealth of text features can be quantified via QNA and their likely non-linear interactive effects can best be analyzed with state-of-the-art predictive modeling techniques which can produce results largely differing from standard general linear model analyses (e.g., van Halteren et al., 2005; Yarkoni & Westfall, 2017). Such techniques can deal with complex interactions difficult to model in a mixed-

effects logistic framework (Tagliamonte & Baayen, 2012) and detect hidden structure in complex data sets, e.g., by recursively scanning and (re-)combining variables (LeCun et al., 2015).

Our results provide evidence for current theoretical discussions which highlight the good reputation regarding the predictive performance of non-linear interactive models (Yarkoni & Westfall, 2017; Cichy & Kaiser, 2019): both non-linear interactive models outperformed the general linear model with higher model fits (mean  $R^2$ ) in the training sets. Regarding the validation sets, again the general linear model performed poorly. Among the two non-linear interactive models, although bootstrap forests produced higher mean  $R^2$  in the training sets, they could not generalize well to the validation set (high  $SD$ ). The poor performance of the general linear model suggests that there are relatively large low-order (e.g., two-way) interactions or other nonlinearities that the non-linear interactive models implicitly captured but that regression did not (cf. Breiman, 2001a; Yarkoni & Westfall, 2017). The good cross-validated performance of our neural nets together with the *FI* analysis offers a considerable heuristic potential for generating hypotheses that can be tested in subsequent experimental designs. Thus, our results suggest that five out of seven surface features (*word length*, *word frequency*, *orthographic neighborhood density*, *sonority score*, and *orthographic dissimilarity index*) are important predictors of mean total reading time, while four (all previous ones minus *orthographic dissimilarity index*) are important for fixation probability, at least in the context of classical poetry.

In line with previous studies, the results from simple linear regressions indicate that longer words with lower *word frequency* and smaller *orthographic neighborhood density* attract longer total reading times and more likely fixations (e.g., Just & Carpenter, 1980; Inhoff & Rayner, 1986; Raney & Rayner, 1995; Pynte et al., 2008; Andrews, 1997). Words with higher *orthographic dissimilarity* also attract longer total reading time. Moreover, a higher *sonority* of a word increased both its total reading time and fixation probability, which is a new finding in poetry reading studies.

Our findings confirm those of previous studies in that longer and low frequency words tend to be fixated more often and longer (e.g., Just & Carpenter, 1980; Inhoff & Rayner, 1986; Raney & Rayner, 1995; Pynte et al., 2008), but also suggest other important predictors, at least for

the reading of poetry: words high in *orthographic neighborhood density* attract less fixations and shorter total reading time supporting the facilitative effect hypothesis of Andrews (1989, 1992). Additionally, words which were more *orthographically dissimilar* (i.e., more salient) attracted longer total reading time. The results concerning the feature *higher frequent neighbors* are inconclusive across the three models which may be due to the fact that in our texts target words had relatively small *higher frequent neighbors* values ( $M = .62$ ,  $SD = 1.24$ ). The effect of this feature requires further investigation using different texts.

Our results also support the hypothesis that through a process of more or less unconscious phonological recoding (Braun et al., 2009; Ziegler & Jacobs, 1995), text sonority may play a role in reading poetic texts: indeed, a higher *sonority* of a word increased both its total reading time and fixation probability supporting our hypothesis. Although replications—e.g. in studies with experimental designs—are required before any conclusions can be drawn, we propose that readers tend to have a more intensive phonological recoding during poetry reading (e.g., Kraxenberger, 2017).

In sum, we take our results as first encouraging evidence that QNA in combination with predictive modeling can be usefully applied to the study of eye tracking behavior in reading complex literary texts. We are also confident that in future studies with bigger samples (i.e., more and longer texts, more readers) and extended feature sets (including interlexical and supralexical ones; Jacobs, 2015b) better generalization performance will be obtained. Here we focused on a few relatively simple QNA-based lexical surface features, but in future studies we will also use computable semantic and syntactic features at the sentence or paragraph levels, as well as predictors related to aesthetic aspects (cf. Jacobs, 2018b).

