

Fachbereich Erziehungswissenschaft und Psychologie
der Freien Universität Berlin

Reading and Rereading Shakespeare's Sonnets:
Combining Quantitative Narrative Analysis and
Predictive Modeling

Dissertation

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaft (Dr. rer. nat.)

vorgelegt von

Shuwei Xue (M.Sc.)

Berlin, 2020

Erstgutachter: Prof. Dr. Arthur M. Jacobs

Zweitgutachter: Prof. Dr. Lars Kuchinke

Tag der Disputation: 10. August 2020

Acknowledgments

I would like to thank my supervisor Prof. Dr. Arthur M. Jacobs for his guidance on this extraordinary journey of interdisciplinary research. In these years, he provided lots of valuable advice and guidance in my research. He taught me writing my first peer-reviewed paper sentence by sentence. I deeply admire his positive work attitude and professional skills. He is more than a perfect supervisor during my life in Berlin. And many thanks to Prof. Dr. Lars Kuchinke for agreeing to be the second reviewer of my dissertation.

I would also like to thank Jana Lüdtkke. She always gives me precious suggestions when I get stuck in my study, and always encourages me when I am upset about myself. I have not only obtained knowledge, but also rigorous academic attitude from her. I would also like to express my thanks to Teresa Sylvester and Sarah Schuster for sharing their expertise as co-authors of publications.

I also thank Teresa Sylvester, Michael Kuhlmann, Eva Fröhlich, Johanna Liebig, David Schmidtke, Sascha Tamm, Adriana Perdigão, Franziska Usée, Wolf Culemann, Frederic Haselon, Silje Tilk and Elena Schmidt for their collegial support in Arthur M. Jacobs' laboratory at the Freie Universität in Berlin.

My studies were generously enabled by the China Scholarship Council, and my formulations were kindly polished by Phil Stanway, Johanna Liebig, and Teresa Sylvester. I am indebted to my parents and other friends for years of encouragement.

Summary

Natural reading is rather like a juggling feat, as our eyes and minds are kept on several things at the same time. Instead, reading texts developed by researchers (so-called “textoids”; Graesser, Millis, & Zwaan, 1997) may be fairly simple, since this facilitates an experimental investigation. It thus provides the chance for clear statements regarding the effect of predefined variables. Likewise, most empirical studies focused only a few selected features while ignoring the great diversity of possibly important others (e.g., Rayner et al., 2001; Reichle, Rayner, & Pollatsek, 2003; Rayner & Pollatsek, 2006; Engbert et al., 2005; Rayner, 2009). However, it is not possible to directly transfer the results generated from textoids to natural reading due to the identification of more than 100 features on different hierarchical levels, which may influence processing a natural text (Graf, Nagler, & Jacobs, 2005; Jacobs, 2015a, b; Jacobs et al., 2017).

The present dissertation differed from past research in that it used a literary text, i.e., Shakespeare’s sonnets, instead of texts constructed by the experimenter. The goal of the present dissertation was to investigate how psycholinguistic features may influence the reading behavior during poem perception. To this end, two problems need to be handled: Firstly, complex natural texts need to be broken up into measurable and testable features by “turning words into numbers” (Franzosi, 2010) for the sake of statistical analysis. Secondly, statistical ways were sought to deal with the non-linear webs of correlations among different features, which has long been a concern of Jacob’s working group (e.g., Willems, 2015; Willems & Jacobs, 2016; Jacobs & Willems, 2018). A quantitative narrative analysis (QNA) based predictive modeling approach was suggested to solve the above problems (e.g., Jacobs et al., 2017; Jacobs, 2017, 2018a, b). Since it is impossible to identify all relevant features of 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)] and including more inter/supra-lexical features also requires extending sample sizes (i.e., more/longer texts and more participants), my dissertation focuses on lexical features. Seven of these are surface features (*word length*, *word frequency*, *orthographic neighborhood density*, *higher frequency neighbors*, *orthographic dissimilarity index*, *consonant vowel quotient*, and the *sonority score*) and two are affective-semantic features (*valence* and *arousal*).

By applying the QNA-based predictive modeling approach, I conducted three eye tracking studies: study 1 (Chapter 5) asked English native speakers to read three of Shakespeare's sonnets (sonnet 27, 60, and 66), aiming to investigate the role of seven surface psycholinguistic features in sonnets reading. Study 2 (Chapter 6) used a rereading paradigm and let another group of English natives read two of the three sonnets (sonnet 27 and 66), to find out whether the roles of the surface psycholinguistic features may be changed in rereading. In study 3 (Chapter 7), I reanalyzed the data of study 2, in which beyond the surface features I started to pay attention to the affective-semantic features, hoping to examine whether the roles of surface and affective-semantic features may be different throughout reading sessions. The three studies show highly reliable data for high feature importance of surface variables, and in rereading an increasing impact of affective-semantic features in reading Shakespeare's sonnets. From a methodological viewpoint, all three studies show a much better sufficiency of neural net approach than the classical general linear model approach in psycholinguistic eye tracking research. For the rereading studies, in general, compared to the first reading, rereading improved the fluency of reading on poem level (shorter total reading times, shorter regression times, and lower fixation probability) and the depth of comprehension (e.g., Hakemulder, 2004; Kuijpers & Hakemulder, 2018). Contrary to the other rereading studies using literary texts (e.g., Dixon et al., 1993; Millis, 1995; Kuijpers & Hakemulder, 2018), no increase in appreciation was apparent.

In summary, this dissertation can show that the application of predictive modeling to investigate poetry might be far more suitable to capture the highly interactive, non-linear composition of linguistic features in natural texts that guide reading behavior and reception. Besides, surface features seem to influence reading during all reading sessions, while affective-semantic features seem to increase their importance in line with processing depth as indicated by higher influence during rereading. The results seem to be stable and valid as I could replicate these novel findings using machine learning algorithms within my dissertation project. My dissertation project is a first step towards a more differentiated picture of the guiding factors of poetry reception and a poetry specific reading model.

Zusammenfassung

Das natürliche Lesen kann man als Kunststück oder auch einen Kraftakt betiteln, da unsere Augen und unser Geist zahlreiche Aufgaben parallel erfüllen muss. Im Gegensatz dazu, mag das Lesen von künstlichen Textelementen (so genannte Texttoide; Graesser, Millis, & Zwaan, 1997), die speziell für die Erforschung bestimmter Variablen des Leseprozesses entwickelt wurden, relativ einfach sein. Der Vorteil dieser Texttoide ist, dass sie die Möglichkeit bieten, experimentelle Fragen zu definierten Variablen klar zu untersuchen und zu beantworten. Hinzu kommt, dass sich die Forschung zumeist nur eine kleine begrenzte Anzahl an Merkmalen untersucht und dabei ignoriert, dass es eine große Bandbreite an möglichen Merkmalen gibt, die den Leseprozess beeinflussen könnten (z.B., Rayner et al., 2001; Reichle, Rayner, & Pollatsek, 2003; Rayner & Pollatsek, 2006; Engbert et al., 2005; Rayner, 2009). Die Ergebnisse der Erforschung des Lesens von Textoiden kann jedoch nicht direkt auf das Lesen von natürlichen Texten übertragen werden, da hierfür über 100 hierarchisch organisierte Merkmale identifiziert wurden, die den Leseablauf beeinflussen könnten (Graf, Nagler, & Jacobs, 2005; Jacobs, 2015a, 2018a; Jacobs et al., 2017).

Die vorliegende Dissertation unterscheidet sich von der zuvor beschriebenen Forschung, da ich literarische Texte, Shakespeares Sonette, untersucht habe, statt konstruierte Texttoide zu verwenden. Das Ziel der Dissertation war es herauszufinden, wie verschiedene psycholinguistische Merkmale das Leseverhalten beeinflussen, wenn wir Gedichte wahrnehmen/ verarbeiten. Um diese Frage zu beantworten, mussten zunächst zwei wichtige Probleme gelöst werden: Zunächst mussten die komplexen Sonette in messbare und falsifizierbare Einheiten aufgeteilt werden—die Wörter mussten also in Zahlen umgewandelt werden (Franzosi, 2010), damit eine statistische Analyse möglich wurde. Zweitens, habe ich, in der Tradition der Arbeitsgruppe um Arthur Jacobs (z.B., Willems, 2015; Willems & Jacobs, 2016; Jacobs & Willems, 2018), verschiedene statistische Verfahren in Betracht gezogen, die die Kapazität haben, die nicht-linearen hoch-korrelierten Verbindungen zwischen den Merkmalen mathematisch abzubilden. Um diese Probleme zu lösen, habe ich auf der “quantitative narrative analysis” (QNA) basierende Vorhersagemodelle untersucht (z.B., Jacobs et al., 2017; Jacobs, 2017, 2018a, 2018b). Da es nicht möglich ist, alle relevanten Merkmale eines natürlichen Textes zu identifizieren [so sind bsw. bereits 50 Merkmale für die Verarbeitung einzelner Worte beschrieben worden (Graf, Nagler, & Jacobs, 2005) und über 100 Merkmale für den Korpus an Shakespeare Sonetten

berechnet worden (Jacobs et al., 2017)], und das Einbeziehen von mehr inter/supralexikalischen Merkmalen erfordert auch das Erweitern von Stichprobengrößen (d. h. mehr/längere Texte und mehr Teilnehmer). Daher habe ich mich entschieden, meine Dissertation den so genannten lexikalischen Merkmalen zu widmen. Spezifischer, habe ich sieben Oberflächenmerkmale (Wortlänge, Wortfrequenz, Orthographische Nachbarschaftsdichte, Höher Frequente Nachbarn, Index der Orthographischen Unterschiedlichkeit, Konsonant-Vokal-Quotient, und Sonorität) und zwei affektiv-semantische Merkmale (Valenz und Erregung) untersucht.

Unter der Anwendung von QNA und Vorhersagemodellierung, habe ich drei Blickbewegungsstudien durchgeführt: In Studie 1 (Kapitel 5) habe ich Englische Muttersprachler gebeten, drei Shakespeare Sonette (Sonett 27, 60, und 66) zu lesen, um die Rolle der sieben Oberflächenmerkmale für literarische Texte zu untersuchen. In Studie 2 (Kapitel 6) habe ich ein Paradigma der Lesewiederholung genutzt, um herauszufinden ob die, in Studie 1 bereits identifizierte Rolle der Merkmale, sich in einem weiteren Lesedurchgang verändert. Hierfür haben erneut Englische Muttersprachler zwei Sonette gelesen (Sonett 27 und 66). Studie 3 (Kapitel 7) ist eine Reanalyse der Daten aus Studie 2. Hier habe ich, neben den sieben Oberflächenmerkmalen, zusätzlich zwei semantisch-affektive Merkmale untersucht. Der Fokus lag erneut darauf herauszufinden, ob sich die Rolle der Merkmale über mehrere Lesedurchgänge hinweg verändert. Alle drei Studien zeigen, dass Oberflächenmerkmale das Leseverhalten beeinflussen, und dass die Wichtigkeit der affektiv-semantischen Merkmale im zweiten Lesedurchgang zunimmt. Methodisch gesehen zeigen meine Studien, dass die Vorhersagemodelle (Neuronale Netze) den komplexen Zusammenhang zwischen den psycholinguistischen Merkmalen und dem Leseverhalten (gemessen durch Blickbewegungen) deutlich besser abbilden können, als herkömmliche lineare Modelle. Wiederholtes Lesen der Sonette führte zu einer höheren Lesegeschwindigkeit des Gedichts und auch zu einer tieferen Verarbeitung (z.B., Dixon et al., 1993; Millis, 1995; Kuijpers & Hakemulder, 2018). Es konnte hingegen kein Zuwachs in der Wertschätzung der Sonette beobachtet werden.

Zusammenfassend, kann meine Dissertation einen Beitrag dazu leisten, lineare Modelle in der Leseforschung durch Vorhersagemodelle abzulösen. Die Vorhersagemodelle können das interaktive, nicht-lineare Zusammenspiel der einzelnen psycholinguistischen Merkmale, die das Lesen von natürlichen Texten lenken, wesentlich besser abbilden. Die untersuchten Oberflächenmerkmale beeinflussen das Leseverhalten sowohl beim ersten als

auch beim wiederholten Lesen, während die semantisch-affektiven Merkmale an Wichtigkeit hinzugewinnen zu scheinen, je tiefer ich die literarischen Texte verarbeite. Die Ergebnisse meiner Studien sind äußerst stabil und valide, da ich sie bereits innerhalb meiner Dissertation replizieren konnte. Damit kann meine Dissertation als erster Schritt dahin gesehen werden, ein differenzierteres Verständnis des Lesens von Gedichten zu erhalten und zu einem gedichtspezifischen Lesemodell zu gelangen.

Abbreviations

AOI	Area of Interest
aro	Arousal
cvq	Consonant Vowel Quotient
FI	Feature Importance
fMRI	Functional Magnet Resonance Imaging
GLEC	Gutenberg Literary English Corpus
hfn	Higher Frequent Neighbors
lineNo.	Line Number
LMM	Linear Mixed Models
logf	Word Frequency
LSA	Latent Semantic Analysis
MDBF	German Multidimensional Mood Questionnaire
NCPM	Neurocognitive Poetics Model
odc	Orthographic Dissimilarity
on	Orthographic Neighborhood Density
QNA	Quantitative Narrative Analysis
sonscore	Sonority Score
val	Valence
VSM	Vector Space Model
wl	Word Length
wordNo.	Word Number

List of Original Publications

The dissertation is based on the following articles.

Xue, S., Lüdtke, 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*.

<https://doi.org/10.16910/jemr.12.5.2>

Xue, S., Jacobs, A. M., & Lüdtke, J. (2020). What Is the Difference? Rereading Shakespeare's Sonnets—An Eye Tracking Study. *Frontiers in psychology*, *11*, 421. <https://doi.org/10.3389/fpsyg.2020.00421>

Xue, S., Jacobs, A. M., & Lüdtke, J. (in prep). The Role of Affective-semantic Features in the (Re-)Reading of Shakespeare's Sonnets: A Reanalysis.

Table of Contents

Acknowledgements.....	III
Summary.....	V
Zusammenfassung.....	VII
Abbreviations.....	XI
Table of Contents	XV
I Theoretical Background	1
Chapter 1: Introduction.....	1
Chapter 2: Reading and Rereading.....	5
Chapter 3: Methodology.....	15
Chapter 4: Research Objectives.....	23
II Empirical Part	27
Chapter 5: Reading Shakespeare Sonnets: Combining Quantitative Narrative Analysis and Predictive Modeling — an Eye Tracking Study.....	29
5.1 Abstract.....	29
5.2 Introduction.....	30
5.3 Method.....	36
5.4 Results.....	42
5.5 Discussion.....	44
5.6 Limitations and Outlook.....	46
Chapter 6: What is the Difference? Rereading Shakespeare’s Sonnets — an Eye Tracking Study.....	49
6.1 Abstract.....	49
6.2 Introduction.....	50
6.3 Method.....	56
6.4 Results.....	63
6.5 Discussion.....	69
6.6 Limitations and Outlook.....	73
Chapter 7: The Role of Affective-semantic Features in the (Re-)Reading of Shakespeare’s Sonnets: A Reanalysis.....	75

7.1 Abstract.....	75
7.3 Method.....	82
7.4 Results	85
7.5 Discussion.....	90
7.6 Limitations and Outlook.....	93
III General Discussion	95
Chapter 8: General Discussion and Outlook.....	97
8.1 Methodological point of view.....	100
8.2 Theoretical point of view.....	102
8.3 Limitation and outlook	106
References	109
List of Figures.....	135
List of Tables	139
Appendices.....	141

I

Theoretical

Background

Chapter 1: Introduction

Reading is a significant part of daily life. We draw information from news bulletins, blogs, brochures, biographies, novels, and poetry. Reading behavior is unique to humankind, and not only fosters general education and cross-cultural understanding, social cognition and cognitive development (e.g., Kidd & Castano, 2013; Koopman, 2016; Marr, 2018; Samur, Tops, & Koole, 2018) but also appeals to our feelings and our sense of beauty, especially when reading literature (e.g., Brewer & Lichtenstein, 1982; Nell, 1988; Oatley, 1995). Scientists from diverse fields have long wished to explore the underlying mechanism in one of two ways: One way is the theoretical approach. It involves summarizing or classifying various kinds of literary texts or their features (e.g., prose or poetry; Jakobson & Jones, 1970; Simonto, 1989; Vendler, 1997) and exploring the emotional or aesthetic side of reading without directly measuring the behavior of readers (e.g., Miall & Kuiken, 1994, 2002). The other way is empirical and involves the investigation of texts often designed by the researchers or chosen from newspapers or journals with little emotional or aesthetic appeal. This empirical approach aims to find out how variables such as the length, frequency, or placing of a word affect comprehension.

Standards of experimentation constrain researchers to use materials that belong to “natural (written) language” (e.g., Grainger & Jacobs, 1996; Just & Carpenter, 1980; Kintsch, 1988). Reading research seems to be having trouble to open itself for empirical studies focusing on more natural and ecologically valid reading acts, e.g., literary reading (Jacobs, 2015b; Radach, Huestegge, & Reilly, 2008; Wallot, Hollis, & van Rooij, 2013). There may be two main challenges ahead.

Firstly, there is a need to cope with the complexity of texts produced in ‘normal’ situations. Processing a word depends on more than 50 lexical and sub-lexical features (Graf et al., 2005), and words combine to make phrases, sentences, stanzas or paragraphs with many features of their own (Jacobs, 2015a, 2018a) like rhetorical ploys (cf. Lausberg, 1960). It is hard enough to deal with the features qualitatively, as shown for instance by wrangles about the classification of metaphors and similes (Schrott & Jacobs, 2011), so the road to quantification may prove to be long and bumpy. The artificial text made by researchers, so called “textoids” (Graesser, Millis, & Zwaan, 1997) is an often-used stimulus material since

this facilitates experimental investigation, and thus, provides the chance for clear statements regarding the effect of predefined variables. As a consequence, most empirical studies focused only a few selected features while ignoring the others (e.g., Rayner et al., 2001; Reichle, Rayner, & Pollatsek, 2003; Rayner & Pollatsek, 2006; Engbert et al., 2005; Reilly & Radach, 2006; Rayner, 2009). However, the comparison to natural text lacks due to the identification of more than 100 features on different hierarchical levels, which may influence processing a text (Graf, Nagler, & Jacobs, 2005; Jacobs, 2015a, 2018a; Jacobs et al., 2017).

Secondly, there is a need to consider non-linear interactions. Diverse researchers agree that all these features influence the reading and interpretation of literary texts in a very interactive and non-linear way. The huge number of variables and the nature of their interactions defy the tools of classical statistics used for instance in the field of psychology. Already in 1982, Kliegl, Olson, and Davidson pointed out that standard accounts of reading in terms of hierarchical regressions do little justice to intercorrelated predictors. So far used (inter) correlation thresholds did not solve the problem, but at least made it manageable (e.g., Balota & Chumbley, 1984). However, standard linear models cannot meet the issue regarding non-linear interactions.

The present work aims to jump in at the deep end by using texts with a literary appeal, and thus, numerous non-linear interactive features. For the sake of consistency and easier comparison, texts were taken from a single genre and had the same style and aim. To put it more technically, they had a similar form, cognitive processing, and social function (Berkenkotter & Huckin, 1993; Freedman, 2003). So far, most empirical studies of natural reading have been based on texts which in libraries would be found in the fact, not fiction section, so poetry has been marginalized. The few studies on poetry have been qualitative and have focused more on the text than on the reader. The aim has been to classify poetry (Hanauer, 1995; Hoffstaedter, 1987), to note the conventions of reading it (Fairley, 1986; Thorne, 1989), to find the effects of certain devices on processing it (Hoorn, 1996, van Peer, 1989, 1990) and how it creates meaning (Hanauer, 1996). It has even been suggested that works of fiction are one thing and works of fact another (R. Jakobson, 1960a), so reading a literary text is special in various ways and has notable benefits (Brewer & Lichtenstein, 1982; Nell, 1988).

For this dissertation, the texts chosen are sonnets by Shakespeare, as their quality and popularity are beyond question. His sonnets appeared in 1609, consist of about 17 000 words

altogether and have inspired countless literary reviews and scientific studies (Jakobson & Jones, 1970; Simonto, 1989; Vendler, 1997). Most of his sonnets were written in iambic pentameters with ten syllables per line, and the lines are grouped by rhyme into three verses with four lines each, followed by two lines. In other words, there are three rhymed quatrains followed by a rhymed couplet. The sonnets are much the same in form and different in content, so what more could a (scientific) researcher hope for? Moreover, all 154 sonnets have been extensively subjected to quantitative narrative analysis (QNA). In contrast to qualitative analysis, QNA tries to quantitatively describe the psycholinguistic features of complex natural verbal materials. In my present work, machine learning tools were used to classify them and to predict how eyes would scan them (Jacobs et al., 2017).

Similar texts may be processed by our minds in similar ways, but the pleasure is not always immediate (Dixon et al., 1993), as pointed out much earlier by Professor Korf in a poem by Christian Morgenstern (1871-1914):

Korf erfindet eine Art von Witzen,
die erst viele Stunden später wirken.
Jeder hört sie an mit Langerweile.

Korf invents a novel kind of humour,
letting jokes fall flat for many hours.
Everyone agrees that they are boring.

Doch als hätt ein Zunder still geglommen,
wird man nachts im Bette plötzlich munter,
selig lächelnd wie ein satter Säugling.

But, as if a fuse were faintly glowing,
listeners begin to beam at bedtime,
smiling blissfully like sated sucklings.

Likewise, the first reading of a literary text may light a mental fuse, whose effects become fully apparent only after rereading and reflection. This may be as true of Shakespeare's sonnets as it is of Professor Korf's jokes, as implied by resource allocation theory (Millis & Simon, 1994; Millis, Simon, & TenBroek, 1998). After a first reading, readers may be able to free more resources for high-level cognitive processes needed to explore the labyrinths of Professor Korf's wit or Shakespeare's puns (Britton et al., 1983).

Eye-tracking is the method commonly used for studying the natural reading and experience of literary texts (Carrol & Conklin, 2014; Dixon & Bortolussi, 2015; Jacobs et al., 2016; van den Hoven et al., 2016) including poetry (e.g., Carminati et al., 2006; Lauwereyns & d'Ydewalle, 1996; Müller et al., 2017; Sun, Morita, & Stark, 1985), as the movements of the eyes reflect those of the mind (e.g., Rayner, 1998; Rayner et al., 2006), so it was chosen as the main method for this dissertation.

As for the statistical approach, two steps must be undertaken: Firstly, complex natural texts were broken up into measurable and testable features by “turning words into numbers” (Franzosi, 2010) for the sake of statistical analysis. Secondly, statistical ways were sought to deal with the non-linear webs of correlations, which has long been a concern of Jacob’s working group (e.g., Willems, 2015; Willems & Jacobs, 2016; Jacobs & Willems, 2018).

Recently, the field of natural reading has been explored through predictive modeling based on QNA, which involves using databases, linguistic corpora, and computer programs to quantify psycholinguistic features and also involves using predictive modeling approaches, such as neural nets or bootstrap forests, to find out how some psycholinguistic features affect the response parameters (e.g., Xue et al., 2019; Xue, Jacobs, & Lüdtke, 2020). This was also the approach used for the present dissertation.

All in all, this dissertation records how naive readers read and reread Shakespeare’s sonnets (Chapter 5, Chapter 6, Chapter 7). With the help of predictive modeling based on QNA, we can work out many psycholinguistic features and check their roles in influencing the movements of readers’ eyes. We can also compare eye-movement in one reading session with those in another, to find out whether they are in line with resource allocation theory (Chapter 6, Chapter 7). In the section below, I give an overview of the concepts and findings underlying the present study as regards conjectures, methods, and aims, then I move on to my empirical findings. In the last section, I discuss more broadly the insights gained and finally offer future perspectives.

Chapter 2: Reading and Rereading

Shall I compare thee to a summer's day?
Thou art more lovely and more temperate:
Rough winds do shake the darling buds of May,
And summer's lease hath all too short a date;
Sometime too hot the eye of heaven shines,
And often is his gold complexion dimm'd;
And every fair from fair sometime declines,
By chance or nature's changing course untrimm'd;
But thy eternal summer shall not fade,
Nor lose possession of that fair thou ow'st;
Nor shall death brag thou wander'st in his shade,
When in eternal lines to time thou grow'st:
 So long as men can breathe or eyes can see,
 So long lives this, and this gives life to thee.

William Shakespeare, Sonnets 18

What happens if we read a natural text in the sense of one not written for research purposes? In the case of the sonnet above, we may have to read it word by word, line by line, stanza by stanza, to get a rough idea of what Shakespeare wished to say and why, then we may read it a second time, to explore more possibilities. The words and thoughts have many facets, whose effects may differ from reading to reading. The effects of rereading on the one hand and the interplay of facets on the other are bound to be themes of research, so let us look at them more closely.

2.1 Reading as something of a feat

Indeed, reading is an artificial cultural achievement that strongly relies on the re-use of domain-general processes (Dehaene & Cohen, 2007) related to action (Pulvermüller, 2005) and emotion (Jacobs et al., 2016; Jacobs, 2015b; Ponz et al., 2013; Ziegler et al., 2018). To have a successful reading experience, multiple linguistic and cognitive processes need to be

combined and orchestrated, for instance, orthographical, phonological, morphological, semantic and syntactical information processing, global text comprehension and affective-aesthetic processes (e.g., Hofmann & Jacobs, 2014; Perry, Ziegler, & Zorzi, 2007; Price, 2012). Indeed, reading is a complex higher-order cognitive activity unique to human beings. Literacy is a key qualification in everyday life, academic education, and career (Kendeou, McMaster, & Christ, 2016). The visual and spoken language system and the neural pathways that link them undergo major developmental changes both in structure and function (Dehaene et al., 2015). Reading comprehension also requires word reading ability, working memory, inference generation, comprehension monitoring, vocabulary, and prior knowledge (Perfetti, Landi, & Oakhill, 2008). To have a better understanding of this multifaceted phenomenon, I will first present the current status of theories of reading.

2.1.1 Classical models of reading

Reading is a highly fascinating, complex, and oblique process (Price, 2012). Thus, it is not surprising that researchers from different disciplines try to approach the essential question: What happens while reading? On single word level several neurocognitive (dual-route model, see Hickok & Poeppel, 2007; interactive activation model, e.g., McClelland & Rumelhart, 1981; Price & Devlin, 2011; Seidenberg, 2005, 2007) and computational (multiple read-out model [MROM], see Jacobs et al., 1998) models have been established. The MROM is based on the interactive activation model, which has successfully predicted many phenomena of human perception (e.g., Jacobs, Graf, & Kinder, 2003). It postulates word recognition is based on three information dimensions—single-word-detector activity, total lexical activity, and time from stimulus onset. The response in a given experimental task is generated (read out) when at least one of the information dimensions appropriate for responding in that task reaches a critical level.

Though word recognition is a fundamental aspect of reading, it does not stop at the single word level, in contrast, it is just the starting point. Conceptualizing reading comprehension of sentences, paragraphs or whole texts is an even greater intellectual challenge for scientists. Not surprisingly, many different models have been proposed in the last decades (e.g., Gough & Tunmer, 1986; McNamara & Magliano, 2009; Perfetti & Stafura, 2014; Ahmed et al., 2016; Oslund et al., 2016, 2018). These models try to portray the relationships and interactions during reading by partitioning the complex process into smaller (testable) pieces and identify the underlying components that influence reading

comprehension. The *Simple View of Reading* (Hoover & Gough, 1990), for example, greatly simplifies the reading process. According to this view, reading comprehension is simply the product of different linguistic decoding processes (i.e., phonology, orthography, lexicon) and language comprehension (i.e., semantic knowledge, inferences).

Kintsch and van Dijk (1978), however, examined the process of reading comprehension and proposed his highly influential construction-integration (CI) model. Coarsely, the model encompasses two phases. During the reading, information from the text is extracted and automatically combined with the reader's general knowledge (construction phase). In the integration phase, the information is incorporated into a bigger context (and an abstract situation model is established. Both processes occur iteratively while reading, i.e., the integrated so-called situation model is constantly updated and modified. The construction phase itself encompasses two levels (Kintsch, 1988, 1998; van Dijk & Kintsch, 1983). At first verbatim information is directly drawn from the text, i.e., words and phrases. Here, perceptual processes are involved as well as lexico-syntactic encoding on a purely linguistic level (surface level). Afterward, the semantic analysis of the propositions of the text determines the meaning of the text. Integrating these sources of information, the microstructure of the text, based on the words and their syntactic relationships, is constructed and first simple local inferences are drawn. The microstructure is then integrated into a higher-order global representation of a whole paragraph or section, i.e., the macrostructure, to identify the global topics and interrelations in the text. Together, the micro- and macrostructure form the so-called *textbase*. In sum, the *textbase* represents what is actually expressed in the text, i.e., the processing level is still quite shallow.

To gain a deeper and comprehensive insight into the text, the information needs to be integrated into a bigger picture. More specifically, the context information needs to be integrated with relevant prior knowledge to form a coherent representation of the overall theme or goal. To achieve that, the reader constructs a *situation model*. As stated before, the *situation model* is constantly updated according to the incoming information as the reading process progresses and new inferences from the micro- and macrostructure are drawn. While it is widely agreed that the construction of a *situation model* is essential for truly comprehending a text. The nature of the *situation model* has been debated. In the early versions of the CI model, Kintsch (1988, 1998) assumed that the *situation model* is fully propositional, just like the *textbase*. These proposition-based accounts have been quite successful as they could explain several behavioral phenomena. For example, it could be

predicted what information will be available during different stages of the reading process (see Gernsbacher & Kaschak, 2013, for an overview). Within the vein of the embodiment movement (e.g., Zwaan, 2014; Meteyard et al., 2012; Andrews, Frank, & Vigliocco, 2014) an alternative representation model has been proposed. In this view, the *situation model* might be grounded in sensorimotor simulation and thus have an analogical nature (e.g., Johnson-Laird, 1983, for an early example; Barsalou, 1999; Zwaan, 2004). There is growing evidence for the analogical approach both from behavioral (e.g., Meteyard, Bahrami, & Vigliocco, 2007; Zwaan & Taylor, 2006) and neuroimaging studies (e.g., Buccino et al., 2005; Glenberg et al., 2008; Speer et al., 2009). The question of the representational format of the *situation model* is still not fully resolved but it might not be restricted to the abstract linguistic, i.e., propositional, domain, analogical or a mix of both but might rather be multimodal involving imagery, emotions, and personal experience (Gernsbacher & Kaschak, 2013; Kintsch & Rawson, 2005). According to Zwaan, Magliano, and Graesser (1995) and Zwaan and Radvansky (1998), the *situation model* might incorporate five dimensions representing information about the protagonist; the time of the event; the spatial relations of characters, events, and objects; causality, and the intentionality and goals of the protagonist. These dimensions might guide the reading process and serve as anchor points during comprehension.

However, the CI lacks assumptions on how different types of texts might influence the construction of the different representational levels and the nature of the respective *situation model*. This is despite the fact that different text genres substantially differ concerning linguistic features such as vocabulary, stylistic means, or syntactic complexity.

2.1.2 Models of Literary reading

When reading research focuses on genre, literary reading, especially poetry reading was widely skipped due to issues when it comes to questions like poetry categorization, conventions of poetry reading and poetry specific effects on processing and meaning construction (for further details, see Hanauer, 1998a).

Following Jakobson's (1960b) formalistic view, poetry guides the reader to linguistic patterns of the text rather than pointing to the content of the message like a research paper. When Jakobson (1960b) characterized poetry, poetry-specific properties are described. These include the importance of the internal structure. Furthermore, the meaning is more used in an ambiguous, associative, and structurally non-linear way. These stylistic elements are uniquely

found in literary texts as the aim is to reach a high aesthetic standard and create a literary world rather than simply providing information. A second important focus is on linguistic features. Thus, when incorporating Jakobson's assumptions into Kintsch' (1988, 1998) and van Dijk's CI model and claims regarding textual schema (Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983), one could hypothesize, that the reader, during its literary process (Schmidt, 1989), built a *situation model* that differs from other genres and thus might be poetry specific. This could mean that the literary *situation model* might not necessarily be propositional but might rather reconstruct information according to the poetry's specific patterns (e.g., regularities, similarities, and repetitions; Hanauer, 1998b). However, besides the readers' need to make meaning out of what was read, Schmidt (1982, 1989) defined conventions for the structure and properties of fictional texts. The two most important conventions of these for the present dissertation are the *aesthetic convention* and the *polyvalence convention* (Schmidt, 1989). More specifically, *aesthetic convention* means a high density of textual specific patterns like repetitions similar as described by Hanauer (1998b) described. These patterns can range from the phonological, semantic up to the morpho-syntactic level (see also Figure 2.1.2, the 4 x 4 matrix). The *polyvalence convention* describes the occurrence of semantically ambiguous words or phrases, like metaphors or similes.

However, Hanauer (1998b) found evidence in line with van Dijk and Kintsch's (1983) assumptions regarding differences of text processing based on their textual schema, which are based on the readers' experience with different genres, media, and texts. Textual schemata may guide the reading process as the reader makes specific predictions about form and content. In fact, reading poetry leads to other behavioral outcomes (higher level of recall, lower reading rate, and lower understanding) compared to encyclopedic texts. For instance, Hanauer (1998b) found evidence that the different perceptual appearance of prose compared to poems influence verbatim recall. Hanauer (1998b) and Fecino, Jacobs, and Lüdtké (2020) found interactions regarding the presentation mode of a text (poem vs. prose). Thus, one can assume that not only textual aspects affect the reader's literary experience, but also contextual variables like predictions about form and style of the text read, which together facilitate or complicate the process of building a situation model.

There to the neurocognitive poetics model of literary reading (NCPM; Jacobs, 2011, 2015a, b; Nicklas & Jacobs, 2017; Willems & Jacobs, 2016) is built on. The NCPM is the first model offering concrete assumptions, regarding behavioral responses of readers on

literary text reception and thus tries to capture the whole reading process on all relevant levels. As illustrated in Figure 2.1.1, the model comprises three components, which interplay in various ways: context, reader, and text. Regarding the text level, the NCPM points to two different reading routes, a fast and automatic route, and a slower route. Both routes contain different densities of so-called back- and foregrounding elements. Backgrounding elements are necessary for building the situation model, including the repertoire of familiar words, contexts, or themes. Foregrounding elements cause defamiliarization effects, including rhetorical devices like alliteration, rhyme, inversion, ellipsis, metaphor, or irony. These are hypothesized to elicit different outcomes on three levels of observation: experiential (e.g., ratings; Jacobs, 2017; Jacobs et al., 2015, 2016, 2017; Jacobs & Kinder, 2017, 2018; Jacobs & Lüdtke, 2017), behavioral (e.g., eye movements; Xue et al., 2017) and neuronal (Hsu, Conrad, & Jacobs, 2014; Hsu, Jacobs, & Conrad, 2015; Hsu et al., 2015a, b). Furthermore, it is hypothesized, that fore- and backgrounding elements are differently processed. For example, the prose may comprise a lower density of foregrounding elements and is thus, faster read, which can be observed by fluent reading. Compared to reading prose, poetry reception is assumed to contain a higher density of foregrounding elements and thus, elicit dysfluent reading (Mukarovský, 1964) which can be operationalized by longer reading time. Why fluent and dysfluent reading? It is hypothesized that foregrounding elements slow down automatic reading, as they violate the constructed situation model, demand more attentional resources resolving foregrounding elements, and thus lead to more consciously (Hanauer, 1998b), but also slower reading.

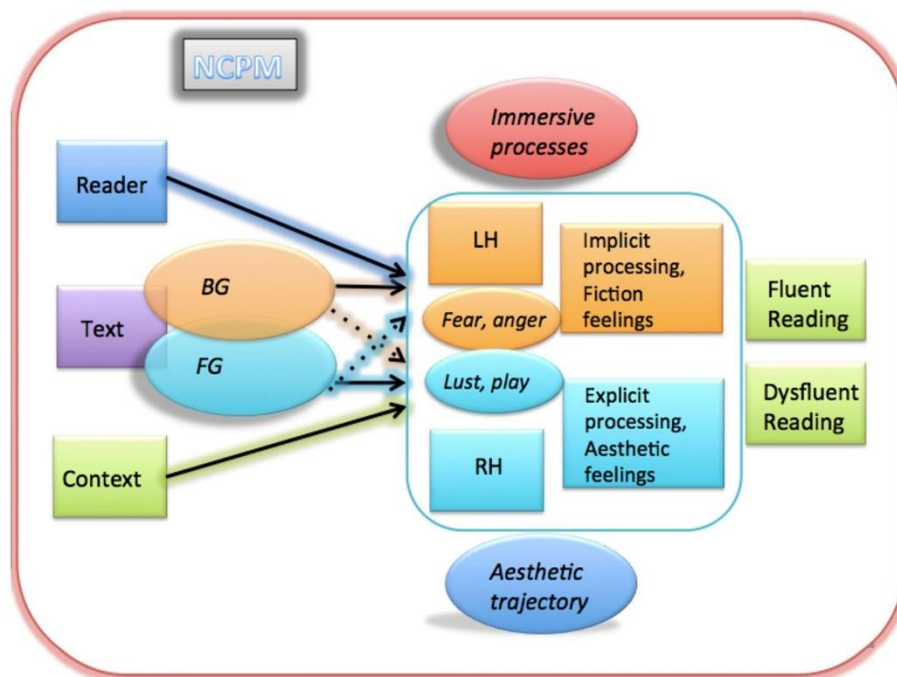


Figure 2.1.1 The Neurocognitive Poetics Model (NCPM) by Jacobs (taken from Jacobs, 2015a)

However, the previous hypothesis gets more precise, when looking closer at the foregrounding elements. These comprise numerous features, which were systematized within a 4x4 matrix. On the x-axis, the feature groups (phonological, semantic, morphosyntactic, and metric) are arranged according to the methodology used by Jakobson and Lévi-Strauss (1962) on the French poem ‘Les chats’ by Charles Baudelaire. On the y-axis four text levels are arranged according to the size of its letter unit, comprising the sublexical level (e.g., phonemes, syllables), the lexical level (e.g., semantic), interlexical level (e.g., sentence or line unit) and supra-lexical level (e.g., whole poem). All combinations of the x- and y-axis describe different features, which influence the reading process. Many of these features are already described, calculated, and analyzed by QNA (see Chapter 3, 3.2). This systematization provides a framework for research on natural literary reading and already influenced several studies ranging from single words and proverbs up to poems (Aryani et al., 2016; Jacobs & Lüdtke, 2017; Jacobs et al., 2015, 2016; Jacobs, Hofmann, & Kinder, 2016; Ullrich et al., 2017).

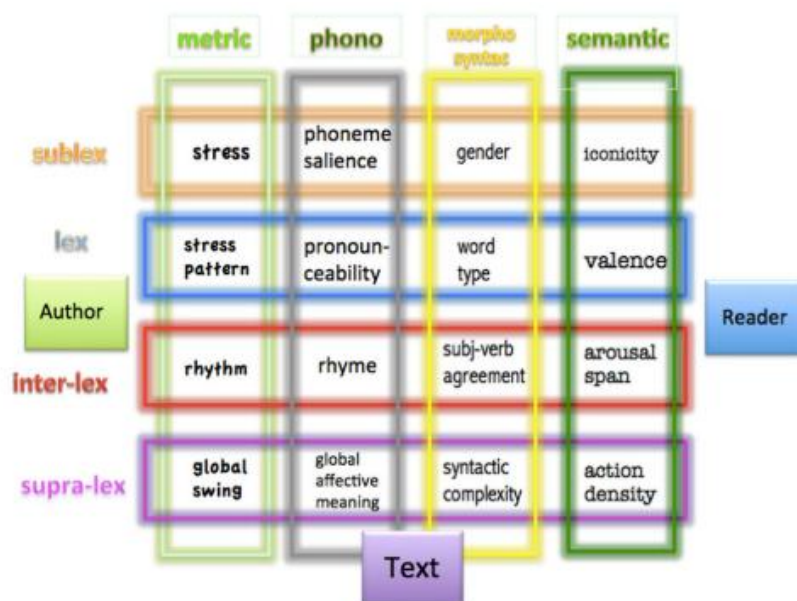


Figure 2.1.2 The 4×4 matrix illustrating four levels of text crossed with four groups of features, with one feature as an example for each cell of the matrix (taken from Jacobs, 2015a)

However, as Kliegl, Olson, and Davidson (1982) already pointed out, research is faced with highly intercorrelated variables and thus, a definition of one feature may rely on the definition of another like *emotional potential*, which is defined as the product of *valence* and *arousal* (e.g., Hsu et al., 2015a; Lüdtke & Jacobs, 2015; cf. Chapter 3, 3.3, for a description of these features). On the other hand, some features work dependently with the other features in influencing the reading behavior. For instance, in the case of sentences, Scott, O'Donnell, and Sereno (2012) found the processing advantage of negative words dependent on word frequency, that is, emotional words (positive or negative) were read consistently faster than neutral words except in the case of negative words with high frequency. Strain, Patterson, and Seidenberg (1995) found an interaction between *word frequency* and *imageability*: words conjuring images up are read faster but only if they are not used often, so no feature is an island unto itself. Thus, the features' effects can best describe by non-linear dynamic functions (see Chapter 3, 3.3).

2.2 Rereading as a reallocation of resources

People tend to read a text more than once, either when they have difficulty to grasp the main points or when they want to deepen their appreciation. A rereading paradigm from

Hyönä and Niemi (1990) has been used in a few studies in various domains (e.g., by Levy, Masson, & Zoubek, 1991; Raney, Therriault, & Minkoff, 2000; Schnitzer & Kowler, 2006; Kaakinen & Hyönä, 2007). How a text is read and reread is assessed during or after each session, be it for instance by eye tracking or self-assessment. Most studies of rereading have been based on expository texts as sources of information for readers to take in and process so have tended to focus on whether or not a reader recalls and grasps more after rereading. The two significant findings on the basis of such rereading paradigms are firstly, rereading increases the reader's recall rate (Amlund, Kardash, & Kulhavy, 1986; Durgunoğlu, Mir, & Ariño-Martí, 1993); and secondly, rereading increases understanding (Rawson, Dunlosky, & Thiede, 2000; Raney, Therriault, & Minkoff, 2000; Brown, 2002; Schnitzer & Kowler, 2006; Kaakinen & Hyönä, 2007; Margolin & Snyder, 2018). While many studies have confirmed the benefit of rereading, only a few of them have investigated the cause. Two researchers worked on a whole theory trying to explain rereading effects. Millis and Simon (1994) offered the Resource Allocation Theory. According to this theory, surface representations (e.g., lexical access) shall be processed automatically and obligatorily. Whereas, high-level features like affective-semantic features, might need extra resources and thus could be processed well in a later reading session.

This could be interpreted as readers may have spare vigor to the “hot” affective-semantic aspects of the poem only after the first reading. Words in texts have to be read and grasped, so the roles of surface features may be consistent across readings, but feelings and deeper implications may or may not be explored, according to interest and effort. Such exploration involves higher-level processes such as emotional empathy, conceptual integration, and elaborative inferences (Millis, Simon, & TenBroek, 1998). After a first reading, less attention may be paid to lower levels and more to higher ones, especially in reading literature, as implied by the finding of Britton et al. (1983) that literary texts call for more processing than do expository ones.

Studies of rereading have seldom been based on literary texts and have mostly involved assessments made after the reading, revealing such classical effects as enhanced comprehension (e.g., Klin, Ralano, & Weingartner, 2007; Kuijpers & Hakemulder, 2018). Especially in the case of literary texts, researchers have wondered whether rereading affects appreciation and aesthetic reactions. They have surmised that these are related to the extent of comprehension (Kuijpers & Hakemulder, 2018). In line with this surmise, the scant studies using literary texts found that rereading does indeed influence readers' appreciation, insofar

as readers tend to like texts more after rereading (e.g., Dixon et al., 1993; Millis, 1995; Kuijpers & Hakemulder, 2018), as shown also by the only study on the rereading of poetry (Hakemulder, 2004). Nevertheless, no study using literary texts has checked the cognitive and emotional processes of comprehension and appreciation. However, whether a literary text is read more fluently the second time is still an open question.