## Limitations and Outlook

A first obvious limitation of the present analyses is the focus on (sub)lexical surface features. There is little doubt that also other sublexical, lexico-semantic, as well as complex interlexical and supralexical features (e.g., syntactic complexity) affect eye tracking parameters during literary reading and, in fact, the multilevel hypothesis of the NCPM—empirically supported by behavioral,

peripheral-physiological and neuronal data predicts just that (e.g., Hsu et al., 2015; Jacobs et al., 2016b). However, for this first study with a relatively small sample size, we felt that using these seven features—several of which are novel to the field of eye tracking in reading—already made things complicated enough. We think that the present five ‘important’ features will also play a role in future extended predictive modeling studies including other features, but this is of course an open empirical question. We are currently working on extending the present research to other lexical and inter/supra-lexical features including qualitative ones like metaphoricity (e.g., Abramo et al., in preparation), but including more features also requires extending sample sizes (i.e., more/longer texts and more participants), a costly enterprise.

Another issue concerns the fact that word repetition or position was not included in the present analyses (i.e., data for words appearing several times in the texts were averaged). In contrast to the immediacy assumption of Just and Carpenter (1980), parafoveal preview effects as predicted by current eye movement control models indicate that both spatial and temporal eye tracking parameters are affected by other factors than the features of the fixated word (for review see Radach & Kennedy, 2013; Reichle et al., 2003). Moreover, since Just and Carpenter’s (1980) study, it is known that words at line beginnings or ends have a special status. This should also be true for rhyming words at line ends in sonnets or similar poem forms. While we think that our averaging procedure might have added some noise to our data without invalidating them, future studies should definitely have a closer look at word position and repetition effects in poetry reading.

Another limitation is the relatively small sample size of our study. In all, only 15 participants read only three Shakespeare sonnets with only 205 words. Even though we used predictive modeling with 1000 iterations, our findings require replication and extension. However, our goal in this study was to reach out to bridge the gap between text based qualitative analyses (dominant in the humanities) and empirical research on literature reading. In the future, we need to check the validity of our findings with larger samples and the generalizability to other literary works.

In sum, with all caution due to the limitations of this first exploratory study, the present results offer the perspective that some psycholinguistic features so far unused

in (or unknown to) the ‘eye tracking in reading community’, in particular *orthographic neighborhood density* and *sonority score* could be important predictors to be looked at more closely in future research. Whether they are specific to the current selection of three sonnets or of more general interest is a valid open research issue not only for neurocognitive poetics but also for research on eye movements in reading in general.

### Ethics and Conflict of Interest

We declare that the contents of the article are in agreement with the ethics described in <http://biblio.unibe.ch/portale/elibrary/BOP/jemr/ethics.html> and that there is no conflict of interest regarding the publication of this paper.

### Acknowledgements

We wish to thank Giordano D., Gambino, R., Pulvirenti G., Mangen, A., Papp-Zipernovszky, O., Abramo, F., Schuster S., and Schmidtke, D. for providing and discussing some ideas in this project. We also thank Tilk S. for helping with carrying out the experiment. The authors would like to acknowledge networking support by COST Action IS1404 E-READ. Xue S. would like to thank Chinese Scholarship Council for supporting her PhD study at Freie Universität Berlin.

### References

- Abramo, F., Gambino, R., Pulvirenti G., Xue, S., Sylvester, T., Mangen, A., Papp-Zipernovszky, O., Lüdtke, J., & Jacobs, A. M. (in preparation). A qualitative-quantitative analysis of Shakespeare’s sonnet 60.
- Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5), 802. doi: 10.1037/0278-7393.15.5.802
- Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(2), 234. doi: 10.1037/0278-7393.18.2.234

- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin and Review*, 4(4), 439–461. doi: 10.3758/BF03214334
- Aryani, A., Conrad, M., Schmidtke, D., & Jacobs, A. (2018a). Why 'piss' is ruder than 'pee'? The role of sound in affective meaning making. *PLoS One*, 13(6), e0198430. doi: 10.1371/journal.pone.0198430
- Aryani, A., Hsu, C. T., & Jacobs, A. M. (2018b). The Sound of Words Evokes Affective Brain Responses. *Brain sciences*, 8(6), 94. doi: 10.3390/brainsci8060094
- Aryani, A., Jacobs, A.M., & Conrad, M. (2013). Extracting salient sublexical units from written texts: “Emophon,” a corpus-based approach to phonological iconicity. *Frontiers in Psychology*, 4, 654. doi: 10.3389/fpsyg.2013.00654
- Aryani, A., Kraxenberger, M., Ullrich, S., Jacobs, A. M., & Conrad, M. (2016). Measuring the basic affective tone of poems via phonological saliency and iconicity. *Psychology of Aesthetics, Creativity, and the Arts*, 10(2), 191. doi: 10.1037/aca0000033
- Berent, I. (2013). The phonological mind. *Trends in cognitive sciences*, 17(7), 319–327. doi: 10.1016/j.tics.2013.05.004
- Boston, M. F., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, 2(1). doi: 10.16910/jemr.2.1.1
- Braun, M., Hutzler, F., Ziegler, J. C., Dambacher, M., & Jacobs, A. M. (2009). Pseudohomophone effects provide evidence of early lexico-phonological processing in visual word recognition. *Human Brain Mapping*, 30(7), 1977–1989. doi: 10.1002/hbm.20643
- Breiman, L. (2001a). Random forests. *Machine Learning*, 45, 5–32. doi: 10.1023/A:1010933404324
- Carroll, G., & Conklin, K. (2014). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition*, 17(4), 784–797. doi: 10.1017/S1366728913000795
- Cichy, R. M., & Kaiser, D. (2019). Deep Neural Networks as Scientific Models. *Trends in cognitive sciences*. doi: 10.1016/j.tics.2019.01.009
- Cichy, R. M., Khosla, A., Pantazis, D., & Oliva, A. (2017). Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks. *Neuroimage*, 153, 346–358. doi: 10.1016/j.neuroimage.2016.03.063
- Clements, G. N. (1990). The role of the sonority cycle in core syllabification. *Papers in Laboratory Phonology*, 1, 283–333. doi: 10.1017/CBO9780511627736.017
- Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye Movements: A Window on Mind and Brain* (pp. 341–371). Amsterdam, Netherlands: Elsevier. doi: 10.1016/B978-008044980-7/50017-3
- Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance VI* (pp. 535–555). New York: Academic Press.
- Dixon, P., & Bortolussi, M. (2016). Measuring Literary Experience: Comment on Jacobs (2015). *Scientific Study of Literature*, 5(2), 178–182. doi: 10.1075/ssol.5.2.03dix
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: a dynamical model of saccade generation during reading. *Psychological Review*, 112(4), 777. doi: 10.1037/0033-295X.112.4.777
- Graf, R., Nagler, M., & Jacobs, A. M. (2005). Faktorenanalyse von 57 Variablen der visuellen Worterkennung. *Zeitschrift für Psychologie/Journal of Psychology*, 213(4), 205–218. doi: 10.1026/0044-3409.213.4.205
- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: a multiple read-out model. *Psychological Review*, 103(3), 518. doi: 10.1037/0033-295X.103.3.518
- Grainger, J., O’regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception and Psychophysics*, 45(3), 189–195. doi: 10.3758/BF03210696
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). Unsupervised learning. In *The Elements of Statistical Learning* (pp. 485–585). Springer, New York, NY. doi: 10.1007/b94608\_14
- Hsu, C. T., Jacobs, A. M., Citron, F. M., & Conrad, M. (2015). The emotion potential of words and passages