Hence another aim of the present dissertation is to examine the effects of rereading poetry by using not only assessments made by readers *after* the sessions but also records of eye movements made *during* the sessions, to find out whether rereading improves their poetry understanding, appreciation and reading fluency. A further aim is to find out whether surface features, like *word frequency*, play a role in changing the eye tracking parameters across reading sessions.

Chapter 3: Methodology

3.1 Eye tracking technique

Eye tracking is widely used in reading research (e.g., Just & Carpenter, 1980; Hyönä & Hujanen, 1997; Rayner, 1998; Rayner et al., 2006; Clifton, Staub, & Rayner, 2007), to track the process of reception indirectly. Just and Carpenter (1980) pointed out, the movements of the eyes reflect those of the mind, so they also reveal the challenges posed by features of a text. Several eye tracking parameters have been defined for the use in analysis, those most often reported being the first fixation duration, the gaze duration, the regression time, and the total reading time.

In reading studies, the first fixation duration is the duration of the first fixation on a certain word; the gaze duration is the sum of all fixation durations on a certain word during the first passage; the regression time is the sum of all fixation durations after the first passage; the total reading time is the sum of all fixation durations on a certain word.

The first fixation duration and gaze duration are thought to belong to the early stages of language processing, less influenced by lexical parameters, as words are identified swiftly and automatically (Hyönä & Hujanen, 1997; Clifton, Staub, & Rayner, 2007). Compared to that, regression time belongs to late stages, as trickier parts of a text are reread, to be reprocessed. The total reading time includes the first processing and reanalysis. The regression time and total reading time may reflect the processing of higher-level linguistic variables (Clifton, Staub, & Rayner, 2007), but no eye movements have yet been specifically linked to a certain phase of cognitive processing (Pickering et al., 2004; Rayner & Liversedge, 2011).

The length and frequency of words influence eye movements, as shown by the fact that words which are shorter and high frequent are fixated shorter and less often than words which are longer and have a low frequency (e.g., Just & Carpenter, 1980; Inhoff & Rayner, 1986; Raney & Rayner, 1995; Pynte, New, & Kennedy, 2008). Scott, O'Donnell, and Sereno (2012) found shorter fixations of emotional words than neutral words during first pass reading. However, its implications on literary reading are widely not understood, since only two studies were done by using eye tracking on literary reading (see, Müller et al., 2017, for a

study on haiku and van den Hoven et al., 2016, as an example for a study on prose). The same applies to the effects of features like *orthographic neighborhood density*, *orthographic dissimilarity*, *consonant vowel quotient*, and *sonority score* on eye movements in reading since no researchers have checked them systematically at the same time. To investigate the relation of these features within eye tracking studies is the main goal of the present dissertation.

Researchers have also considered whether rereading expository texts increases fluency by letting less time be taken on reading single words or the whole text and have come to the following findings: After a first reading, less time is taken to read the whole text (Millis & King, 2001). Likewise, most eye-tracking parameters on the level of words improve the total reading time is reduced, as is the regression time (the sum of fixations on a certain word after the first reading), and the rate of skipping is higher (Hyönä & Niemi, 1990; Raney & Rayner, 1995; Raney, Therriault, & Minkoff, 2000; Kaakinen & Hyönä, 2007).

In general, the rereading benefits may be due to a change in the roles played by lexical, interlexical, or supralexic features. Levy and his colleagues (1992, 1993) surmised that readers may even benefit from reading one text before reading another with a similar meaning or context. They checked by replacing some of the words with synonyms and by changing the syntactic structure, and the results were positive. Raney, Therriault, & Minkoff (2000), however, found that replacing some words with others of similar meaning, only shortened the gaze duration and the total reading time on certain words. Therefore, they concluded that rereading had a stronger influence on later processing stages compared to early ones.

To clarify the role of some lexical features in the rereading of expository texts, Raney and Rayner (1995) manipulated the frequency of words, but the decrease in fixation durations was the same for low- and high-frequency words across readings. Likewise, Chamberland et al. (2013) found that the benefit of rereading was the same for content and function words and for low- and high-frequency words, except that the second time the duration of gaze stronger reduced for function words than for content words, though some studies have found that low-frequency words benefit more than others from multiple readings (see Kinoshita, 2006, for a review). All in all, the effects of rereading on eye tracking parameters in the early stages of the process (e.g., on gaze duration) have been inconsistent, especially in the case of some psycholinguistic features, which are thereby in need of further investigation.

3.2 Quantitative narrative analysis

As shown in Figure 2.1.2, several psycholinguistic features may influence the process of natural reading. To examine their particular impact on reading empirically, they have to be quantified. A text feature may be defined as objective and identifiable if we can specify a rule for checking its presence in a given text. The quantitative narrative analysis (QNA) provides a framework to turn these psycholinguistic features into numbers (e.g., Jacobs, 2015a, 2017, 2018a, 2019; Jacobs, Hofmann, & Kinder, 2016; Jacobs et al., 2017; Jacobs & Kinder, 2017, 2018). QNA is based on the 4 x 4 matrix (described in Chapter 2, Figure 2.1.2) with its different text levels and groups of features. In the last decades, the different cells of the matrix have been extensively examined on different levels of analysis. For instance, information on different lexical features (e.g., arousal rating) have been used to build databases such as BAWL (Vö, Jacobs, & Conrad, 2006; Vö et al., 2009), DENN-BAWL (Briesemeister, Kuchinke, & Jacobs, 2011), kidBAWL (Sylvester et al., 2016) and ANGST (Schmidtke et al., 2014). Other studies have taken a multi-level approach and examined the role of features from different levels (Hsu et al., 2015a; Jacobs et al., 2016; Ullrich et al., 2017). While the features of the lexical and interlexical levels are well described, quantifying variables at the supralexical level is still an open issue. Here, to operationalize narrative structure and complexity, the use of appropriate tools from QNA (e.g., Jacobs, 2015a; Franzosi, 2010) or advanced qualitative-quantitative narrative analysis (Q2NA, see Jacobs, 2018a) has been suggested.

3.2.1 Surface features

Surface features define (psycho)linguistic properties on the sublexical and lexical levels. It is relatively easy to specify simple, local features of a text, like the number of words, but features may also be abstract and general. The set of possible features of texts is usually huge [e.g., over 50 features mentioned for single word recognition (Graf et al., 2005) or over 100 features computed for the corpus of Shakespeare's sonnets (Jacobs et al., 2017)]. Nonetheless, it is impossible to identify all relevant features of a natural text, so most empirical studies seem to check only a few features chosen (e.g., Rayner et al., 2001; Reichle, Rayner, & Pollatsek, 2003; Rayner & Pollatsek, 2006; Engbert et al., 2005; Reilly & Radach, 2006; Rayner, 2009), and these include several surface features (e.g., *word length*, *word frequency*, *higher frequency neighbors*, *sonority score* on the lexical and the *number of syllables* on the sublexical level).

The most widely discussed surface features are *word length* and *word frequency*, often used as proxies for word difficulty (Breland, 1996). Mostly, the shorter and more common a word, the faster it is read. These word-length and -frequency effects are often shown by tasks involving word naming, lexical decisions, semantic decisions, and memory (see Barton et al., 2014; Brysbaert, Mander, & Keuleers, 2018, for a review). For instance, as shown by eye-tracking studies with non-literary texts, a word takes more time to be read and is reread more often (e.g., Just & Carpenter, 1980; Inhoff & Rayner, 1986; Raney & Rayner, 1995; Pynte, New, & Kennedy, 2008) the longer and less common it is.

Apart from the research of these basic surface features, there has been a lot of research on *orthographic neighborhood density* (the number of valid words produced by changing a single letter of a certain word). For instance, *cat* can be changed into *bat*, *fat*, *mat*, or *cab* (Coltheart et al., 1977) (see Andrews, 1997, for a review). In general, the higher this density, the more easily a word is read, unless the presence of its *higher frequent neighbors* in the hypothetical mental lexicon inhibits processing of a target word (its neighbors are more common than the targeted word itself) (Grainger et al., 1989; Grainger & Jacobs, 1996; Perea & Pollatsek, 1998), but there are no clear conclusions about the combined effects of both features (*orthographic neighborhood density* and *higher frequent neighbors*) on reading (Williams, Perea, Pollatsek, & Rayner, 2006). Furthermore, by using the Levenshtein distance metric and going beyond the standard operationalization based on words of the same length, we can also compute an *orthographic dissimilarity index*, not yet empirically studied.

While the above features are basically ‘orthographic’, the effects of sublexical and lexical phonological features in some studies of silent reading (e.g., Aryani, Jacobs, & Conrad, 2013; Aryani et al., 2016, 2018; Aryani, Hsu, & Jacobs, 2018; Braun et al., 2009; Schmidtke, Conrad, & Jacobs, 2014; Jacobs, 2015b, c; Ullrich et al., 2017; Ziegler & Jacobs, 1995) and the wide use of phonetic rhetorical devices in natural language have led us to include two phonological features: the *consonant vowel quotient* and the *sonority score*. The former is a simple proxy for how easily a word can be spoken—which in principle is related to how easily it can be phonologically recoded (H.-W. Lee, Rayner, & Pollatsek, 2001). To rank English phonemes in terms of acoustic energy or sonority (Ladefoged, 1993), we have used a *sonority scale*, beginning with the most sonorous (e.g., Clements, 1990) and thereby possibly the most beautiful as subjectively experienced (Jacobs, 2017). The scale has also been used in studying aphasia (Stenneken et al., 2005), and there is evidence that sonority and

the qualities of consonants play a role in silent reading (Maïonchi-Pino et al., 2008; Berent, 2013), especially of poetic texts (Kraxenberger, 2017).

The above surface features that can directly be extracted from text corpora and the target words themselves by the help of computer algorithms, i.e., independent from subjective rating data but purely objective. For this dissertation, we computed these features as follows: The *word length* (*wl*) is the number of letters per word; the *word frequency* (*logf*) is the logarithm of the number of appearances of a word in the Gutenberg Literary English Corpus as a reference (GLEC; Jacobs, 2018b; Xue et al., 2019). The *orthographic neighborhood density* (*on*) is the number of words of the same length as a certain word and differing by only one letter in GLEC; *higher frequency neighbors* (*hfn*) is the number of orthographic neighbors with a higher word frequency than the word in GLEC; *orthographic dissimilarity* (*odc*) is the word's mean Levenshtein distance from all other words in the corpus (GLEC), a metric which can be generalized to apply to words of different lengths; the *consonant vowel quotient* (*cvq*) is the quotient of consonants and vowels in one word; the *sonority score* (*sonscore*) is the sum of the phonemes' positions in the sonority hierarchy, divided by the square root of *wl* (the sonority hierarchy of English phonemes has 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), For instance, in sonnets, the word “ART” got the *sonscore* of $10 \times 1 [a] + 7 \times 1 [r] + 1 \times 1 [t] = 18 / \text{SQRT}(3) = 10.39$.

3.2.2 Affective-semantic features

Till now, we only centered on “cold” superficial features like the lengths and frequencies of words, so it is time to be broadened to include “hot” affective and semantic features. After all, the language may also be meaningful and evocative (e.g., Brewer & Lichtenstein, 1982; Nell, 1988; Oatley, 1995). The affective-semantic features most often studied are *valence* (the degree of positive or negative affect) and *arousal* (the degree of internal activation). Most studies of these two factors have shown that emotional words are more easily processed than neutral words. For instance, highly arousing words are recognized faster in lexical decision-making (e.g., Hofmann et al., 2009; Schacht & Sommer, 2009; Scott et al., 2009) and to be recalled more often (e.g., Kissler et al., 2007).

A further affective-semantic feature often discussed is the *semantic similarity*, as a word may be processed more easily and swiftly if semantically like another in the same context (Roland et al., 2012). Recently, many approaches based on a model of distributional

lexical semantics have been proposed to quantify semantic similarity. These include latent semantic analysis (LSA; Deerwester et al., 1990; Landauer & Dumais, 1997), Bayesian models (Griffiths, Steyvers, & Tenenbaum, 2007) and neural networks (Mikolov, Chen, Corrado, & Dean, 2013). The model of distributional lexical semantics is based on the notion that in a given context the likeness of one word to previous words may be assessed in terms of their interchangeability (co-occurrence) in a big set of texts. These affective-semantic features have yet to be checked in studies of reading literature.

In contrast to the surface features, the calculation of affective-semantic features on the lexical level, like *valence* (*val*) and *arousal* (*aro*), is based on a hybrid method combining the traditional methods based on a dictionary or list of words and the computational method based on a vector space (the VSM or Vector Space Model) (Taboada et al., 2011). We used the training corpus (GLEC) to estimate the semantic similarity of a certain word in the poem to each of 12 word labels (seven positive labels, such as HAPPINESS or PRIDE, and five negative ones such as DISGUST or FEAR; Ekman, 2005; Westbury et al., 2015) for which *valence* & *arousal* rating-data are available. The *valence* and *arousal* value of a certain word is the average of the ratings of its *k* nearest neighbors in the vector space (Jacobs, 2019). This means in this dissertation, word-based *valence* and *arousal* are not entirely based on objective information but also a subjective list of words.

3.2.3 *Interlexical and supralelexical features*

According to the NCPM (Jacobs, 2011, 2015a, b; Nicklas & Jacobs, 2017; Willems & Jacobs, 2016), there are also interlexical and supralelexical (the synergy of words comprising a text) features may play important roles in reading. For instance, Jacobs (2015c) showed that the interlexical feature, *arousal span*, which is produced by the contrast between the word of the lowest arousal and the word of the highest arousal in one sentence, can account for about 25% of the variance in suspense ratings from readers. Moreover, *Surprisal*, as a supralelexical index, can be used to predict the “literariness” of metaphors (Jacobs & Kinder, 2018). However, unlike the lexical and interlexical features, quantifying relevant features at the supralelexical level still presents big challenges because of the lack of similar databases or lists that could provide the relevant information.

3.3 Predictive modeling

As mentioned earlier, machine learning tools, such as neural nets or random forests, can be used in modeling big sets of data with complex interactions and intercorrelations. With the predictive models and computational means now available, we can analyze human cognition, emotion and behavior, such as eye movements, in naturally rich settings (Lappi, 2015) such as literature (e.g., Willems, 2015; Willems & Jacobs, 2016; Jacobs & Willems, 2018).

Given that features of texts influence the reading and interpretation of literary texts in a very interactive and nonlinear way, two machine-learning models (neural nets and bootstrap forests) were compared with a standard least square regression, to find out which was better at predicting the eye movement parameters in reading literary texts. The neural net chosen makes use of an architecture inspired by the neurons in the human brain (LeCun, Bengio, & Hinton, 2015) and was a multi-layer perceptron (one or two layers) able to predict one or more response variables by using a flexible function of the input variables. It can implicitly detect all possible (nonlinear) interactions between predictor variables and perform a dimension reduction on correlated predictors. Therefore, the approach appears advantageous for studies on natural reading in which multiple psycholinguistic and context features may play a role (Jacobs, 2015a, 2018a). Bootstrap forests predict a response value by averaging the predicted response values across many decision trees, each of which is grown on a bootstrap sample of the training data (Hastie, Tibshirani, & Friedman, 2009). All models were evaluated through predictive modeling, comparing the goodness of fit indices (R^2 , misclassification rate, ROC, and AUC) for training and test sets.

Recent developments especially in the fields of bioinformatics (Strobl, Malley, & Tutz, 2009), ecology (e.g., Manel et al., 1999; Were, Bui, Dick, & Singh, 2015), geology and risk analysis (Nefeslioglu, Gokceoglu, & Sonmez, 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), fMRI data analysis (e.g., Cichy et al., 2017) or applied reading research (Lou et al., 2017; Matsuki, Kuperman, & Van Dyke, 2016) highlight the use of machine learning tools, such as neural nets or random forests, in modeling big data sets with complex interactions and -correlations. Predictions based on the models can then be checked.

Thanks to the new models and computational methods, it is now possible to analyze human cognition, emotion and behavior in rich naturalistic settings (Lappi, 2015), so even the movements of eyes over literature can be tracked and analyzed (e.g., Willems, 2015; Willems & Jacobs, 2016; Jacobs & Willems, 2018).

Chapter 4: Research Objectives

4.1 Limitations of Previous Research

Literary (re-)reading happens in our daily lives and has a significant number of benefits. Since the invention of letters or other signs, human beings have created many great and enduring literary works. One of the most successful and popular pieces of verbal art in the world is Shakespeare's works, which have been widely studied in the last 50 years (e.g., Jakobson & Jones, 1970; Simonto, 1989; Vendler, 1997). Although scholars and researchers are very interested in this kind of cultural heritage, nearly all studies on literary works focus on text-based qualitative aspects.

At the same time, in the last 20 years reading has been extensively investigated in empirical studies, e.g., by using eye tracking (Rayner et al., 2001; Reichle, Rayner, & Pollatsek, 2003; Rayner & Pollatsek, 2006; Engbert et al., 2005; Reilly & Radach, 2006; Rayner, 2009). Within the field of eye tracking research, though, single sentences from non-literary materials appear to be the most extensively investigated text material (e.g., Clifton, Staub, & Rayner, 2007; Radach & Kennedy, 2013; Rayner, 2009). Although natural reading takes place most often at the level of longer text units like newspaper articles, short stories or novels, eye tracking 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, Huestegge, & Reilly, 2008; Wallot, Hollis, & van Rooij, 2013). It seems that there is a gap between mainly theoretical qualitative literary studies and empirical research in psychology.

The aim of my three eye-tracking studies of the (re-)reading of Shakespeare's sonnets was to bridge the gap between text-based qualitative analyses and empirical research on literature reading, so as to grasp more of the mechanism used in reading poetry. In all these studies, I applied QNA-based predictive modeling. However, since the number of features is overwhelming, the present dissertation focuses on seven lexical features, which I call 'surface features' (*word length, word frequency, orthographic neighborhood density, higher frequency neighbors, orthographic dissimilarity index, consonant vowel quotient, and the*

sonority score) and two are affective-semantic features (*valence* and *arousal*). In the following section, they are presented in detail.

4.2 Research Questions and Hypotheses

Within the present dissertation, I asked: 1) which predictive modeling approach can be successfully used in the study of poetry reading; 2) what are the potential effects of psycholinguistic features (*word length (wl)*, *word frequency (logf)*, *orthographic neighborhood density (on)*, *number of higher frequency neighbors (hfn)*, *orthographic dissimilarity (odc)*, *consonant vowel quotient (cvq)*, *sonority score (sonscore)*, *valence (val)*, and *arousal (aro)*) on eye movement parameters; 3) do the roles of psycholinguistic features change across reading sessions; and finally, 4) does rereading improve understanding, appreciation on supralexical level, i.e., poem level?

Since only non-linear interactive models can deal with complex interactions and detect hidden structures in complex data sets (LeCun, Bengio, & Hinton, 2015), I assume that predictive modeling approaches would outperform the general linear model and produce satisfactory model fits.

According to the resource allocation theory, the processing of the surface features is automatic and obligatory, thus their roles may be consistent across readings. Since after the first reading session readers may have more resources to analyze the affective-semantic aspects of the poem, the importance of the two affective-semantic features may be increased in the rereading session compared to the first reading session. Former studies had shown that rereading improved readers' comprehension and increased their appreciation of literary texts (Klin, Ralano, & Weingartner, 2007; Dixon et al., 1993; Millis, 1995; Kuijpers & Hakemulder, 2018), so I expect to get similar results with poetry. In other words, I expect that readers would identify the topic better (measured by higher understanding ratings) and appreciate the poem more after the last session.

4.3 Conceptualization of the Empirical Part

Study 1 (Chapter 5) analyzed the eye movements of readers of three of the 154 sonnets as a function of seven lexical features, extracted by QNA. Using predictive modeling based on machine learning, five 'surface' features (*word length*, *orthographic neighborhood density*, *word frequency*, *orthographic dissimilarity*, and *sonority score*) were found to be

important in predicting the total reading time and fixation probability in poetry reading. A phonological feature, the *sonority score*, was likewise found to play a role, as currently surmised. Our findings based on eye movements open new possibilities for future research into the reading of poetic texts and other complex literature.

Study 2 (Chapter 6) was based on two of Shakespeare's sonnets. Eye movements were recorded during the reading of a sonnet then were recorded a few minutes later during a second reading. After each reading, comprehension, and appreciation were measured using a questionnaire. In general, reading was more fluent (shorter total reading times, shorter regression times, and lower fixation probability) and the depth of comprehension was higher the second time of reading. No increase in appreciation was apparent, despite claims to the contrary on the basis of former studies. Moreover, predictive modeling analysis showed that readers' eye movements were determined by the same psycholinguistic features in both sessions, so even in reading poetry, the process highly depends on surface features unaffected by repetition.

Study 3 (Chapter 7) consisted of reanalyzing the data of study 2 regarding the roles played by seven surface psycholinguistic features in poetry reading and rereading. By applying predictive modeling based on QNA, two affective-semantic psycholinguistic features were likewise calculated and added, to predict eye movements. I confirmed that irrespective of how often a poem is read, certain surface features always stand out from other features, but affective-semantic features are crucial after the first reading. Apparently, surface features are basic for understanding, but once the basis becomes clear, readers focus more on affective-semantic aspects.

In the following section, my three studies are explained in greater detail.

II

Empirical

Part

Chapter 5: Reading Shakespeare Sonnets: Combining Quantitative Narrative Analysis and Predictive Modeling — an Eye Tracking Study¹

Shuwei Xue, Jana Lüdtke, Teresa Sylvester, and Arthur M. Jacobs

5.1 Abstract

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

¹ This Chapter is published as "Reading Shakespeare Sonnets: Combining Quantitative Narrative Analysis and Predictive Modeling — an Eye Tracking Study". *Journal of Eye Movement Research*, 12(5). <https://doi.org/10.16910/jemr.12.5.2>

5.2 Introduction

Not marble, nor the gilded monuments

Of princes, shall outlive this powerful rhyme;

William Shakespeare, Sonnets 55 (ll. 1-2)

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, Tops, & Koole, 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, Staub, & Rayner, 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, Hollis, & van Rooij, 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, Huestegge, & Reilly, 2008; Wallot, Hollis, & van Rooij, 2013).

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 that can handle the intercorrelated psycholinguistic features and the nonlinear relationship between them and reading behavior. In the following sections, we expound the two challenges separately, and at the end put forward our hypotheses.