- in reading Harry Potter—An fMRI study. *Brain and Language*, 142, 96–114. doi: 10.1016/j.bandl.2015.01.011
- Hyönä, J., & Hujanen, H. (1997). Effects of case marking and word order on sentence parsing in Finnish: An eye fixation analysis. *The Quarterly Journal of Experimental Psychology Section A*, 50(4), 841–858. doi: 10.1080/713755738
- Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception and Psychophysics*, 40(6), 431–439. doi: 10.3758/BF03208203
- Jacobs, A. M. (2011). Neurokognitive Poetik: Elemente eines Modells des literarischen Lesens [Neurocognitive poetics: Elements of a model of literary reading]. In *Gehirn und Gedicht: Wie wir unsere Wirklichkeiten konstruieren [Brain and Poetry: How We Construct Our Realities]*, Edited by R. Schrott and A.M. Jacobs, (pp. 492–520). Munich: Carl Hanser.
- Jacobs, A. M. (2015a). Neurocognitive poetics: methods and models for investigating the neuronal and cognitive–affective bases of literature reception. *Frontiers Human Neuroscience*, 9, 186. doi: 10.3389/fnhum.2015.00186
- Jacobs, A. M. (2015b). Towards a neurocognitive poetics model of literary reading. In R. M. Willems (Ed.), *Cognitive Neuroscience of Natural Language Use* (pp. 135–159). New York, NY, US: Cambridge University Press. doi: 10.1017/CBO9781107323667.007
- Jacobs, A. M. (2015c). The scientific study of literary experience: sampling the state of the art. *Scientific Study of Literature*, 5, 139–170. doi: 10.1075/ssol.5.2.01jac
- Jacobs, A. M. (2017). Quantifying the beauty of words: a neurocognitive poetics perspective. *Frontiers in Human Neuroscience*, 11, 622. doi: 10.3389/fnhum.2017.00622
- Jacobs, A. M. (2018a). The Gutenberg English Poetry Corpus: exemplary Quantitative narrative analyses. *Frontiers in Digital Humanities*, 5, 5. doi: 10.3389/fdigh.2018.00005
- Jacobs, A. M. (2018b). (Neuro-) Cognitive Poetics and Computational Stylistics. *Scientific Study of Literature*, 8(1), 165–208. doi: 10.1075/ssol.18002
- Jacobs, A. M., & Kinder, A. (2017). The brain is the prisoner of thought: a machine-learning assisted quantitative narrative analysis of literary metaphors for use in Neurocognitive Poetics. *Metaphor and Symbol*, 32, 139–160. doi: 10.1080/10926488.2017.1338015
- Jacobs, A. M., & Kinder, A. (2018). What makes a metaphor literary? Answers from two computational studies. *Metaphor and Symbol*, 33(2), 85–100. doi: 10.1080/10926488.2018.1434943
- Jacobs, A. M., & Lüdtke, J. (2017). Immersion into narrative and poetic worlds. *Narrative Absorption*, 27, 69. doi: 10.1075/nal.27.05jac
- Jacobs, A. M., & Willems, R. M. (2018). The fictive brain: neurocognitive correlates of engagement in literature. *Review of General Psychology*, 22(2), 147–160. doi: 10.1037/gpr0000106
- Jacobs, A. M., Hofmann, M. J., & Kinder, A. (2016a). On elementary affective decisions: to like or not to like, that is the question. *Frontiers Psychology*, 7, 1836. doi: 10.3389/fpsyg.2016.01836
- Jacobs, A. M., Lüdtke, J., Aryani, A., Meyer-Sickendiek, B., & Conrad, M. (2016b). Mood- empathic and aesthetic responses in poetry reception: a model-guided, multilevel, multimethod approach. *Scientific Study of Literature*, 6, 87–130. doi: 10.1075/ssol.6.1.06jac
- Jacobs, A. M., Rey, A., Ziegler, J. C., & Grainger, J. (1998). MROM-p: An interactive activation, multiple readout model of orthographic and phonological processes in visual word recognition. In J. Grainger & A. M. Jacobs (Eds.), *Scientific Psychology Series. Localist Connectionist Approaches to Human Cognition* (pp. 147–188). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers. <https://www.researchgate.net/publication/284659394>
- Jacobs, A. M., Schuster, S., Xue, S., & Lüdtke, J. (2017). What’s in the brain that ink may character ....: a quantitative narrative analysis of Shakespeare’s 154 sonnets for use in neurocognitive poetics. *Scientific Study of Literature*, 7, 4–51. doi: 10.1075/ssol.7.1.02jac
- Jacobs, A. M., Vö, M. L. H., Briesemeister, B. B., Conrad, M., Hofmann, M. J., Kuchinke, L., & Braun, M. (2015a). 10 years of BAWLing into affective and aesthetic processes in reading: what are the echoes? *Frontiers in Psychology*, 6, 714. doi: 10.3389/fpsyg.2015.00714

- Jakobson, R., & Jones, L. G. (1970). *Shakespeare's verbal art in Th'expençe of spirit* (No. 35). Walter de Gruyter.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329. doi: 10.1037/0033-295X.87.4.329
- Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science*, 342(6156), 377–380. doi: 10.1126/science.1239918
- Kliegl, R., Olson, R. K., & Davidson, B. J. (1982). Regression analyses as a tool for studying reading processes: Comment on Just and Carpenters eye fixation theory. *Memory and Cognition*, 10(3), 287–296. doi: 10.3758/BF03197640
- Klitz, T. S., Legge, G. E., & Tjan, B. S. (2000). Saccade planning in reading with central scotomas: Comparison of human and ideal performance. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a Perceptual Process* (pp. 667–682). Amsterdam, Netherlands: North-Holland/Elsevier Science Publishers. doi: 10.1016/B978-008043642-5/50031-0
- Koopman, E. M. E. (2016). Effects of “literariness” on emotions and on empathy and reflection after reading. *Psychology of Aesthetics, Creativity, and the Arts*, 10(1), 82. doi: 10.1037/aca0000041
- Kraxenberger, M. (2017). On Sound-Emotion Associations in Poetry. Ph.D. Thesis, Freie Universität Berlin.
- Kuperman, V., Dambacher, M., Nuthmann, A., & Kliegl, R. (2010). The effect of word position on eye-movements in sentence and paragraph reading. *The Quarterly Journal of Experimental Psychology*, 63(9), 1838–1857. doi: 10.1080/17470211003602412
- Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. *The Quarterly Journal of Experimental Psychology*, 66(3), 563–580. doi: 10.1080/17470218.2012.658820
- Ladefoged, P. (1993). *A Course in Phonetics*. (Third Edition). Fort Worth: Harcourt, Brace Jovanovich College Publishers.
- Lappi, O. (2015). Eye Tracking in the Wild: the Good, the Bad and the Ugly. *Journal of Eye Movement Research*, 8(5). doi: 10.16910/jemr.8.5.1
- Lausberg, H. (1960). *Handbuch der literarischen Rhetorik: eine Grundlegung der Literaturwissenschaft. [Hauptbd.]*. Hueber.
- Lauwereyns, J., & d'Ydewalle, G. (1996). Knowledge acquisition in poetry criticism: The expert's eye movements as an information tool. *International Journal of Human-Computer Studies*, 45(1), 1–18. doi: 10.1006/ijhc.1996.0039
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436. doi: 10.1038/nature14539
- Lee, H.W., Rayner, K., & Pollatsek, A. (2001). The relative contribution of consonants and vowels to word identification during reading. *Journal of Memory and Language*, 44(2), 189–205. doi: 10.1006/jmla.2000.2725
- Leech, G. N. (1969). *A linguistic guide to English poetry*. London, United Kingdom: Longman.
- Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. Chips: an ideal-observer model of reading. *Psychological Review*, 104(3), 524. doi: 10.1037/0033-295X.104.3.524
- Lou, Y., Liu, Y., Kaakinen, J. K., & Li, X. (2017). Using support vector machines to identify literacy skills: Evidence from eye movements. *Behavior Research Methods*, 49(3), 887–895. doi: 10.3758/s13428-016-0748-7
- Maïonchi-Pino, N., Cara, B. D., Magnan, A., & Ecalle, J. (2008). Roles of consonant status and sonority in printed syllable processing: Evidence from illusory conjunction and audio-visual recognition tasks in French adults. *Current Psychology Letters. Behaviour, Brain and Cognition*, 24(2). <https://journals.openedition.org/cpl/4033>
- Maïonchi-Pino, N., de Cara, B., Ecalle, J., & Magnan, A. (2012). Are French dyslexic children sensitive to consonant sonority in segmentation strategies? Preliminary evidence from a letter detection task. *Research in Developmental Disabilities*, 33(1), 12–23. doi: 10.1016/j.ridd.2011.07.045
- Manel, S., Dias, J. M., Buckton, S. T., & Ormerod, S. J. (1999). Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology*, 36(5), 734–747. doi: 10.1046/j.1365-2664.1999.00440.x