5.2.1 *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 a 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, 2018a) 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's 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., 2015, 2016, 2017; Jacobs & Kinder, 2017, 2018; Jacobs & Lüdtke, 2017), the behavioral (e.g., eye movements; Xue et al., 2017), and the neuronal (Hsu et al., 2015a).

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; Simonto, 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, 2015; Jacobs et al., 2016; van den Hoven et al., 2016; Lauwereyns & d’Ydewalle, 1996; Müller et al., 2017; Sun, Morita, & Stark, 1985) examined the eye movement behavior of Shakespeare’s 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’s 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, Rayner, & Pollatsek, 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 supralexic 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 non-literary texts:

In eye tracking studies of reading non-literary texts it is widely acknowledged that long 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, New, & Kennedy, 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 the 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, the 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, Jacobs, & Conrad, 2013; Aryani et al., 2016, 2018; Aryani, Hsu, & Jacobs, 2018; Braun et al., 2009; Schmidtke, Conrad, & Jacobs, 2014; 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 (H.-W. 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 estimating 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 (Maïonchi-Pino et al., 2008; Berent, 2013), especially of poetic texts (Kraxenberger, 2017). Both features have not been examined in literary reading studies using eye tracking.

5.2.2 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, 2018a; Leech, 1969; Schrott & Jacobs, 2011). Kliegl, Olson, and Davidson (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, Malley, & Tutz, 2009), ecology (e.g., Manel et al., 1999; Were, Bui, Dick, & Singh, 2015), geology and risk analysis (Nefeslioglu, Gokceoglu, & Sonmez, 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, Kuperman, & Van Dyke, 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., Willems, 2015; Willems & Jacobs, 2016; Jacobs & Willems, 2018).

For the 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 poetry. The neural net model is a multilayer perceptron which can predict one or more response variables using a flexible function of the input variables. It can implicitly detect all possible (nonlinear) interactions between predictor variables and many 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 the goodness of fit index (R^2) for training and test 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).

5.2.3 Hypotheses

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

Based on the previous eye tracking studies and existent models of eye movement control (e.g., Engbert et al., 2005; Klitz, Legge, & Tjan, 2000; Legge, Klitz, & Tjan, 1997; Reichle, Rayner, & Pollatsek, 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., long 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*, we hypothesized 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 more exigent processing (e.g., Jacobs et al., 1998; Maionchi-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, Staub, & Rayner, 2007).

5.3 Method

5.3.1 Participants

Fifteen native English participants (five females; $M_{\text{age}} = 31.5$ years, $SD_{\text{age}} = 14.1$, age range: 18–68 years) were recruited from an announcement released at Freie Universität Berlin. All participants had a 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.

5.3.2 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 $1,024 \times 768$ 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.

5.3.3 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 Sonnets'

corpus but also within European poetry. The motifs are: love as a 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's sonnets (see Introduction). Inspired by our previous QNA study on Shakespeare's sonnets (Jacobs et al., 2017), we conducted a fine-grained lexical analysis of all words used in the present three sonnets, summarized in Table 5.1. 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 5.1 Number of Words per Category within Each Sonnet and within all Three Sonnets

Sonnet	Closed-class words		Adjective/ Adverb		Nouns		Verbs		Total word number
	count	%	count	%	count	%	count	%	
	27	49	44.14	20	18.02	28	25.23	14	
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. % is the percentage of each word category within each sonnet or all three sonnets

5.3.4 Procedure

The experiment was conducted in a dimly lit and sound-attenuated room. The data acquisition for each sonnet was split into 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 understanding. Before 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). 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, keywords, 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, Lüdtkke & 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. 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 5.1 for an illustration of the procedure).

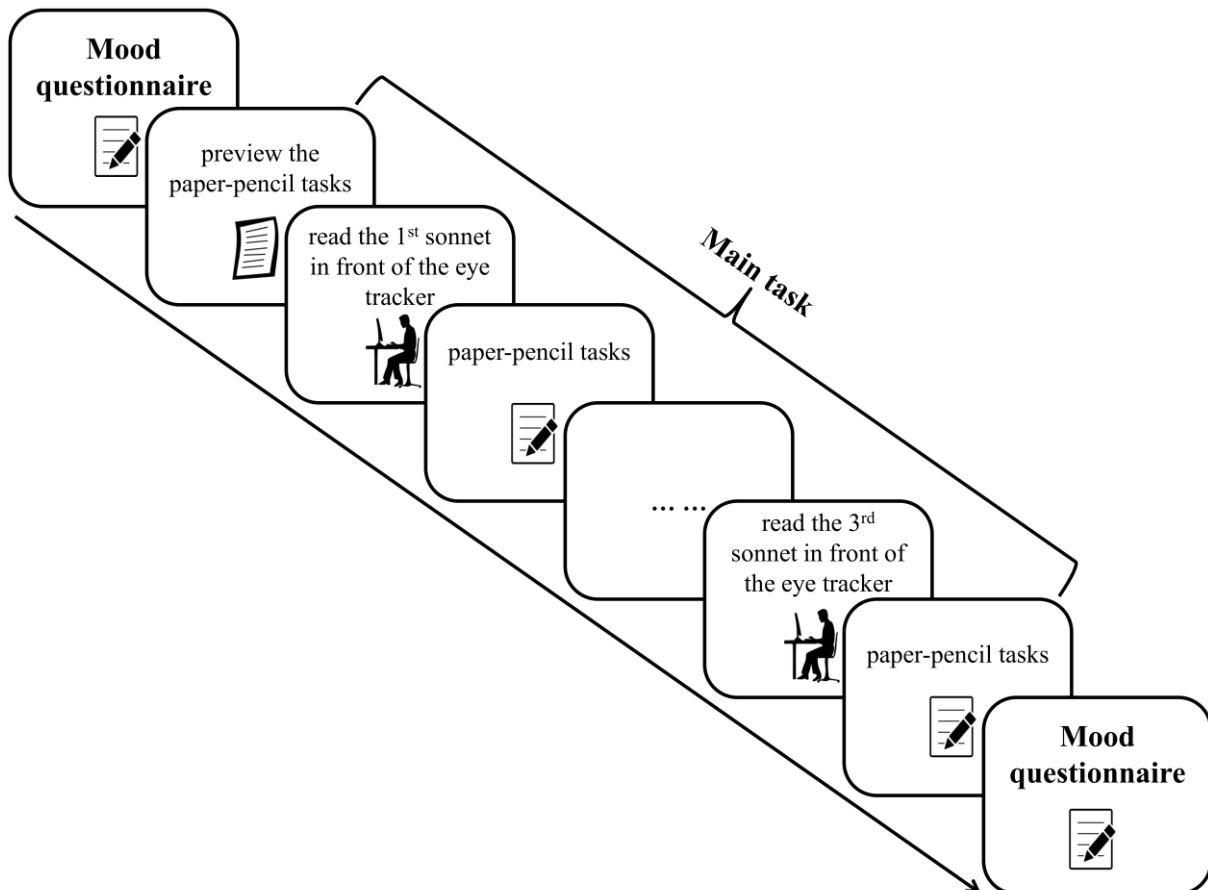


Figure 5.1 The Procedure of the Experiment

5.3.4 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, 2018b): *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* value 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 5.2. 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 5.2 Correlations between Seven QNA Features

Variables	1	2	3	4	5	6	7
1. Word length (<i>wl</i>)	–						
2. Log frequency (<i>logf</i>)	-.75	–					
3. Orthographic neighbours (<i>on</i>)	-.81	.68	–				
4. Higher frequency neighbours (<i>hfn</i>)	-.31	.00	.36	–			
5. Orthographic dissimilarity based on corpus (<i>odc</i>)	.74	-.48	-.39	-.18	–		
6. Consonant vocal quotient (<i>cvq</i>)	.19	-.10	-.24	-.05	.10	–	
7. Sonority score (<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/>). 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 the first fixation on the target word) as a measure of word identification, gaze duration (the sum of all fixations on the target word during the first pass), re-reading time (sum of fixations on the target word after the 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 the 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 ‘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. Because 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 5.3).

Table 5.3 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) was used to run all statistical analyses. The values of all variables (seven predictors and three eye movement parameters) were standardized before modeling. To counter possible overfitting, for all three models we used a cross-validation procedure using 90% of the data as a training set and the remaining 10% as a test set. Given the intrinsic probabilistic nature of two of the models and the limited sample size ($N = 205$ words, i.e., about 20 in the test sets), predictive modeling results varied across repeated runs, depending on which words were selected as training or test 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). 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, Malley, & Tutz, 2009). To make our results better comparable with previous work, we also tested the effects of ‘important predictors’ (FI s $> .10$) 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 the general linear model, i.e., standard least squares regression, got acceptable model fit as described above, instead of reporting FI s 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.

5.4 Results

Figure 5.2 shows the overall mean R^2 s (averaged across 1000 iterations) for the three eye tracking parameters for both the training and test sets using all three modeling approaches. Figure 5.3 shows the seven FI s 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.

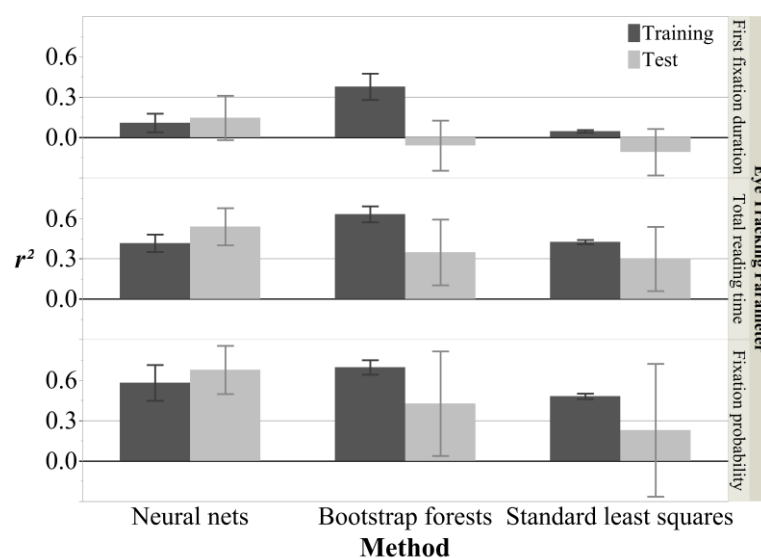


Figure 5.2 Model Fits of Different Measure Groups via Different Modeling Methods

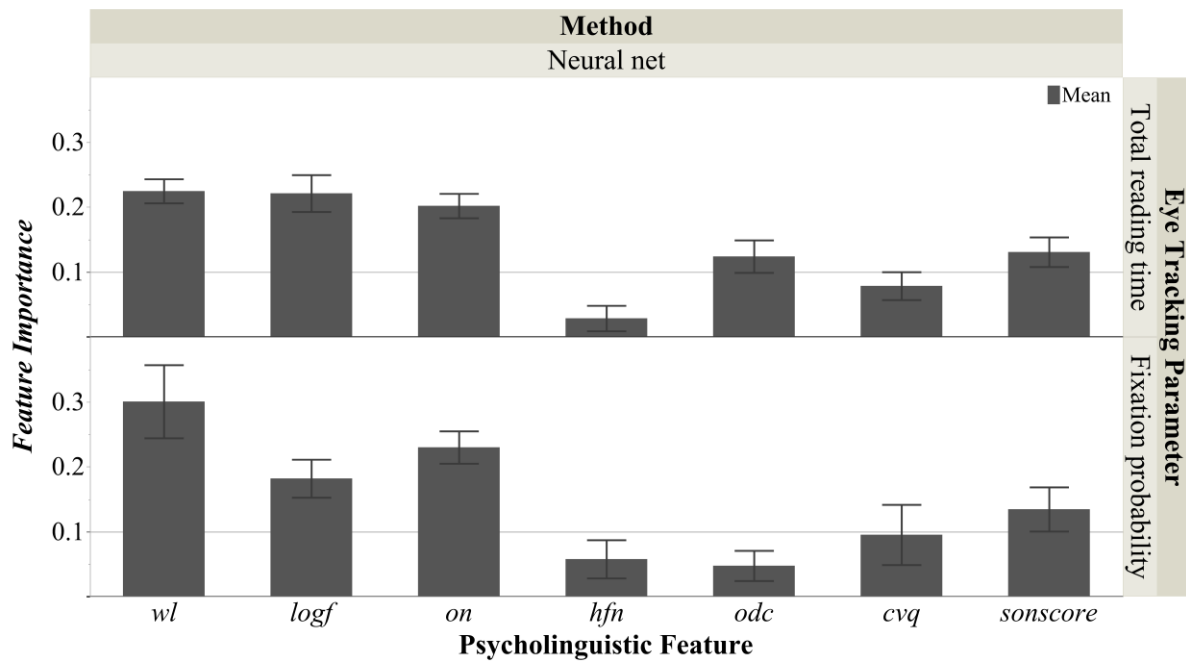


Figure 5.3 Feature Importances for Total Reading Time and Fixation Probability

Mean First Fixation Duration. Figure 5.2 shows that while in the training set (*train*) the bootstrap forests model's fit was satisfactory (mean $R^2_{train} = .38$, $SD_{train} = .10$), it did not generalize to the test set (*test*) at all (mean $R^2_{test} = -.10$, $SD_{test} = .19$). The neural nets model and standard least squares regression also showed poor fits for both training (neural nets: mean $R^2_{train} = .11$, $SD_{train} = .07$; standard least squares: mean $R^2_{train} = .05$, $SD_{train} = .01$) and test set (neural nets: mean $R^2_{test} = .15$, $SD_{test} = .16$; mean $R^2_{test} = -.10$, $SD_{test} = .17$). 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 5.2, all three model fits in the training set were good (neural nets: mean $R^2_{train} = .42$, $SD_{train} = .07$; bootstrap forests: mean $R^2_{train} = .63$, $SD_{train} = .06$; standard least squares: mean $R^2_{train} = .43$, $SD_{train} = .02$). However, only the neural net model performed well for both the training and test sets (mean $R^2_{test} = .54$, $SD_{test} = .14$), while bootstrap forests' and standard least squares regression's fits in the test set were smaller and had higher standard deviations (bootstrap forests: mean $R^2_{test} = .35$, $SD_{test} = .25$; standard least squares: mean $R^2_{test} = .30$, $SD_{test} = .24$).

The *FI* analysis of the optimal neural nets model, shown in Figure 5.3, 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 5.2 also shows that the fits for the training set of all three models were good (neural nets: mean $R^2_{train} = .58$, $SD_{train} = .13$; bootstrap forests: mean $R^2_{train} = .70$, $SD_{train} = .05$; standard least squares: mean $R^2_{train} = .48$, $SD_{train} = .02$). Again, only the neural nets performed well for both the training and test sets (mean $R^2_{test} = .68$, $SD_{test} = .18$), while the model fits in the test sets of bootstrap forests and standard least squares regression were insufficient (bootstrap forests: mean $R^2_{test} = .43$, $SD_{test} = .39$; standard least squares: mean $R^2_{test} = .23$, $SD_{test} = .49$).

For the *FIs* of neural net model shown in Figure 5.3, 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 $R^2_{train} = .37$, $SD_{train} = .02$; mean $R^2_{test} = .29$, $SD_{test} = .27$; fixation probability: mean $R^2_{train} = .33$, $SD_{train} = .01$; mean $R^2_{test} = .14$, $SD_{test} = .75$), lower *logf* (total reading time: mean $R^2_{train} = .36$, $SD_{train} = .02$; mean $R^2_{test} = .25$, $SD_{test} = .26$; fixation probability: mean $R^2_{train} = .27$, $SD_{train} = .02$; mean $R^2_{test} = .06$, $SD_{test} = .66$) and smaller *on* (total reading time: mean $R^2_{train} = .26$, $SD_{train} = .01$; mean $R^2_{test} = .18$, $SD_{test} = .23$; fixation probability: mean $R^2_{train} = .33$, $SD_{train} = .02$; mean $R^2_{test} = .09$, $SD_{test} = .73$) had longer total reading time and a higher fixation probability. Words with lower *odc* (total reading time: mean $R^2_{train} = .17$, $SD_{train} = .02$; mean $R^2_{test} = .07$, $SD_{test} = .26$) attracted longer total reading time. The linear relationship between *sonscore* and the two eye movement parameters was positive: total reading time: mean $R^2_{train} = .19$, $SD_{train} = .01$; mean $R^2_{test} = .11$, $SD_{test} = .20$; fixation probability: mean $R^2_{train} = .15$, $SD_{train} = .001$; mean $R^2_{test} = .02$, $SD_{test} = .41$.

5.5 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's 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 nonlinear 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, Bengio, & Hinton, 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 test 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 test 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, 2001; 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, New, & Kennedy, 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 long and low-frequency words tend to be fixated more often and longer (e.g., Just & Carpenter, 1980; Inhoff & Rayner, 1986; Raney & Rayner, 1995; Pynte, New, & Kennedy, 2008), but also suggest other important predictors, at least for the reading of poetry: words high in *orthographic neighborhood density* attract fewer fixations and shorter total reading time supporting the facilitative effect hypothesis of Andrews (1989, 1992). Additionally, words that were more orthographically dissimilar (i.e., more salient) also attracted longer total reading time. The results concerning the feature *higher frequent neighbors* are inconclusive across the three models which may be because 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).

5.6 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., 2015a; Jacobs et al., 2016). However, for this first study with 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, Rayner, & Pollatsek, 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’s 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 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.

Author Contributions

All authors contributed to the design of the experiment. Xue S. carried out the experiment, analyzed the data, and wrote the first draft of the manuscript; Lüdtke J. modified the manuscript; Jacobs A. M. improved the manuscript. All authors have contributed to and approved the final manuscript.

Chapter 6: What is the Difference? Rereading Shakespeare's Sonnets — an Eye Tracking Study²

Shuwei Xue, Arthur M. Jacobs, and Jana Lüdtke

6.1 Abstract

Texts are often reread in everyday life, but most studies of rereading have been based on expository texts, not on literary ones such as poems, though literary texts may be reread more often than others. To correct this bias, the present study is based on two of Shakespeare's sonnets. Eye movements were recorded, as participants read a sonnet then read it again after a few minutes. After each reading, comprehension, and appreciation were measured with the help of a questionnaire. In general, compared to the first reading, rereading improved the fluency of reading (shorter total reading times, shorter regression times, and lower fixation probability) and the depth of comprehension. Contrary to the other rereading studies using literary texts, no increase in appreciation was apparent. Moreover, results from a predictive modeling analysis showed that readers' eye movements were determined by the same critical psycholinguistic features throughout the two sessions. Apparently, even in the case of poetry, the eye movement control in reading is determined mainly by surface features of the text, unaffected by repetition.

Keywords: rereading, poetry reading, eye movements, QNA, predictive modeling

² This chapter is published as "What is the Difference? Rereading Shakespeare's Sonnets — an Eye Tracking Study". *Frontiers in psychology*, 11, 421.
<https://doi.org/10.3389/fpsyg.2020.00421>

6.2 Introduction

*When to the sessions of sweet silent thought
I summon up remembrance of things past,*

William Shakespeare, Sonnets 30 (ll. 1-2)

What happens if you read a text for the second time? You may read it faster, remember more details, and understand it better. This improvement, widely known as the rereading benefit or rereading effect, has been noted in many studies (see Raney, 2003, for a review). Most of them, however, have been based on the rereading of expository texts (e.g., Hyönä & Niemi, 1990; Levy, Masson, & Zoubek, 1991; Levy, di Persio, & Hollingshead, 1992; Raney & Rayner, 1995; Rawson, Dunlosky, & Thiede, 2000; Raney, Therriault, & Minkoff, 2000; Schnitzer & Kowler, 2006; Kaakinen & Hyönä, 2007; Margolin & Snyder, 2018), only a few of them on the rereading of literary texts (e.g., Dixon et al., 1993; Millis, 1995; Kuijpers & Hakemulder, 2018) and only one of these on the rereading of poetry (Hakemulder, 2004). None of those based on literary texts used direct or indirect methods to record the cognitive processes associated with comprehension and appreciation while they were happening. Researchers have relied on assessments made by readers after, not during, the process of reading. We wished to overcome this limitation by relying not only on assessments made later but also on eye-movements made during the reading of poetry. Here we shall begin by glancing at earlier studies showing the benefit of rereading, go on to present our own approach, put forward hypotheses and finally check them empirically.

6.2.1 *The Effect of Rereading Expository and Literary Texts*

Ever since the rereading paradigm was introduced by Hyönä and Niemi (1990), it has been used in a few studies in various domains (e.g., by Levy, Masson, & Zoubek, 1991; Raney, Therriault, & Minkoff, 2000; Schnitzer & Kowler, 2006; Kaakinen & Hyönä, 2007). Readers have to read a text more than once, and their way of reading is assessed during or after each session (e.g., by eye tracking or self-assessment). In other studies, particular attention was paid to the effect of reading words or phrases repeated within a text (e.g., Kamienkowski et al., 2018), but this is not our concern.

As mentioned above, most studies of rereading have used expository texts as a basis. Expository texts are treated as sources of information stipulating reading processes directed

to information intake, so studies using such texts have tended to focus on whether a reader remembers and understands more after the second compared to the first session. The main findings are: firstly, readers who read an expository text twice recalled significantly more than those who read it only once (Amlund, Kardash, & Kulhavy, 1986; Durgunoğlu, Mir, & Ariño-Martí, 1993); secondly, rereading facilitated readers to build a better comprehension of the topic (Rawson, Dunlosky, & Thiede, 2000; Raney, Therriault, & Minkoff, 2000; Brown, 2002; Schnitzer & Kowler, 2006; Kaakinen & Hyönä, 2007; Margolin & Snyder, 2018). Meanwhile, researchers were also interested in the influence of rereading on reading fluency, e.g., whether the reading time spent on the text or on single words within that text would be saved. The answers to these questions were positive. That is, after a first reading, not only was the overall time spent on reading the expository text less (Millis & King, 2001), but rereading also improved most eye tracking parameters on the word level: total reading time (the sum of all fixation durations on a certain word) was less, regression time (the sum of fixations on a certain word after the first passage) was less, and the rate of skipping was higher (Hyönä & Niemi, 1990; Raney & Rayner, 1995; Raney, Therriault, & Minkoff, 2000; Kaakinen & Hyönä, 2007).

Many studies have confirmed the benefit of rereading, but only a few of them have sought the cause. In general, the rereading benefit may have been due to a change in the roles played by lexical, interlexical, or supralexical features in the course of reading. Levy and his colleagues (1992, 1993) have assumed that the rereading benefit could be observed not only when rereading the same text but also when reading another text with a similar meaning or context. They checked this assumption by replacing some words with synonyms, by changing the syntactic structure of the text and by using a paraphrased text in the rereading session. The results confirmed their hypotheses. However, Raney, Therriault, and Minkoff (2000) found that when for the second reading a paraphrased version of the original text (words from the related texts were replaced by synonyms) was used, only gaze duration (the sum of all fixation durations on a certain word during first passage) and total reading time were less. They assumed, that rereading had a stronger influence on later processing stages compared to early ones. To clarify at least the role of some lexical features, Raney and Rayner (1995) have tried changing the frequency of words in expository texts, but the decrease in fixation durations was the same for low- and high-frequency words across readings. Likewise, Chamberland et al. (2013) found that the benefit of rereading was the same for content and function words and for low- and high-frequency words, except in the case of gaze duration,

when the rereading effect was greater for function than for content words. However, some studies have found that low-frequency words benefit more from multiple readings than high-frequency words (see Kinoshita, 2006, for a review). In other words, results have been inconsistent regarding the effects of rereading on eye tracking parameters in the early stages of the process (e.g., on gaze duration), especially in the case of various psycholinguistic features. The exact roles played by psycholinguistic features on various eye tracking parameters in rereading need further investigation.

Moreover, all the above findings are based on the rereading of expository texts. There have been only a handful of studies on rereading of literary texts, and these have relied only on assessments made after reading. Not surprisingly, these studies also found classical rereading effects, e.g., enhanced comprehension (e.g., Klin, Ralano, & Weingartner, 2007; Kuijpers & Hakemulder, 2018). Especially in the case of literary texts, researchers have also been interested in whether rereading affects a reader's appreciation and aesthetic emotional reactions as a result of 'literary/foregrounding effects. They assumed that 'literary/foregrounding effects' might be related to the level of comprehension (Kuijpers & Hakemulder, 2018), so increased during second reading (Dixon et al., 1993). In line with this hypotheses, the scant studies using literary texts found that rereading indeed influenced readers' appreciation, insofar as readers tended to rate texts as more likable after the rereading session (e.g., Dixon et al., 1993; Millis, 1995; Kuijpers & Hakemulder, 2018). The only study on the rereading of poetry has confirmed this hypothesis (Hakemulder, 2004). Nevertheless, none of the studies based on literary texts have checked cognitive and emotional processes associated with comprehension and appreciation while they were happening, by for instance recording the movements of a reader's eyes on the single word level. Whether a literary text is read more fluently the second time round is still an open question.

Hence the main aim of the present study is to examine the effects of rereading poetic texts by using not only assessments made by readers *after* the sessions but also records of eye-movements made *during* the sessions, to find out whether rereading affects a reader's understanding and appreciation and increases the fluency of reading. A further aim is to check whether surface psycholinguistic features, like *word frequency*, may play a role in changing eye tracking parameters across reading sessions.

6.2.2 Eye Movement Research on Poetry Reading

As we all know, it is not easy to do research using natural texts, as they are mostly very complex (Jacobs et al., 2017; Xue et al., 2017; 2019). Especially if we use literary texts such as poems not specially designed for research (Bailey & Zacks, 2011; Willems & Jacobs, 2016), simple or complex text features, seldom occur without interacting with many other features on various levels. Although there have been studies on reading literary texts or poems (e.g., Carrol & Conklin, 2014; Dixon & Bortolussi, 2015; Jacobs et al., 2016; van den Hoven et al., 2016; Lauwereyns & d'Ydewalle, 1996; Müller et al., 2017; Sun, Morita, & Stark, 1985), the big majority of eye tracking studies on reading constrained to experimental textoids and tested only a few selected features while ignoring many others (Rayner et al., 2001; Reichle, Rayner, & Pollatsek, 2003; Rayner & Pollatsek, 2006; Engbert et al., 2005; Rayner, 2009).

Within the framework of neurocognitive poetics (Jacobs, 2011, 2015a, b; Nicklas & Jacobs, 2017; Willems & Jacobs, 2016), two steps have been suggested to cope with the innumerable features of texts and/or readers and their many (non-linear) interactions. Firstly, a way should be found to break the complex literary works up into simpler, measurable features, for instance by Quantitative Narrative Analysis (QNA; e.g., Jacobs, 2015a, 2017, 2018a, 2019; Jacobs, Hofmann, & Kinder, 2016; Jacobs et al., 2017; Jacobs & Kinder, 2017, 2018; Xue et al., 2019). Secondly, proper statistical and machine learning modeling tools should be chosen to cope with intercorrelated, non-linear relationships between the many features which may affect the (re)reading of poetry (e.g., Jakobson & Lévi-Strauss, 1962; Schrott & Jacobs, 2011; Jacobs, 2015a, b, c, 2019; Jacobs, Hofmann, & Kinder, 2016; Jacobs et al., 2016).

Recently, a QNA-based predictive approach was successfully applied to account for eye tracking parameters in the reading of three Shakespeare's sonnets (sonnet 27, 60, and 66) using multiple psycholinguistic features (Xue et al., 2019). In the study of Xue et al. (2019), seven surface psycholinguistic features, a combination of well-studied (*word length*, *word frequency*, and *higher frequent neighbors*) and less-studied and novel features (*orthographic neighborhood density*, *orthographic dissimilarity*, *consonant vowel quotient*, and *sonority score*), were computed based on the Neurocognitive Poetics Model (NCPM, Jacobs, 2011, 2015a, b; Nicklas & Jacobs, 2017; Willems & Jacobs, 2016) and recent proposals about QNA (e.g., Jacobs et al., 2017; Jacobs, 2017, 2018a, b). Besides, two non-linear interactive approaches, i.e., neural nets and bootstrap forests, were compared with a general linear approach (standard least squares regression), to look for the best way to predict three eye

tracking parameters (first fixation duration, total reading time, and fixation probability) using the seven above mentioned features. For the prediction of first fixation duration, none of the three approaches yielded appropriate model fits, as the first fixation duration may have been due more to fast and automatic reading behavior rather than to lexical parameters (Hyönä & Hujanen, 1997; Clifton, Staub, & Rayner, 2007). For the other two parameters total reading time and fixation probability, neural nets outperformed the general linear approach and also the bootstrap forests. This might be due to the fact, that within this context neural nets could best deal with the complex interactions and nonlinearities in the data (Coit, Jackson, & Smith, 1998; Francis, 2001; Breiman, 2001; Yarkoni & Westfall, 2017). Most importantly, the feature importance analysis of the optimal neural nets approach detected that the two well-known basic features, *word length* and *word frequency*, were most important in accounting for the variance in total reading time and fixation probability. Moreover, also two of the novel features were important predictors. One of the two phonological features, the *sonority score*, was important for predicting both total reading time and fixation probability. *Orthographic neighborhood density* and *orthographic dissimilarity* proved to be important for predicting total reading time, whereas *orthographic neighborhood density* proved to be important for predicting fixation probability.

For the present study, which is a first attempt to evaluate the effects of surface psycholinguistic features in poetry rereading investigation using eye tracking, we also want to compare the predictive performance of neural nets as an example of a non-linear interactive approach with a general linear approach, including the same seven predictors used in Xue et al. (2019), but with a new larger sample of readers. Thus, in the context of the 'replication crisis' debate (Maxwell, Lau, & Howard, 2015; Earp & Trafimow, 2015; Shrout & Rodgers, 2018), the present study also served as a replication (Xue et al., 2019), i.e., whether a neural nets approach could build satisfactory models in a rereading study and whether the same 'important features' in predicting relevant eye tracking parameters would be detected again.

To summarize, the current study examined the general validity of findings of rereading by using two of Shakespeare's sonnets. We asked: 1) whether rereading improves understanding and appreciation; 2) whether rereading increases reading fluency; 3) whether the roles of surface features change across reading sessions. We used the terms *first session* and *last session* to denote the two reading sessions, each of which consisted of reading a sonnet then filling a questionnaire in. The terms have been redefined because poems, unlike expository prose, are seldom read straight through from beginning to end (Müller et al., 2017;

Xue et al., 2017), so a lot of rereading took place within each session. For the sake of improvement in appreciation (Kuijpers & Hakemulder, 2018), we also updated the rereading paradigm by inserting a paraphrasing session between the two sessions.

6.2.3 Hypotheses

Former studies had shown that rereading improved readers' comprehension and increased their appreciation of literary texts (Klin, Ralano, & Weingartner, 2007; Dixon et al., 1993; Millis, 1995; Kuijpers & Hakemulder, 2018), so we expected to get similar results with poetry. In other words, we expected that readers would identify the topic better (showing more understanding) and appreciate the poem more after the last session.

To determine the effect of rereading on fluency, we concentrated on changes in eye tracking parameters on the word level. Mostly, in the case of expository texts, fluency increased after a first reading session (e.g., Levy et al., 1991, 1993), so we expected the same to be true in the case of poetry. However, we also thought that rereading may mostly affect eye tracking parameters related to later stages of processing (e.g., Raney & Rayner, 1995), so regression time and total reading time would be less for the last session. We also expected that the skipping rate in the last session would be higher, lessening the fixation probability. We had no clear expectations about parameters related to early processing, such as first fixation duration and gaze duration, since rereading involves an interplay of several psycholinguistic features, whose effects had not fully been clarified by earlier investigations (Raney & Rayner, 1995; Chamberland et al., 2013; Kinoshita, 2006).

Using poetic materials for reading and rereading, this study wanted to not only replicate effects long familiar from studies using expository texts but also replicate findings from Xue et al. (2019). They had successfully applied QNA-based predictive modeling approaches to the reading of poetic texts, to cope with the intercorrelated, non-linear relationships between the many text features. Since Xue et al. (2019) indicated that neural nets outperformed bootstrap forests, here we only included one non-linear interactive approach (neural nets) and one general linear approach (standard least squares regression). We expected that neural nets would provide the best fits to the data of the cross-validation test sets.

Moreover, we were also interested in the causes of the rereading effect. For instance, which surface psycholinguistic features may affect reading fluency across sessions? Or may

different features affect it in different sessions? There had been no studies of most of them, so in this sense our study was exploratory.

6.3 Method

6.3.1 Participants

English native speakers were recruited through an announcement released at the Freie Universität Berlin. Altogether 25 people took part (eleven females; $M_{\text{age}} = 23.9$ years, $SD_{\text{age}} = 4.3$, age range: 19–33 years). They were neither trained literature scholars of poetry nor aware of the purpose of the experiment. All speakers had a normal or corrected-to-normal vision and gave their informed, written consent before taking part. They were given eight euros as compensation. This study followed the guidelines of the ethics committee of the Department of Education and Psychology at the Freie Universität Berlin. Some eye movement data were removed, as the eye tracker had failed to record them in full. The data finally used for analyzing the eye movements and predictive modeling came from 22 participants for sonnet 27 (eleven females; $M_{\text{age}} = 23.45$ years, $SD_{\text{age}} = 4.1$, age range: 19–32 years) and 23 participants for sonnet 66 (nine females; $M_{\text{age}} = 24.22$ years, $SD_{\text{age}} = 4.36$, age range: 19–33 years).

6.3.2 Apparatus

Eye movements were collected by a remote EYELINK eye tracker (SR Research Ltd., Mississauga, Ontario, Canada). The sampling frequency was 1000 Hz, and only the right eye was tracked. Readers' heads were kept still by a chin-and-head rest. 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 refreshment rate of 60 Hz and a resolution of $1,024 \times 768$ pixels, 50 cm away from the reader. Each tracking session began with a standard 9-point calibration and validation procedure, to ensure a spatial resolution error of less than 0.5° of the angle of vision.

6.3.3 Materials

For this rereading experiment, only two of the three Shakespeare's sonnets used by Xue et al. (2019) were presented, to let readers concentrate without getting tired. The two sonnets were: 27 ('Weary with toil...') and 66 ('Tired with all these...'). Both sonnets

covered different topics, "love as a tension between body and soul" (sonnet 27) and "social evils during the period Shakespeare lived" (sonnet 66). To increase statistical power for all levels of analysis we collapsed the data across the two sonnets.

6.3.4 Procedure

The reading was done in a quiet and dimly lit room and consisted of two tasks: the general mood state task and the main task. Readers were told about the whole procedure at the start.

The general mood state task was used at the beginning and at the end of the experiment, to assess any changes in the reader's moods. They were asked to fill in an English version of the German multidimensional mood questionnaire (MDBF; Steyer et al., 1997), to let three bipolar dimensions of subjective feeling (depressed vs. elevated, calmness vs. restlessness, sleepiness vs. wakefulness) on a 7-point rating scale be assessed. The results showed that they were in a neutral mood of calmness and wakefulness throughout. According to the results of paired-samples *t*-tests, there was no significant change of mood before and after the experiment (all $t(24)s < 2$, $ps > .1$), as if reading sonnets caused no lingering changes in the global dimensions assessed by MDBF.

The main task was made up of five parts: a. a first reading session in front of the eye tracker; b. a paper-pencil task for the first session; c. an oral paraphrasing line by line; d. a last reading session in front of the eye tracker; e. a paper-pencil task for the last session. For the first session, participants were free to read the sonnet at their own speed. Rereading in the course of one session was allowed. Before each sonnet appeared onscreen, readers were presented with a black dot fixation marker (0.6° of the angle of vision) to the left of the first word in line 1, the distance between the dot and first word being 4.6° . When they fixated on the marker, the sonnets appeared automatically. After the first session, readers went to another desk to fill in our self-developed paper-pencil task (see Papp-Zipernovszky et al., in preparation). They got no feedback on their answers. Following this step, they orally paraphrased the sonnet, line by line, according to their understanding, and again, no feedback or fixed answer was given by the experimenter. The paraphrasing process was recorded by a digital voice recorder. Readers were then asked to reread the sonnet at their own speed before the eye tracker again. Before the last reading session, recalibration was needed. In the end, readers worked on the paper-pencil task for the second time. After answering the questionnaire for the first sonnet, they went on to read the second sonnet in front of the eye

tracker. The two sonnets were presented left-aligned in the center of the monitor (distance: 8.0° from the left margin of the screen) by using a font (Arial) with variable width and a letter-size of 22-points (approximately 4.5×6.5 mm, $0.5^\circ \times 0.7^\circ$ of the angle of vision). One reader would be shown sonnet 27 first and the next be shown sonnet 66 first and so on, to cancel out any effect due to the sequence. Likewise, a questionnaire was presented before the last session, so a sample questionnaire was also presented before the first.

Altogether, the experiment took about 50 minutes (see Figure 6.1 for an illustration of the procedure).

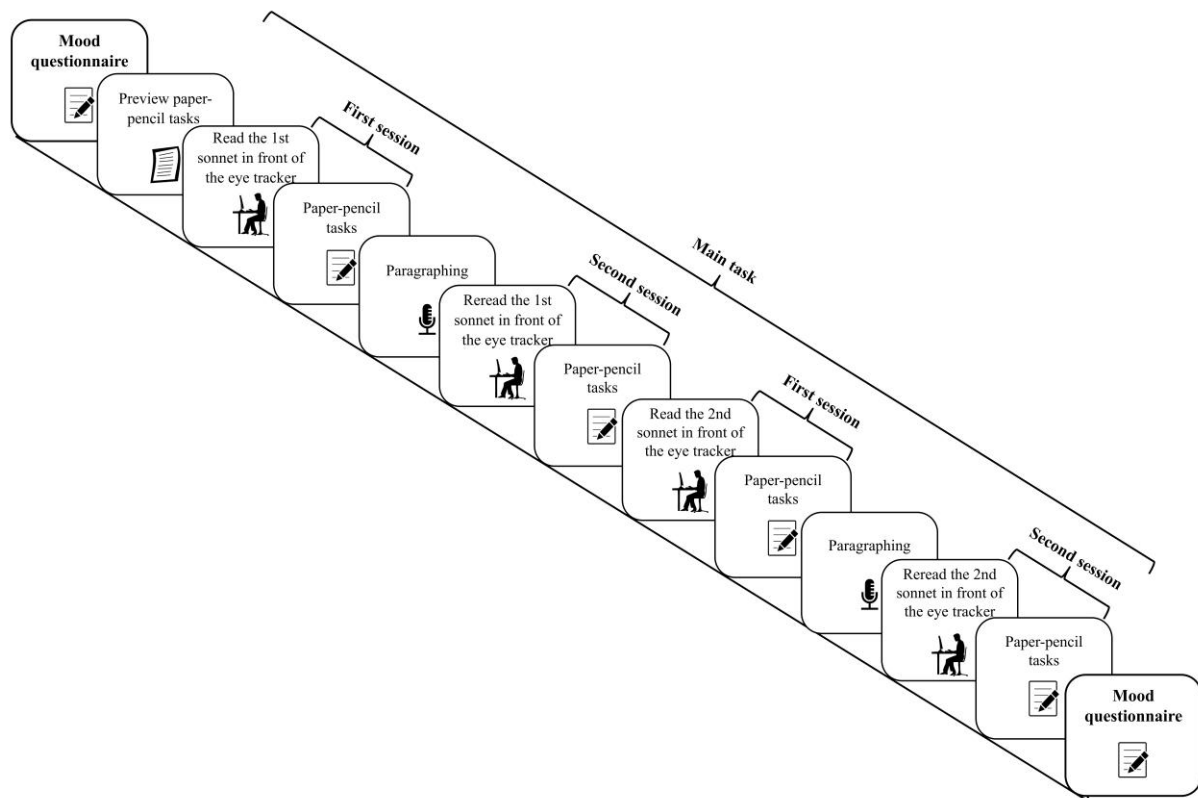


Figure 6.1 The Procedure of the Experiment

6.3.5 Data Analysis

Paper-pencil Task.

Unlike in the paper-pencil task used by Papp-Zipernovszky et al. (in preparation) and Xue et al. (2019), the question about rhyme pairs was included in only the questionnaire used for the last session, so as not to divert attention from comprehension, so in this respect there could be no comparison between the first and the last session. Otherwise, all parts of the questionnaire were the same for the first and last session.

In the present study, we focused on three questions, one related to the general willingness to do any rereading, another one related to comprehension, and a third one related to appreciation. Since a lot of “rereading” was involved in reading the questionnaire presented after each session, the question about willingness (“I would like to read this poem again”) was used as a control question. After the last session, participants should have reported less willingness to do any rereading, in being weary and less motivated. The question, “I like this poem”, was used to evaluate participants’ appreciation of it (Lüdtke, Meyer-Sickendieck, & Jacobs, 2014; Kraxenberger & Menninghaus, 2017). For both questions, readers indicated their agreement with the statements on a 5-point rating scale ranging from 1 = totally disagree to 5 = totally agree. The topic identification question was meant to find out whether readers successfully grasped the main topic of each poem (“Which is the main topic of this poem?”). Six choices were offered, but only one was right. If readers agreed with none of the choices, they could put forward another, which was later evaluated by two experts from the humanities. In the two sessions, 20% of the answers were formulated by the participants themselves (first session: 10 answers; last session: 10 answers). Answers, which were not clear or did not exceed the explanation of surface meaning, were evaluated as wrong. For instance, for sonnet 27, answers like “Never resting” or “A journey both physically and mentally” were coded as wrong. None of the self-formulated answers in the first session was right, but 40% of them (4 answers) were right in the last session.

JMP 14 Pro (https://www.jmp.com/en_us/software/predictive-analytics-software.html) was used for the statistical analyses. For the two questions about appreciation and a general willingness to do any rereading, we used paired-samples *t* tests, to check the differences between the first session and the last. Since we evaluated and recoded the answers for the topic identification question as “yes” or “no” (categorical variable), we then used a non-parametric test, i.e., Bowker’s test, to check the difference between sessions.

Eye Tracking Parameters.

Pre-processing of the raw data was done by EyeLink Data Viewer (<https://www.sr-research.com/data-viewer/>). As mentioned earlier, data from three readers of sonnet 27 and two readers of sonnet 66 were removed, because the eye tracker had failed to record their eye movements. From the data, we then determined first fixation duration (the duration of the first fixation on a certain word), gaze duration (the sum of all fixations on a certain word during first passage), regression time (the sum of fixations on a certain word after first

passage) and total reading time (the sum of all fixation durations on a certain word) for each word, participant and sonnet.

For all analyses predicting eye tracking parameters, we focused on the effect of text-based features on rereading. We also decided to use the same pre-processed data for all analyses. To reliably test the effect of the surface features used in Xue et al. (2019) in predicting eye tracking parameters in first and last reading by neural nets, the eye tracking data have to be aggregated at the word level. We therefore cumulated the data over all participants to obtain the mean values for each word within each sonnet and each session. To take the amount of skipping into account, the fixation probability was calculated. Skipped words were thus treated as missing values (skipping rate: $M_{\text{first-session}} = 13\%$, $SD_{\text{first-session}} = .34$; $M_{\text{last-session}} = 20\%$, $SD_{\text{last-session}} = .40$). For instance, in the last session, words fixated by all participants, like 'expired' (sonnet 27) or 'jollity' (sonnet 66) had a probability of 100%, whereas words fixated by only one or two participants like 'To' (sonnet 27) or 'I' (sonnet 66) had fixation probabilities below 20%. Altogether, in the first session over 38% of the words had a fixation probability of 100% and in the last session the amount decreased to 25%, which led to a highly asymmetric distribution. However, unlike Xue et al. (2019), we did not aggregate the eye tracking data for words appearing twice or more often. Instead, here we included positional information (line number: *lineNo.*; word number in each line: *wordNo.*) as a feature in the predictive modeling analysis. For each reading session, the total sample size entering in the models was $N = 202$ words. The correlations between the five aggregated eye tracking parameters are shown in Table 6.1.

To test for the rereading effects at the word-level, linear mixed models (LMM) with one fixed effect (session) and one random effect (word nested within sonnet) were applied to the five eye tracking parameters using JMP 14 Pro.