- Marr, M. J. (2018). Overtaking the Creases: Introduction to Francis Mechner's "A Behavioral and Biological Analysis of Aesthetics". *The Psychological Record*, 68(3), 285. doi: 10.1007/s40732-018-0307-y
- Matsuki, K., Kuperman, V., & Van Dyke, J. A. (2016). The Random Forests statistical technique: An examination of its value for the study of reading. *Scientific Studies of Reading*, 20(1), 20–33. doi: 10.1080/10888438.2015.1107073
- Müller, H. J., Geyer, T., Günther, F., Kacian, J., & Pierides, S. (2017). Reading English-language haiku: processes of meaning construction revealed by eye movements. *Journal of Eye Movement Research*, 10(1). doi: 10.16910/jemr.10.1.4
- Nefeslioglu, H. A., Gokceoglu, C., & Sonmez, H. (2008). An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. *Engineering Geology*, 97(3–4), 171–191. doi: 10.1016/j.enggeo.2008.01.004
- Nicklas, P., & Jacobs, A. M. (2017). Rhetoric, Neurocognitive Poetics, and the Aesthetics of Adaptation. *Poetics Today*, 38(2), 393–412. doi: 10.1215/03335372-3869311
- Perea, M., & Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 767. doi: 10.1037/0096-1523.24.3.767
- Pynte, J., New, B., & Kennedy, A. (2008). A multiple regression analysis of syntactic and semantic influences in reading normal text. *Journal of Eye Movement Research*, 2(1). doi: 10.16910/jemr.2.1.4
- Radach, R., & Kennedy, A. (2013). Eye movements in reading: Some theoretical context. *The Quarterly Journal of Experimental Psychology*, 66(3), 429–452. doi: 10.1080/17470218.2012.750676
- 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. doi: 10.1007/s00426-008-0173-3
- Raney, G. E., & Rayner, K. (1995). Word frequency effects and eye movements during two readings of a text. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 49(2), 151. doi: 10.1037/1196-1961.49.2.151
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3), 372. doi : 10.1037/0033-2909.124.3.372
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506. doi: 10.1080/17470210902816461
- Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific studies of reading*, 10(3), 241–255. doi: 10.1207/s1532799xssr1003\_3
- Rayner, K., & Pollatsek, A. (2006). Eye-movement control in reading. In *Handbook of Psycholinguistics* (Second Edition) (pp. 613–657). doi: 10.1016/B978-012369374-7/50017-1
- Rayner, K., Binder, K. S., Ashby, J., & Pollatsek, A. (2001). Eye movement control in reading: Word predictability has little influence on initial landing positions in words. *Vision Research*, 41(7), 943–954. doi: 10.1016/S0042-6989(00)00310-2
- Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26(4), 445–476. doi: 10.1017/S0140525X03000104
- Reilly, R. G., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research*, 7(1), 34–55. doi: 10.1016/j.cogsys.2005.07.006
- Saltelli, A. (2002). Sensitivity analysis for importance assessment. *Risk Analysis*, 22(3), 579–590. doi: 10.1111/0272-4332.00040
- Samur, D., Tops, M., & Koole, S. L. (2018). Does a single session of reading literary fiction prime enhanced mentalising performance? Four replication experiments of Kidd and Castano (2013). *Cognition and Emotion*, 32(1), 130–144. doi: 10.1080/02699931.2017.1279591
- Schmidtke, D.S., Conrad, M., & Jacobs, A.M. (2014b). Phonological iconicity. *Frontiers in Psychology*, 5, 80. doi: 10.3389/fpsyg.2014.00080
- Schrott, R., & Jacobs, A.M. (2011). *Gehirn und Gedicht: Wie wir unsere Wirklichkeiten konstruieren (Brain and Poetry: How We Construct Our Realities)*. München, Germany: Hanser.