Table 6.1 Correlations between the Five Eye Tracking Parameters

Session	Variables	1	2	3	4	5
First	1. First fixation duration	–				
	2. Gaze duration	.65	–			
	3. Regression time	.14	.46	–		
	4. Total reading time	.34	.72	.95	–	
	5. Fixation probability	.16	.34	.62	.61	–
Last	1. First fixation duration	–				
	2. Gaze duration	.77	–			
	3. Regression time	.20	.45	–		
	4. Total reading time	.53	.82	.88	–	
	5. Fixation probability	.23	.44	.55	.59	–

Predictors for Predictive Modeling.

Positional information. As mentioned earlier, several words are repeated in the sonnets (e.g., *mind*), so we added the positional information (*lineNo.* and *wordNo.*) of the words in each sonnet.

Psycholinguistic features. Seven psycholinguistic features were calculated for all words (word-token, 202 words) in the two sonnets: *word length (wl)* is the number of letters per word; *word frequency (logf)* is the log transformed number of times that a word appears in the Gutenberg Literary English Corpus as a reference (GLEC; Jacobs, 2018b; Xue et al., 2019); *orthographic neighborhood density (on)* is the number of words of the same length as a certain word and differing by only one letter in GLEC; *higher frequent neighbors (hfn)* is the number of orthographic neighbors with a higher word frequency than the word in GLEC; *orthographic dissimilarity (odc)* is the word's mean Levenshtein distance from all other words in the corpus (GLEC), a metric that generalizes orthographic similarity 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). For example, in our two sonnets, ART got the *sonscore* of $10 \times 1 [a] + 7 \times 1 [r] + 1 \times 1 [t] = 18 / \text{SQRT}(3) = 10.39$. As shown in Table 6.2, some of these

psycholinguistic features were highly correlated, hence the need to apply machine-learning tools in a predictive approach (e.g., Coit, Jackson, & Smith, 1998; Francis, 2001; Tagliamonte & Baayen, 2012; Yarkoni & Westfall, 2017).

Table 6.2 Correlations between the Seven Psycholinguistic Features

Variables	1	2	3	4	5	6	7
1. Word length (<i>wl</i>)	–						
2. Log frequency (<i>logf</i>)	-.81	–					
3. Orthographic neighbors (<i>on</i>)	-.85	.72	–				
4. Higher frequency neighbors (<i>hfn</i>)	-.23	-.06	.28	–			
5. Orthographic dissimilarity based on corpus (<i>odc</i>)	.62	-.44	-.28	-.12	–		
6. Consonant vowel quotient (<i>cvq</i>)	.30	-.10	-.36	-.14	.04	–	
7. Sonority score (<i>sonscore</i>)	.74	-.64	-.61	-.23	.58	.07	–

Predictive Modeling.

We also utilized the JMP 14 Pro to run all predictive modeling analyses³. As described above, nine predictors (*lineNo.*, *wordNo.*, *wl*, *logf*, *on*, *hfn*, *odc*, *cvq*, and *sonscore*) and five eye tracking parameters (first fixation duration, gaze duration, regression time, total reading time, and fixation probability) were included in these analyses. The values of all eye movement parameters and psycholinguistic features were standardized before being analyzed in predictive modeling.

Cross-validation was used as a solution to the problem of overfitting. Among the methods of cross-validation, K-fold appears to work better than hold-out in the case of a small sample size, because it uses data more efficiently (Refaeilzadeh, Tang, & Liu, 2009). It divides the original data into K subsets. In turn, each of the K sets is used to test the model fit on the rest of the data, fitting a total of K models. The model giving the best test statistic is chosen as the final model. The 10-fold cross-validation is usually recommended as the best method since it provides the least biased estimation of accuracy (Kohavi, 1995). Therefore, in the present study, instead of the 10% hold-out cross-validation method (i.e., taking 90% of

³For the neural nets 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 standard least squares regression, we only specified the nine fixed effects (*lineNo.*, *wordNo.*, *wl*, *logf*, *on*, *hfn*, *odc*, *cvq*, and *sonscore*) and predicted each eye tracking parameter using the same nine predictors.

the data as a training set and the remaining 10% as a test set) used in Xue et al. (2019), we used 10-fold cross-validation.

Given the intrinsic probabilistic nature of neural nets, predictive modeling results vary across repeated runs. These differences depend also on the splitting into training and test set during cross-validation (total sample size = 202 words, i.e., about 20 cases in each fold during cross-validation). To cover the potential disadvantages of splitting small samples, the k-fold cross-validation procedure was repeated 100 times and the model fit scores were averaged (e.g., Were et al., 2015). Note that for the standard least square regression, JMP 14 Pro only provides the 100 model fit scores for the test sets, which, of course, is the relevant piece of information.

Following the procedure of Xue et al. (2019) for comparing neural nets and linear regression, our criterion for a satisfactory model fit score was a mean $R^2 > .30$ and a low SD). When the non-linear interactive approach, proved to be satisfactory, we determined feature importance (FI), an index of effect strength used in machine learning (https://scikit-learn.org/stable/modules/feature_selection.html). In the current study, FIs were computed as the total effect of each predictor as 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 together with other features (for details, see also Saltelli, 2002). This measure is interpreted as an ordinal value on a scale of 0 to 1, FI values $> .1$ being considered as 'important' (cf. Strobl, Malley, & Tutz, 2009). If the general linear approach proved to be satisfactory, the parameter estimates were reported instead of FIs .

6.4 Results

6.4.1 Paper-pencil Task

The results of the rereading effects on rating data are shown in Figure 6.2: Firstly, there was a significant effect on readers' willingness to do any rereading ($t(49) = 3.32, p = .002$). After the last session, readers were less willing to reread the sonnet than after the first session (first session: $M = 3.78, SD = 1.04$; last session: $M = 3.18, SD = 1.04$). Secondly, the rereading effect on topic identification was also significant ($\chi^2 = 8, df = 1, p = .005$). Readers were more able to choose the right topic after the last session than after first session (first session: $N_{\text{right}} = 30, N_{\text{wrong}} = 20$; last session: $N_{\text{right}} = 42, N_{\text{wrong}} = 8$).

Readers tended to appreciate a sonnet in the last session more than in the first (first session: $M = 3.32$, $SD = .94$; last session: $M = 3.52$, $SD = 1.02$), but the difference was not statistically significant ($t(49) = -1.81$, $p = .077$). We also checked for each sonnet separately by applying a paired-samples t test. For sonnet 27, there was no significant difference in appreciation, whether it was read in the first or last session ($t(24) = -.30$, $p = .77$; first session: $M = 3.88$, $SD = .67$; last session: $M = 3.92$, $SD = .81$), but there was a significant difference for sonnet 66 ($t(49) = -2.09$, $p = .047$), which was appreciated more if read in the last session (first session: $M = 2.76$, $SD = .83$; last session: $M = 3.12$, $SD = 1.05$).

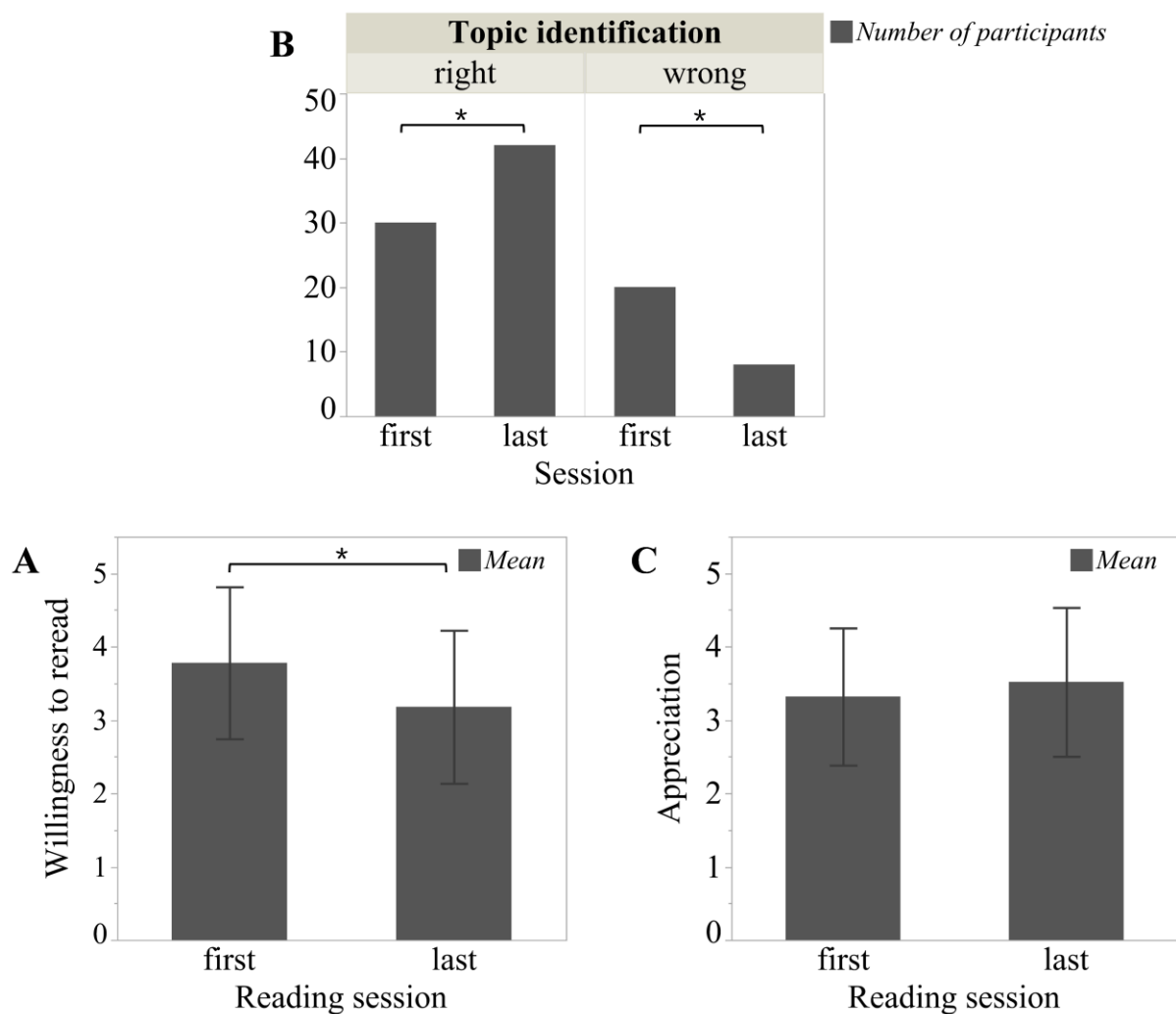


Figure 6.2 Rereading Effect on Rating Data

6.4.2 Eye Tracking Parameters

As illustrated in Figure 6.3, linear mixed models (LMM) with one fixed effect (session) and one random effect (word nested within sonnet) showed significant rereading effects on regression time ($t(1) = 22.34$; $p < 0.0001$), total reading time ($t(1) = 20.28$; $p < 0.0001$), and fixation probability ($t(1) = 6.54$; $p < 0.0001$). In the last session as compared to the first, readers tended to spend less time on regressions (first session: $M = 414.90$ ms, $SD = 243.78$; last session: $M = 149.85$ ms, $SD = 120.12$) and to shorten their total reading time (first session: $M = 739.80$ ms, $SD = 309.48$; last session: $M = 474.45$ ms, $SD = 187.05$). Moreover, the probability of fixating a word was likewise smaller in the last session (first session: $M = 86.81\%$, $SD = 17.49$; last session: $M = 80.35\%$, $SD = 21.85$).

However, for first fixation duration (first session: $M = 256.84$ ms, $SD = 40.43$; last session: $M = 259.25$ ms, $SD = 43.04$) and gaze duration (first session: $M = 324.91$ ms, $SD = 108.51$; last session: $M = 324.60$ ms, $SD = 99.19$), we found no significant differences between the two sessions (first fixation duration: $t(1) = -.83$; $p = 0.41$; gaze duration: $t(1) = .06$; $p = .95$).

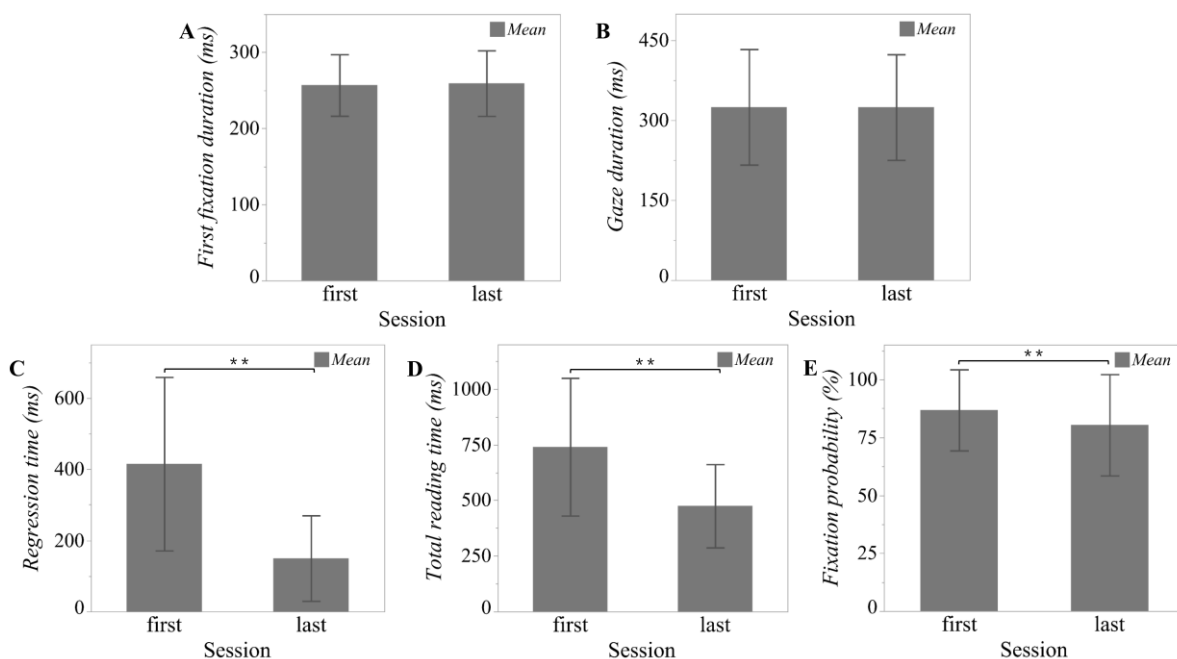


Figure 6.3 Rereading Effect on Eye Tracking Parameters

6.4.3 Predictive Modeling

Figure 6.4 shows the overall R^2 (100 iterations) for predicting the five eye tracking parameters using the two modeling approaches. As mentioned above, for the standard least square regression the R^2 for the whole data set and the mean R^2 for the test sets were computed. As illustrated in Figure 6.4, generally neural nets produced acceptable models for all five eye tracking parameters (mean $R^2 > .30$), and they also produced much higher model fits than standard least squares regression. Therefore, the nine FIs for the neural nets were computed (see Figure 6.5). Below we illustrate our results for the five eye tracking parameters, respectively.

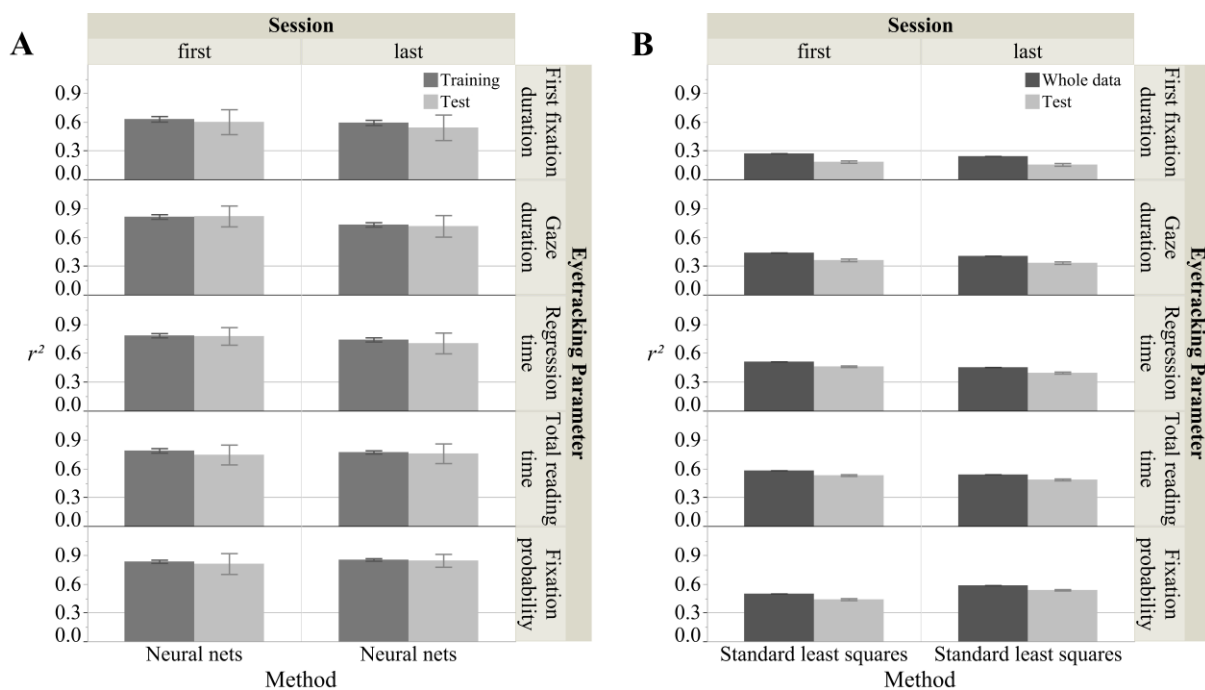


Figure 6.4 Fit Scores for Different Models and Measures

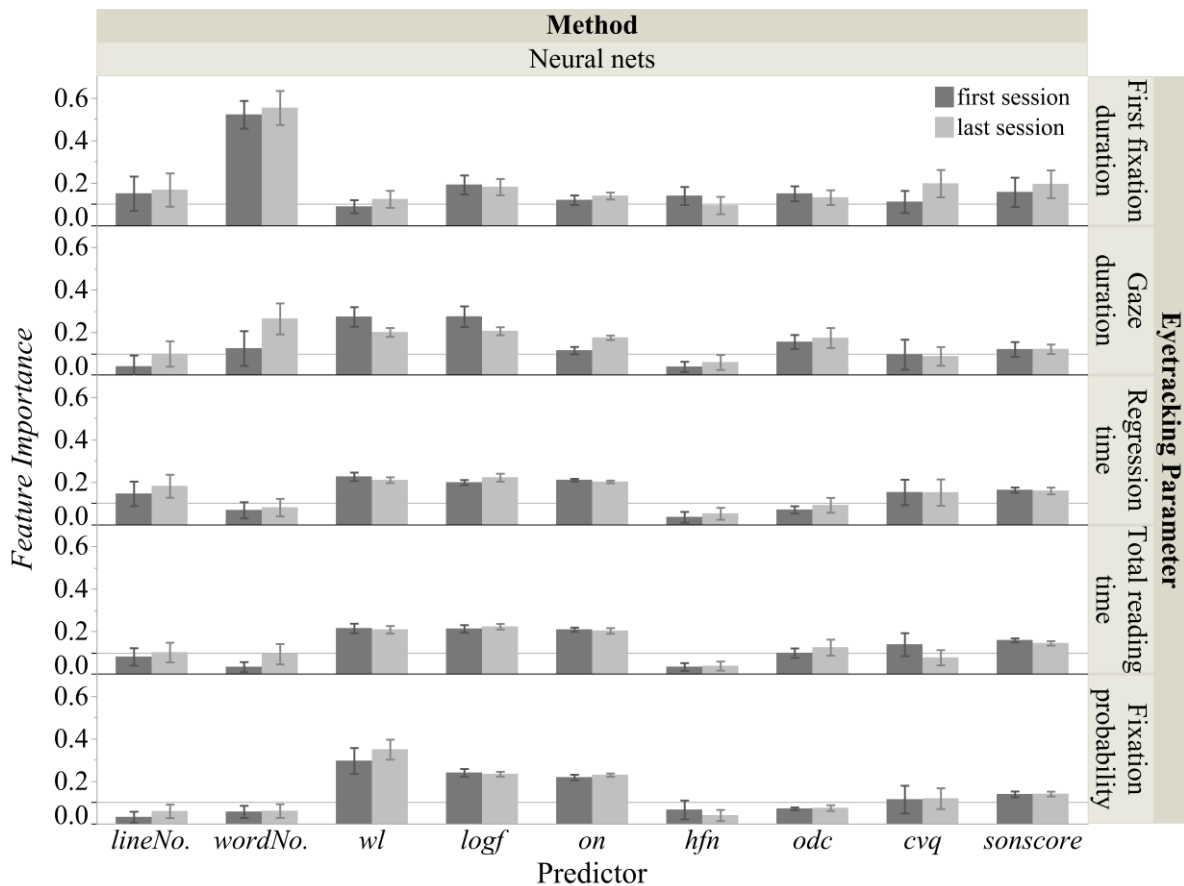


Figure 6.5 Feature Importance for the Five Eye Tracking Parameters

First fixation duration.

As shown in Figure 6.4, in the first session, neural nets produced good fits for both the training and test sets (mean $R^2_{train} = .63$, $SD R^2_{train} = .03$; mean $R^2_{test} = .60$, $SD R^2_{test} = .13$) in contrast to standard least squares ($R^2_{whole} = .27$; mean $R^2_{test} = .18$, $SD R^2_{test} = .01$). The same was true for the last session. Only neural nets produced good fits (neural nets: mean $R^2_{train} = .59$, $SD R^2_{train} = .03$; mean $R^2_{test} = .54$, $SD R^2_{test} = .13$; standard least squares: mean $R^2_{whole} = .24$; mean $R^2_{test} = .16$, $SD R^2_{test} = .01$).

The *FI* analysis of the optimal neural nets approach in Figure 6.5 suggested that in the first session nearly all the predictors were important for predicting first fixation duration (*wordNo.* [.52], *logf* [.19], *sonscore* [.16], *lineNo.* [.15], *odc* [.15], *hfn* [.14], *on* [.12], *cvq* [.11]), except for *wl* (.09). Similarly, in the last session also all predictors were important (*wordNo.* [.55], *cvq* [.20], *sonscore* [.20], *logf* [.18], *lineNo.* [.17], *on* [.14], *odc* [.13], *wl* [.12], *hfn* [.10]).

Gaze duration.

Figure 6.4 shows that in the first session, neural nets and standard least squares both produced acceptable fits, but those of neural nets were clearly higher (mean $R^2_{train} = .82$, $SD R^2_{train} = .02$; mean $R^2_{test} = .82$, $SD R^2_{test} = .11$) than standard least squares ($R^2_{whole} = .44$; mean $R^2_{test} = .36$, $SD R^2_{test} = .01$). The same was true for the last session. Although both approaches yielded acceptable models, neural nets again produced clearly better fits (mean $R^2_{train} = .73$, $SD R^2_{train} = .02$; mean $R^2_{test} = .72$, $SD R^2_{test} = .11$) than standard least squares ($R^2_{whole} = .41$; mean $R^2_{test} = .33$, $SD R^2_{test} = .01$).

The *FI* analysis of the optimal neural nets approach shown in Figure 6.5 suggested that in the first session, seven predictors were important for predicting gaze duration (*logf* [.27], *wl* [.27], *odc* [.16], *wordNo.* [.12], *sonscore* [.12], *on* [.11], *cvq* [.10]), while *lineNo.* (.04) and *hfn* (.04) were less important. For the last session, there were also seven important predictors (*wordNo.* [.26], *logf* [.21], *wl* [.20], *on* [.18], *odc* [.17], *sonscore* [.12], *lineNo.* [.10]), while this time the less important ones were *cvq* (.09) and *hfn* (.06).

Regression time.

As illustrated in Figure 6.4, similar to gaze duration, in the first session, neural nets and standard least squares again were both acceptable; but neural nets produced higher model fits (mean $R^2_{train} = .78$, $SD R^2_{train} = .02$; mean $R^2_{test} = .78$, $SD R^2_{test} = .09$) than standard least squares ($R^2_{whole} = .51$; mean $R^2_{test} = .46$, $SD R^2_{test} = .01$). The same was true for the last session with neural nets (mean $R^2_{train} = .74$, $SD R^2_{train} = .02$; mean $R^2_{test} = .70$, $SD R^2_{test} = .11$) being better than standard least squares ($R^2_{whole} = .45$; mean $R^2_{test} = .40$, $SD R^2_{test} = .01$).

Figure 6.5 shows the *FI* analysis of the optimal neural nets approach suggesting that in the first session, six predictors were important for regression time (*wl* [.23], *on* [.21], *logf* [.20], *sonscore* [.16], *cvq* [.15], *lineNo.* [.15]), while *odc* (.07), *wordNo.* (.07), and *hfn* (.04) were less important. For the last session, the important predictors were the same (*logf* [.22], *wl* [.21], *on* [.20], *lineNo.* [.18], *sonscore* [.16], *cvq* [.15]), as were the less important ones: *odc* (.09), *wordNo.* (.08), and *hfn* (.05).

Total reading time.

Likewise, Figure 6.4 shows results for neural nets (mean $R^2_{train} = .79$, $SD R^2_{train} = .02$; mean $R^2_{test} = .75$, $SD R^2_{test} = .10$) and standard least squares ($R^2_{whole} = .58$; mean $R^2_{test} = .53$, $SD R^2_{test} = .01$) during the first session and for the last session: neural nets (mean $R^2_{train} = .77$,

$SD R^2_{train} = .02$; mean $R^2_{test} = .76$, $SD R^2_{test} = .10$) and standard least squares ($R^2_{whole} = .54$; mean $R^2_{test} = .49$, $SD R^2_{test} = .01$).

The *FI* analysis of the optimal neural nets approach shown in Figure 6.5 suggested that in the first session, six predictors were important for total reading time (*wl* [.22], *logf* [.21], *on* [.21], *sonscore* [.16], *cvq* [.14], *odc* [.10]), while *lineNo.* (.08), *hfn* (.03), and *wordNo.* (.03) were less important. For the last session, there were also six important predictors (*logf* [.22], *wl* [.21], *on* [.20], *sonscore* [.14], *odc* [.12], *lineNo.* [.10]), and three less important ones: *wordNo.* (.09), *cvq* (.08), and *hfn* (.04).

Fixation probability.

Finally, Figure 6.4 also gives results for the first session for both neural nets (mean $R^2_{train} = .84$, $SD R^2_{train} = .02$; mean $R^2_{test} = .81$, $SD R^2_{test} = .02$) and standard least squares ($R^2_{whole} = .50$; mean $R^2_{test} = .44$, $SD R^2_{test} = .01$). For the last session, again, neural nets produced better model fits (mean $R^2_{train} = .86$, $SD R^2_{train} = .01$; mean $R^2_{test} = .85$, $SD R^2_{test} = .07$) than standard least squares ($R^2_{whole} = .59$; mean $R^2_{test} = .54$, $SD R^2_{test} = .01$).

The *FI* analysis of the optimal neural nets approach in Figure 6.5 suggested that in the first session, five predictors were important for fixation probability (*wl* [.30], *logf* [.24], *on* [.22], *sonscore* [.14], *cvq* [.11]), while *odc* (.07), *hfn* (.07), *wordNo.* (.06), and *lineNo.* (.03) were less important. For the last session, the important predictors were the same (*wl* [.35], *logf* [.23], *on* [.23], *sonscore* [.14], *cvq* [.12]), as were the less important ones: *odc* (.07), *wordNo.* (.06), *lineNo.* (.06), and *hfn* (.04).

6.5 Discussion

Every day we all read many kinds of texts such as news reports, blogs, brochures, biographies, reviews, instructions and regulations, novels, or poetry for the sake of being informed or entertained. Usually, we read a text or parts of a text more than once to grasp all the main points or to deepen our enjoyment, and this is especially true in the case of literature. Once a text is familiar from a first reading, it may be read faster. All of these effects are familiar and are known as the classical reading benefit found in many studies based on expository texts, but few looked at literary ones such as poetry. Arguably no writer of classical literature is more eminent than Shakespeare, so we chose two of his sonnets as our materials. We compared the rating data and the eye tracking data in the first session with those in the later and analyzed the difference, then we also analyzed the roles played by seven

surface psycholinguistic features in predicting five eye tracking measures in both sessions with the help of predictive modeling.

6.5.1 *The rereading benefit or rereading effect*

In line with previous studies (e.g., Hakemulder, 2004; Kuijpers & Hakemulder, 2018), our questionnaire data indicated that readers identified the main topic more reliably after the last session. This shows that rereading Shakespeare's sonnets does indeed enhance readers' understanding. As assumed by Hyönä and Niemi (1990), a first reading conjures up in readers a mental representation, which rereading may activate for the sake of easier understanding, even in the case of poetry. Moreover, as shown by answers to the question about a willingness to read the poem again, readers were less willing to do so after the last session. Each sonnet involved a lot of rereading, so readers may have felt more fatigued after the last session.

Unlike former studies (e.g., Dixon et al., 1993; Millis, 1995; Kuijpers & Hakemulder, 2018), in our study rereading did not significantly affect readers' appreciation. However, when we checked the results for each sonnet separately, the effect reappeared, insofar as readers liked sonnet 66 slightly more after the last session than after the first (first session: $M_{\text{sonnet66}} = 2.76$, $SD_{\text{sonnet66}} = .83$; last session: $M_{\text{sonnet66}} = 3.12$, $SD_{\text{sonnet66}} = 1.05$). For sonnet 27 the difference was not significant, though (first session: $M_{\text{sonnet27}} = 3.88$, $SD_{\text{sonnet27}} = .67$; last session: $M_{\text{sonnet27}} = 3.92$, $SD_{\text{sonnet27}} = .81$). Whether this difference is the result of a ceiling effect (sonnet 27 was already well appreciated after the first session) or the result of different levels of general comprehensibility (sonnet 66 has longer and less frequent words than sonnet 27, e.g., standardized word length: $M_{\text{sonnet66}} = .24$, $SD_{\text{sonnet66}} = 1.10$; $M_{\text{sonnet27}} = -.20$, $SD_{\text{sonnet27}} = .87$; standardized word frequency: $M_{\text{sonnet66}} = -.18$, $SD_{\text{sonnet66}} = 1.13$; $M_{\text{sonnet27}} = .15$, $SD_{\text{sonnet27}} = .86$) has to be tested in future studies.

Besides assessing reading behavior by ratings, we also applied eye tracking as an indirect online method to measure ongoing cognitive and affective processes associated with comprehension and appreciation. Linear mixed model analyses confirmed that rereading increases reading fluency, even in the case of poetry, as shown by a decrease in regression time and total reading time, which are typical of later stages of the process of reading and comprehension. The skipping rate was likewise higher in the last session, so the probability of fixating on any word was smaller during the last session. Rereading seemed to have no effect on first fixation and gaze durations, though. As already mentioned, analysis of eye

tracking parameters associated with the early stages of the process has not led to consistent findings, especially when various psycholinguistic features were taken into account (Raney, Therriault, & Minkoff, 2000; Kinoshita, 2006; Chamberland et al., 2013). In our study, first fixation and gaze durations were nearly the same in the last session as in the first, likely because these parameters reflect fast and automatic initial word recognition processes (cf. Hyönä & Hujanen, 1997; Clifton, Staub, & Rayner, 2007) hardly affected by rereading.

6.5.2 QNA-based predictive modeling approaches

By using machine-learning tools, complex relationships in and between data sets can be disentangled and identified (e.g., Coit, Jackson, & Smith, 1998; Francis, 2001; Breiman, 2001; Tagliamonte & Baayen, 2012; Yarkoni & Westfall, 2017; LeCun, Bengio, & Hinton, 2015). Among the many machine-learning tools, neural nets may be the most suitable for psychological studies, since they make use of an architecture inspired by the neurons in the human brain (LeCun, Bengio, & Hinton, 2015). In neural nets, data are transmitted from an input layer over one or more hidden layer(s) to the output layer assigning different weights to all connections between layers during the learning/training phase. The neural nets' hidden layer(s) also performs a dimension reduction on correlated predictors. Therefore, the approach appears advantageous for studies on natural reading in which multiple psycholinguistic and context features may play a role (Jacobs, 2015a, 2018a). In Xue et al. (2019), the neural nets approach proved to be the optimal one in predicting two eye tracking parameters (total reading time and fixation probability) using seven surface features.

In the present study we successfully replicated the findings of Xue et al. (2019) about reading Shakespeare's sonnets: 1) the neural nets approach was the best way to predict the total reading time and fixation probability using a set of nine psycholinguistic features; 2) *word length*, *word frequency*, *orthographic neighborhood density* and *sonority score* were most important in predicting total reading time and fixation probability for poetry reading, and *orthographic dissimilarity* proved to be important for total reading time. Nevertheless, comparing the results of this study with those of Xue et al. (2019) uncovers some differences. In this present rereading study also the *consonant vowel quotient* was indicated as a potentially important feature for total reading time (first session) and fixation probability (first and last session). This finding of two important phonological features, *sonority score*, and *consonant vowel quotient*, is in line with the assumption that consonant status and

sonority also play a role in silent reading (Maïonchi-Pino et al., 2008; Berent, 2013), especially of poetic texts (Kraxenberger, 2017).

In contrast to Xue et al. (2019), neural nets also produced acceptable model fits for the first fixation duration. That was also true for gaze duration and regression time, two eye tracking parameters not tested in Xue et al. (2019). For all three parameters, neural nets outperformed the standard least square analysis. The calculation of the *FIs* indicated that *word length*, *word frequency*, *orthographic neighborhood density*, and *sonority score* were important in predicting first fixation duration, gaze duration and regression time for poetry reading, except that *word length* was less important for predicting first fixation duration in the first reading session ($FI = .09$). Crucially, we found that the positional information, i.e., word number in a certain line, was important in predicting first fixation and gaze durations, which again supports the idea that these measures reflect fast and automatic reading behavior and are less sensitive to lexical features (Hyönä & Hujanen, 1997; Clifton, Staub, & Rayner, 2007).

By applying the predictive modeling approach, we also wanted to find out which psycholinguistic features may cause potential differences in eye tracking parameters for the first and last sessions. The comparison of five eye tracking parameters for the first and last reading indicated a significant decrease in regression time, total reading time, and fixation probability for the last session. More interestingly, the basic features most important in the first session were also the most important ones in the last. Surface features like *word length*, *word frequency*, *orthographic neighborhood density*, and *sonority* thus seem to be basic to eye movement behavior in reading and to remain so, no matter how many times a text is read. However, since most of the surface features important in one session were also important in the other, it remains unclear why total reading and regression times decreased in the last session. Perhaps this was due to changes in the importance of other lexico-semantic or complex interlexical and supralexical features (e.g., syntactic complexity; Lopopolo, Frank, & Willems, 2019) across reading sessions. As illustrated in Figure 6.4, the overall model fits were slightly decreased across sessions for all eye tracking parameters except for fixation probability. This could indicate that while surface features play a lesser role, other features become more important, leaving a lot to explore in future research on eye movements in poetry reading.

In conclusion, by using a rereading paradigm, we examined the effects of reading and rereading Shakespeare's sonnets. Besides assessing reading behavior by rating and examining cognitive processes by using the eye tracking technique, we also checked the roles of surface psycholinguistic features across reading sessions by using predictive modeling. Our study confirmed not only the benefit of rereading a text usually obtained with non-literary materials, but also the advantages of neural nets modeling, as well as the key importance of surface psycholinguistic features in all sessions of reading.

6.6 Limitations and Outlook

In this study, we remedied two shortfalls of Xue et al. (2019). Firstly, we included positional information (line number and the position of the word in the line) in the predictive modeling, to compensate for potential position effects (Pynte, New, & Kennedy, 2008; Pynte, New, & Kennedy, 2009; Kuperman et al., 2010). We found that they were indeed important ($FIs > .10$) for predicting first fixation duration and gaze duration, but not for predicting regression and total reading time or fixation probability. Secondly, we enlarged our sample size by recruiting more readers. Despite the changes, results were much the same: the neural nets approach was the most suitable one, and the key features again were *word length*, *word frequency*, *orthographic neighborhood density*, and *sonority score*.

Of course, there is still room for further improvement. Firstly, we used only two sonnets, not to strain readers, but for some predictors (e.g., *higher frequent neighbors*, $M = .55$, $SD = 1.11$) two short texts may not produce sufficient variation. In future studies, our findings should therefore be checked with more and different poems (Fechino, Jacobs, & Lüdtké, 2020). Secondly, according to the multilevel hypothesis of the NCPM (e.g., Hsu et al., 2015a; Jacobs et al., 2016), many fore- and backgrounding features, especially on the interlexical and supralexical levels, also contribute to the highly complex literary reading process. Before we can efficiently include them in empirical eye tracking studies, we still have to identify, define, and classify them more reliably, though. However, existing classification schemes often overlap or are inconsistent or incomplete (cf. Leech, 1969). Certainly, there are some promising approaches to quantifying the occurrence of rhetorical figures (Gambino & Pulvirenti, 2018; Jakobson & Lévi-Strauss, 1962; Jacobsen, 2006; Jacobs, 2015a, 2017, 2018a; Jacobs & Kinder, 2017, 2018), but many questions remain open, as regards, for instance, possible weightings. Thirdly, for predictive modeling, we aggregated the eye tracking data over participants, which may inflate certain psycholinguistic effects

(Kliegl, Olson, & Davidson, 1982; Lorch & Myers, 1990). However, in neural nets, it is not possible to consider subject effects as a random effect like in linear mixed models (e.g., Baayen, Davidson, & Bates, 2008). To make model comparisons possible, we thus had to use the aggregated values for both approaches.

Author Contributions

All authors contributed to the design of the experiment. Xue S. carried out the experiment, analyzed the data, and wrote the first draft of the manuscript; Lüdtke J. modified the manuscript; Jacobs A. M. improved the manuscript. All authors have contributed to and approved the final manuscript.

Chapter 7: The Role of Affective-semantic Features in the (Re-)Reading of Shakespeare's Sonnets: A Reanalysis⁴

Shuwei Xue Arthur M. Jacobs, and Jana Lüdtke

7.1 Abstract

Literary reading is an important activity in leisure time, while the nature of it is largely unknown. In the present study, we reanalyzed the data of one study examining the roles of seven lexical surface psycholinguistic features in poetry (re-)reading. By applying quantitative narrative analysis (QNA), two lexical affective-semantic features, *valence* and *arousal*, were also computed and added to predict the eye movements of readers while reading. Using neural nets as a machine learning-based predictive modeling approach, we replicated former results that no matter how many times readers had read the poem, the surface features always stand out when predicting aggregated measures of gaze duration, regression time, total reading time and fixation probability. Moreover, the reanalysis shows that word-based *valence* and *arousal* also played an important role. For both features, we observed an increase in the importance from the first to the last reading. Especially for *valence*, feature importance values observed in the last reading were as high as the values for the most important surface features. We assume that surface features lay the foundation for poetry reception, and once a first understanding of the meaning is established, readers start to pay more attention to the affective-semantic aspects.

Keywords: poetry reading, rereading, affective-semantic features, QNA, predictive modeling, eye tracking

⁴ This chapter is preparation as “Once Known, Twice Hedonic: Enjoying Shakespeare’s Sonnets Through Rereading: A Reanalysis” at the time of dissertation submission.

7.2 Introduction

*Music to hear, why hear'st thou music sadly?
Sweets with sweets war not, joy delights in joy.*

William Shakespeare, Sonnets 8 (ll. 1-2)

The process of a poetry reading is something of a juggling feat, in which various text features are dealt with at nearly the same time. It is still an open question which features drive the process of poetry reading and understanding (Jacobs, 2015a, b), which features are more important for first reading (e.g., Ullrich et al., 2017; Xue et al., 2019) and whether the role of these features may change throughout multiple reading sessions (e.g., Xue, Jacobs, & Lüdtkke, 2020). To explore the assumed interaction between text and reader especially in poetry reading, scholars from humanities focussed on the qualitative analysis of the poem and the possible effects on an ideal reader (e.g., Jakobson & Jones, 1970; Simonto, 1989; Vendler, 1997); few empirical studies focussed on literary experiences observed in real readers using offline methods like questionnaires, memory tasks or interpretations (e.g., Lüdtkke, Meyer-Sickendieck, & Jacobs, 2014; Jacobs et al., 2016; Kuiken, Campbell, & Sopčák, 2012; Kraxenberger & Menninghaus, 2017) and even less empirical studies used online methods like thought protocols, eye tracking or fMRI to get more information about the underlining processes (e.g., Hoffstaedter, 1987; van't Jagt, Hoeks, Dorleijn, & Hendriks, 2014; Gao & Guo, 2018). Nevertheless, it seems that there is a gap between mainly theoretical qualitative literary studies and empirical research about poetry reading.

To foster empirical investigations using (more) natural and complex materials like poems, recently a quantitative narrative analysis (QNA) based predictive modeling approach has been introduced (e.g., Jacobs, 2015a, 2017, 2018a, 2019; Jacobs, Hofmann, & Kinder, 2016; Jacobs et al., 2017; Jacobs & Kinder, 2017, 2018). Xue and colleagues (2019, 2020) tested this approach for reading and rereading Shakespeare's sonnets following a two-step procedure: Firstly, the complex natural texts, in these cases of Xue et al., the Sonnets 27, 60 and 66, were broken up into measurable and testable features by using databases, linguistic corpora and computer programs (Franzosi, 2010) for the sake of statistical analysis. Secondly, a machine learning-based predictive modeling approach was used to test how different features influence reading behavior measured online while reading and rereading. In

comparison to classical multiple linear regression analysis, machine learning-based predictive modeling approaches, like neural nets, sought to deal with the non-linear webs of correlations of these features (e.g., Willems, 2015; Willems & Jacobs, 2016; Jacobs & Willems, 2018). However, the two studies by Xue and colleagues (2019, 2020) only concentrated on seven “cold” surface psycholinguistic features, like *word length* and *word frequency*, which are well established as important to predict eyetracking measures like total reading times or fixation probability. To get a better understanding of poetry reading and rereading and to foster a fruitful exchange between theoretical qualitative literary studies and empirical research about poetry reading, the list of features needs to be broadened. One important candidate are “hot” affective and semantic features, which are assumed to be relevant for the effect of poems to express meaning and evoke feelings (e.g., Brewer & Lichtenstein, 1982; Nell, 1988; Oatley, 1995). In the present study, to remedy the above deficiency, we reanalysed the rereading data of Xue, Jacobs, and Lüdtke (2020), while using a broader range of features including not only the seven surface features (*word length*, *word frequency*, *orthographic neighborhood density*, *higher frequent neighbors*, *orthographic dissimilarity*, *consonant vowel quotient* and *sonority score*) but also two well established affective-semantic ones, namely word-based *valence* and *arousal*.

7.2.1 Lexical features in reading

Reading, in general, is influenced by numerous features, including characteristics of readers, characteristics of the situation in which reading takes, and characteristics of the text (Jacobs, 2015a, b). Many empirical studies in the field of reading focused on the last aspects. For instance, empirical research identified over 50 features of words that influence single word recognition (Graf, Nagler, & Jacobs, 2005). The QNA helps to compute over 100 features for the corpus of Shakespeare’s sonnets (Jacobs et al., 2017). In the emerging field of Neurocognitive Poetics, a 4×4 feature matrix has been developed to guide research in this field (Jacobs, 2011, 2015a, b, 2018a, b; Nicklas & Jacobs, 2017; Willems & Jacobs, 2016), which states that the processing of literary texts like narratives or poetry is influenced by a whole set of sublexical, lexical, interlexical, and supralexical features at the metric, phonological, morpho-syntactic and semantic levels. Examples of the features have been proved by several studies using materials from the various length and complexity like single words, proverbs or whole poems (Aryani et al., 2016; Jacobs & Lüdtke, 2017; Jacobs et al., 2015, 2016; Jacobs, Hofmann, & Kinder, 2016; Ullrich et al., 2017). Since it is not possible

to identify all relevant features characterizing a natural text, the big majority of reading studies tested only a few selected features while ignoring the others, for instance in eye tracking studies using literary materials (Müller et al., 2017; van den Hoven et al., 2016; Xue et al., 2019; Xue, Jacobs, & Lüdtke, 2020). Especially eye tracking studies focused on word-based measures and lexical features. Besides, including more features also requires extending sample size (i.e., more/longer texts and more participants), so we followed this approach and concentrated on lexical features.