- Simonton, D. K. (1989). Shakespeare's Sonnets: A Case of and for Single-Case Historiometry. *Journal of Personality*, 57, 695–721. doi: 10.1111/j.1467-6494.1989.tb00568.x
- Stenneken, P., Bastiaanse, R., Huber, W., & Jacobs, A. M. (2005). Syllable structure and sonority in language inventory and aphasic neologisms. *Brain and Language*, 95, 280–292. doi: 10.1016/j.bandl.2005.01.013
- Steyer, R., Schwenkmezger, P., Notz, P., & Eid, M. (1997). Der Mehrdimensionale Befindlichkeitsfragebogen (MDBF). Hogrefe, Göttingen.
- Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, 14(4), 323. doi: 10.1037/a0016973
- Sun, F., Morita, M., & Stark, L. W. (1985). Comparative patterns of reading eye movement in Chinese and English. *Perception and Psychophysics*, 37(6), 502–506. doi: 10.3758/BF03204913
- Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, 24(2), 135–178. doi: 10.1017/S0954394512000129
- Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*, 49(11), 1225–1231. doi: 10.1016/S0895-4356(96)00002-9
- Ullrich, S., Aryani, A., Kraxenberger, M., Jacobs, A. M., & Conrad, M. (2017). On the relation between the general affective meaning and the basic sublexical, lexical, and inter-lexical features of poetic texts—a case study using 57 poems of HM Enzensberger. *Frontiers in Psychology*, 7, 2073. doi: 10.3389/fpsyg.2016.02073
- 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. doi: 10.1525/collabra.39
- Van Halteren, H., Baayen, H., Tweedie, F., Haverkort, M., & Neijt, A. (2005). New machine learning methods demonstrate the existence of a human stylome. *Journal of Quantitative Linguistics*, 12(1), 65–77. doi: 10.1080/09296170500055350
- Vendler, H. (1997). *The Art of Shakespeare's Sonnets*. Cambridge, MA: Harvard University Press.
- Wallot, S., Hollis, G., & van Rooij, M. (2013). Connected text reading and differences in text reading fluency in adult readers. *PloS One*, 8(8), e71914. doi: 10.1371/journal.pone.0071914
- Were, K., Bui, D. T., Dick, Ø. B., & Singh, B. R. (2015). A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. *Ecological Indicators*, 52, 394–403. doi: 10.1016/j.ecolind.2014.12.028
- Willems, R. M. (Ed.). (2015). *Cognitive neuroscience of natural language use*. Cambridge University Press.
- Willems, R. M., & Jacobs, A. M. (2016). Caring about Dostoyevsky: the untapped potential of studying literature. *Trends in Cognitive Sciences*, 20(4), 243–245. doi: 10.1016/j.tics.2015.12.009
- Williams, C. C., Perea, M., Pollatsek, A., & Rayner, K. (2006). Previewing the neighborhood: The role of orthographic neighbors as parafoveal previews in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 32(4), 1072. doi: 10.1037/0096-1523.32.4.1072
- Xue, S., Giordano, D., Lüdtke, J., Gambino, R., Pulvirenti, G., Spampinato, C., & Jacobs, A.M. (2017). Weary with toil, I haste me to my bed: Eye tracking Shakespeare sonnets. Talk given at the 19th European Conference on Eye Movements, Wuppertal, Germany.
- Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12(6), 1100–1122. doi: 10.1177/1745691617693393
- Ziegler, J. C., & Jacobs, A. M. (1995). Phonological information provides early sources of constraint in the processing of letter strings. *Journal of Memory and Language*, 34(5), 567. doi: 10.1006/jmla.1995.1026