Lexical surface features used in studies of Xue et al. (2019; 2020).

Among all the lexical features, *word length* and *word frequency* are the most widely discussed as important variables predicting word-based measures like first fixation duration gaze duration or fixation probability. It is said that short and frequent words would be easily recognized and processed, leading to shorter reading time and fewer chances being fixated (e.g., Just & Carpenter, 1980; Inhoff & Rayner, 1986; Raney & Rayner, 1995; Pynte, New, & Kennedy, 2008). As for the features related to orthography, it is acknowledged that reading would be facilitated if a word has more orthographic neighbors (Coltheart et al., 1977; see Andrews, 1997, for a review). However, if its orthographic neighbors were more frequent than the word itself, the effect may turn to be inhibitory (Grainger et al., 1989; Grainger & Jacobs, 1996; Perea & Pollatsek, 1998). While testing the machine learning modeling tools Xue and colleagues (2019, 2020) used also some less studied lexical features. The feature, *orthographic dissimilarity*, is calculated using the Levenshtein distance metric based on words of the same lengths in a certain corpus, whose role in reading has not been widely investigated. Moreover, Xue and colleagues (2019, 2020) also included two surface features related to phonological aspects, the *consonant vowel quotient* and the *sonority score*, it is supposed that consonant status and sonority play a role in silent reading (Maionchi-Pino et al., 2008; Berent, 2013), especially of poetic texts (Kraxenberger, 2017).

In line with other studies, Xue et al. (2019) found that *word length* and *word frequency* were important features to predict the eye movements of readers in poetry reading. Besides, *orthographic neighborhood density*, *orthographic dissimilarity*, and *sonority score* were important features, too. Moreover, in the second study focussing on rereading poetry, Xue, Jacobs, and Lüdtke (2020) found that these predictors were consistently important in both the first and the last reading sessions, showing the fundamental roles of these surface features. However, as mentioned above, the two studies only considered the “cold” side of

the poetic materials, i.e., the roles of surface features. To the best of our knowledge, no empirical study using online methods to explore the reading of poems had checked the influence of multiple surface features and the affective-semantic ones in one study, although affective-semantic features are assumed to be very important for literary reading (Brewer & Lichtenstein, 1982; Nell, 1988; Oatley, 1995).

Affective-semantic features in reading and reading poetry.

Readers are quite sensitive to the affective-semantic information since they can guide our attention and help to construct the representation of the situation described in a text (van Dijk & Kintsch, 1983; Megalaki, Ballenghein, & Baccino, 2019). The most frequently investigated affective-semantic features are *valence* (often defined as the degree of positive or negative affect) and *arousal* (often defined as the degree of internal activation). Generally, studies focussing on word-based measures showed that emotional words identified by lexical valence and arousal values have a processing advantage over neutral words. For instance, emotional words with high arousal are recognized more quickly in lexical decision tasks (e.g., Hofmann et al., 2009; Schacht & Sommer, 2009; Scott et al., 2009) and recalled more often in memory tasks (e.g., Kissler et al., 2007) than non-emotional words. Scott, O'Donnell, and Sereno (2012) conducted one of the first eye tracking studies on processing differences of emotional and neutral words in sentences. Following the well-known processing advantage of emotional words in studies on single word processing, first fixation duration and gaze duration on emotion words were shorter than those on neutral words. Since Scott, O'Donnell, and Sereno (2012), the processing advantage for emotional words is replicated by several other eye tracking studies (for an overview see Lüdtko Kaakinen & Jacobs, in press).

However, empirical studies about the influence of affective-semantic features on reading poetry are only constrained to offline methods, like answering rating questions after reading. For example, Kraxenberger and Menninghaus (2017) tested the differences in appreciation between joyful and sad poems. They found that compared to joyful poems sad poems got higher aesthetic appreciation. While this result nicely demonstrated how affective-semantic features at the supralexical level influenced the reading and interpretation of a poem, they say nothing about the role of affective-semantic features at the lexical level like word-based valence and arousal. To make sure which lexical features may influence the valence ratings at the supralexical level, Ullrich et al. (2017) collected ratings (at supralexical level) on eight different *general affective meaning* scales—valence, arousal, friendliness,

sadness, spitefulness, poeticity, onomatopoeia, and liking—for 57 German poems (“*die verteidigung der wölfe*”) which the contemporary author H. M. Enzensberger had labeled as either “friendly”, “sad”, or “spiteful”. They found that word-based valence as a lexical feature accounts for a major amount of up to 50% of the variance in affective ratings at supralexical level. If word-based valence influences the supra-lexical ratings in poetry reading, we assume that word-based valence (and also arousal) also influence online reading behavior as demonstrated for non-poetic texts (e.g., Scott, O’Donnell, & Sereno, 2012). Besides the fact that until today no study focused on the influence of word-based *valence* and *arousal* on reading using online methods like eye tracking, also no study tested the influence of affective-semantic lexical features in a rereading paradigm.

And more notably, in the present study, we only extracted the content words from the data of Xue, Jacobs, and Lüdtke (2020) to add two lexical affective-semantic features, *valence* and *arousal*, as these cannot be computed for grammatical words.

7.2.2 A machine learning-based predictive modeling approach

In natural reading, as mentioned, one critical point needs to be addressed: psycholinguistic features especially at the lexical level do not play their role independently. Instead, all of them are nonlinearly intercorrelated with each other in influencing the reading behavior of readers. For example, in sentence reading, Scott, O’Donnell, and Sereno (2012) found that the processing advantage of negative words was dependent on word frequency, that is, emotion words (positive or negative) were read consistently faster than neutral words except in the case of negative words with high frequency. Although many studies have noticed the complexity of the relationship between psycholinguistic features, nearly all empirical studies we know of concentrated only on a few selected features and controlled the others by using experimentally designed materials (e.g., Rayner et al., 2001; Reichle, Rayner, & Pollatsek, 2003; Rayner & Pollatsek, 2006; Engbert et al., 2005; Rayner, 2009). This procedure does not allow the usage of natural reading material like poetry. That means especially for research on reading poetry, it is necessary to find out an approach that could take multiple psycholinguistic features into account and handle the complex relationships between them.

A machine learning-based predictive modeling approach has been proposed to handle the above problems in studies of literary reading (Jacobs, 2017, 2018b; Jacobs & Kinder, 2017, 2018; Jacobs et al., 2017). 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). Actually, this approach has already been successfully applied in many fields, e.g., in the fields of bioinformatics (Strobl, Malley, & Tutz, 2009), ecology (e.g., Manel et al., 1999; Were, Bui, Dick, & Singh, 2015), geology and risk analysis (Nefeslioglu, Gokceoglu, & Sonmez, 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 & Kaiser, 2019) or applied reading research (Lou et al., 2017; Matsuki, Kuperman, & Van Dyke, 2016).

For poetry reading, Xue et al. (2019) found that a machine-learning-based predictive modeling approach (i.e., neural nets model) outperformed the other approaches when using seven surface psycholinguistic features to predict relevant eye movement parameters. Xue, Jacobs, and Lüdtkke (2020) replicated this finding in a rereading study of poetry. The neural nets model is a multilayer perceptron which can predict one or more response variables using a flexible function of the input variables. It can implicitly detect all possible (nonlinear) interactions between predictor variables and many other advantages over general linear regression models when dealing with complex stimulus-response environments (e.g., Tu, 1996). It would also be evaluated in a predictive modeling approach comparing the goodness of fit index (R^2) for training and test sets. Unlike traditional linear approaches, this approach can build models with multiple predictors and provide the relative importance of each predictor in predicting the response parameters.

As revealed by the two studies from our group (Xue et al., 2019; Xue, Jacobs, & Lüdtkke, 2020), five surface features (*word length*, *word frequency*, *orthographic neighborhood density*, *consonant vowel quotient*, and *sonority score*) were important in influencing relevant eye parameters. Moreover, the important features in influencing relevant eye tracking parameters were the same ones across reading sessions, although the eye tracking parameters themselves changed significantly across sessions. Based on the resource allocation theory (Millis & Simon, 1994; Millis, Simon, & TenBroek, 1998), the processing of the surface features is automatic and obligatory, thus their roles should be consistent across readings. However, the processing of high-level features, such as the “hot” affective-semantic features, may be effortful and optional. After a prior reading, readers might have more free-up resources which could be redistributed to high-level processes. This might be extremely

true in the case of literary reading. Britton et al. (1983) found that literary texts demanded greater investment in processing resources than expository texts.

7.2.3 Aims of the current study

In the present study, we reanalyzed the data obtained by Xue, Jacobs, and Lüdtké (2020) by adding two well established affective-semantic features in the analyses. That is we predict the eye movements of readers at the lexical level using both the seven surface features (*word length, word frequency, orthographic neighborhood density, higher frequent neighbors, orthographic dissimilarity, consonant vowel quotient, and sonority score*) and the two features about positional information (line number and the position of the word in the line) extended by the two the affective-semantic features *valence* and *arousal*. Specifically, unlike the other studies focusing on the interplay or the time course of two or three features, we were interested in the relative importance of all the eleven features in the single reading and rereading of poetry. We specifically determined to find out whether the affective-semantic features are important for predicting eye movement measures at the lexical level.

Based on the findings of Xue et al. (2019) and Xue, Jacobs, and Lüdtké (2020), we supposed that neural nets could build satisfactory models even when focusing on the content words only. Moreover, we assumed that the important surface features would again pop out even when two new features were added in the analyses. Additionally, we were not sure about how the aesthetic-semantic features may influence the eye movements of readers. Although previous studies detected some significant effects of these features, none of them used poetic texts in a rereading paradigm. Their effects might be decreased since there was no experimental manipulation. According to the resource allocation theory (Millis & Simon, 1994; Millis, Simon, & TenBroek, 1998), we also assumed that after the first reading session readers might have more resources to the affective-semantic aspects of the poem. Therefore, the importance of the two affective-semantic features might be increased in the last reading session compared to the first reading session.

7.3 Method

7.3.1 Participants & Apparatus & Materials & Procedure

These sections are the same as study 2.

7.3.2 Data analysis

Paper-pencil task.

The general emotional responses on poem level could be revealed by two questions about *supralexical valence* and *supralexical arousal*, extracted from the paper-pencil task (Papp-Zipernovszky et al., in preparation). They were respectively evaluated by two questions: “In general, poems can express positive or negative emotions. While reading this poem for the first time, did it feel negative or positive?”; “In general, poems can evoke feelings ranging from boredom to excitement. While reading this poem for the first time, did it feel calming or exciting?”. For the question about *supralexical valence*, readers indicated their agreement with the statements on a 7-point rating scale ranging from -3 = extremely negative to 3 = extremely positive. For *supralexical arousal*, readers indicated their agreement with the statements on a 5-point rating scale ranging from 1 = very calming to 5 = very exciting.

JMP 14 Pro (https://www.jmp.com/en_us/software/predictive-analytics-software.html) was used for the statistical analyses. We used paired-samples *t* tests, to check the differences between the first session and the last.

Predictors for the machine learning-based predictive modeling approach.

We got the data from Xue, Jacobs, and Lüdtkke (2020), including the positional information, the data of seven surface psycholinguistic features for all words, and the data of all eye tracking parameters. In the present study, we added the two affective-semantic features, word-based *valence* and *arousal*, to compensate for the missing of this kind of feature in the two studies on Shakespeare’s sonnets (Xue et al., 2019; Xue, Jacobs, & Lüdtkke, 2020). However, we only calculated the affective-semantic features for all open-class words (adjective, adverb, noun, and verb), because there is no point in calculating the affective-semantic values for the close-class words (refers to the category of function words).

Positional information. Several words are repeated in the sonnets (e.g., mind), so we added the positional information (*lineNo.* and *wordNo.*) of the words in each sonnet.

Surface features. *Word length (wl)* is the number of letters per word; *word frequency (logf)* is the log transformed number of times that a word appears in the Gutenberg Literary English Corpus as a reference (GLEC; Jacobs, 2018b); *orthographic neighborhood density (on)* is the number of words of the same length as a certain word and differing by only one

letter in GLEC; *higher frequent neighbors (hfn)* is the number of orthographic neighbors with a higher word frequency than the word in GLEC; *orthographic dissimilarity (odc)* is the word's mean Levenshtein distance from all other words in the corpus (GLEC), a metric that can be generalized to apply 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 two sonnets, ART got the *sonscore* of $10 \times 1 [a] + 7 \times 1 [r] + 1 \times 1 [t] = 18 / \text{SQRT}(3) = 10.39$.

Affective-semantic features. The calculation of the two affective-semantic features, *valence (val)* and *arousal (aro)*, is based on a hybrid method combining the traditional method based on a dictionary or list of words and the computational method based on a vector space (the VSM or Vector Space Model; Taboada et al., 2011; Jacobs, 2019). We used the training corpus (GLEC) to estimate the semantic similarity of a certain word in the poem to each of 12 word labels (seven positive labels, such as HAPPINESS or PRIDE, and five negative ones such as DISGUST or FEAR; Ekman, 2005; Westbury et al., 2015) for which *valence* and *arousal* rating-data are available. The *valence* and *arousal* value of a certain word is the average of the ratings of its *k nearest neighbors* in the vector space.

Eye tracking parameters. We decided only to include five aggregated eye tracking parameters (the data over all subjects to obtain the mean values for each word-token within each sonnet): the first fixation duration (the duration of the first fixation on a certain word), the gaze duration (the sum of all fixations on a certain word during first passage), the total reading time (the sum of all fixation durations on a certain word), the regression time (the sum of fixation durations on a certain word after the first passage), the fixation probability (the amount of skipping is taken into account in calculating the fixation probability; fixation probability = [number of subjects fixated on a certain word/ total amount of the subjects] × 100%).

Machine learning-based predictive modeling approach.

We only used the approach of neural nets model, because compared to bootstrap forests model and standard least squares regression, it got the most satisfactory model fits for both studies (Xue et al., 2019; Xue, Jacobs, & Lüdtke, 2020). Altogether eleven predictors (*lineNo.*, *wordNo.*, *wl*, *logf*, *on*, *hfn*, *odc*, *cvq*, *sonscore*, *val*, and *aro*) were used to predict

five eye tracking parameters (first fixation duration, gaze duration, total reading time, regression time and fixation probability). Most importantly, we only did the predictive modeling analyses for all open-class words (121 words). The values of all psycholinguistic features and eye tracking parameters were standardized before putting them into predictive modeling analysis.

Feature importances (*FIs*) were also calculated. *FI* was a term used in machine learning (https://scikit-learn.org/stable/modules/feature_selection.html). In the current study, 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 was 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 was interpreted as an ordinal value on a scale of 0 to 1 with *FI* values $> .1$ considered ‘important’ (Strobl, Malley, & Tutz, 2009).

7.4 Results

7.4.1 Paper-pencil task

On poem level, in general, readers thought that the poems did not express strong positive or negative emotions (*supralexical valence*: $M_{\text{first-session}} = -.64$, $SD_{\text{first-session}} = 1.47$; $M_{\text{last-session}} = -.48$, $SD_{\text{last-session}} = 1.64$) and did not evoke very calming or very exciting feelings (*supralexical arousal*: $M_{\text{first-session}} = 3.22$, $SD_{\text{first-session}} = .84$; $M_{\text{last-session}} = 3.22$, $SD_{\text{last-session}} = .84$). The results of the rereading effects on *valence* and *arousal* are shown in Figure 7.1: there was no significant difference between sessions in both *supralexical valence* ($t(49) = -1.24$, $p = .22$) and *supralexical arousal* ($t(49) = .00$, $p = 1.00$).

Taken into account the result of Xue et al. (2019) demonstrating that the effect of rereading on appreciation ratings was visible only for sonnet 60, we also checked for each sonnet separately by applying a paired-samples *t* test. For sonnet 27, there was no significant difference in the rating of *supralexical valence* and *supralexical arousal*, whether it was read in the first or last session (*supralexical valence*: $t(24) = -1.78$, $p = .09$; $M_{\text{first-session}} = .40$, $SD_{\text{first-session}} = 1.26$; $M_{\text{last-session}} = .68$, $SD_{\text{last-session}} = 1.25$; *supralexical arousal*: $t(24) = .00$, $p = 1.00$; $M_{\text{first-session}} = 3.40$, $SD_{\text{first-session}} = .96$; $M_{\text{last-session}} = 3.40$, $SD_{\text{last-session}} = .96$). This is also true for sonnet 66 that there was no significant difference between session in both *supralexical valence* ($t(24) = -.20$, $p = .85$; $M_{\text{first-session}} = -1.68$, $SD_{\text{first-session}} = .75$; $M_{\text{last-session}} =$

-1.64, $SD_{\text{last-session}} = 1.08$) and *supralexicalexical arousal* ($t(24) = .00$, $p = 1.00$; $M_{\text{first-session}} = 3.04$, $SD_{\text{first-session}} = .68$; $M_{\text{last-session}} = 3.04$, $SD_{\text{last-session}} = .68$).

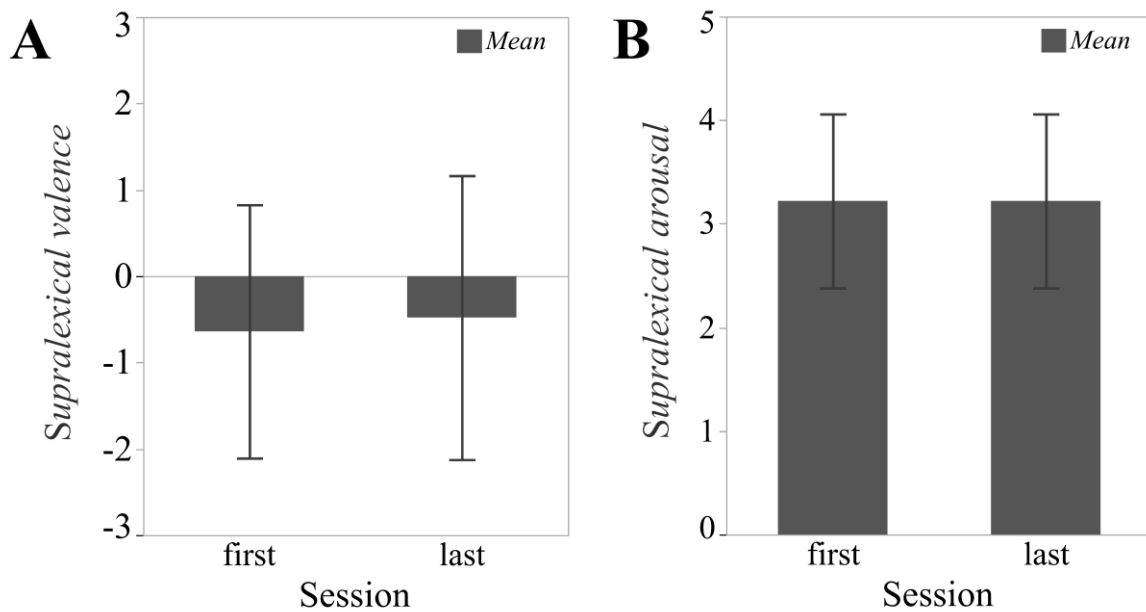


Figure 7.1 Rereading Effect on Rating Data

7.4.2 Machine learning-based predictive modeling approach

Figure 7.2 shows the overall R^2 (100 iterations) for predicting the five eye tracking parameters using neural nets. As illustrated in Figure 1, generally neural nets produced acceptable models for all five eye tracking parameters (mean $R^2 > .30$). Therefore, the eleven FIs for the neural nets were computed (see Figure 7.3). Below we illustrate our results for the five eye tracking parameters, respectively.

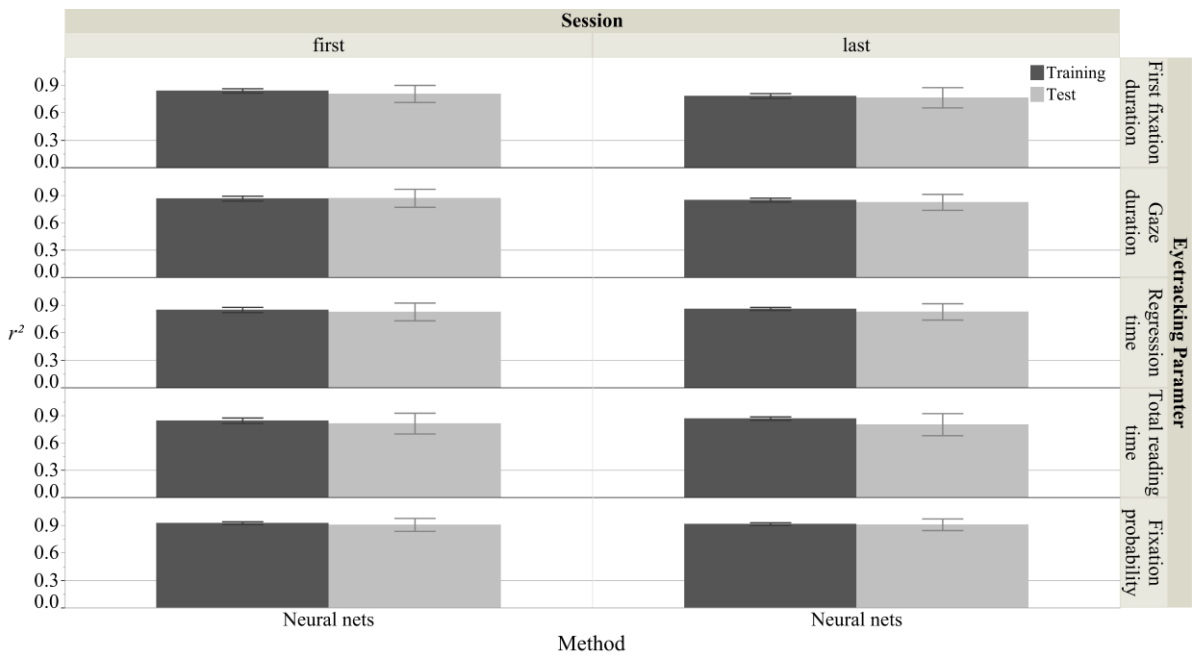


Figure 7.2 Fit Scores for Different Models and Measures

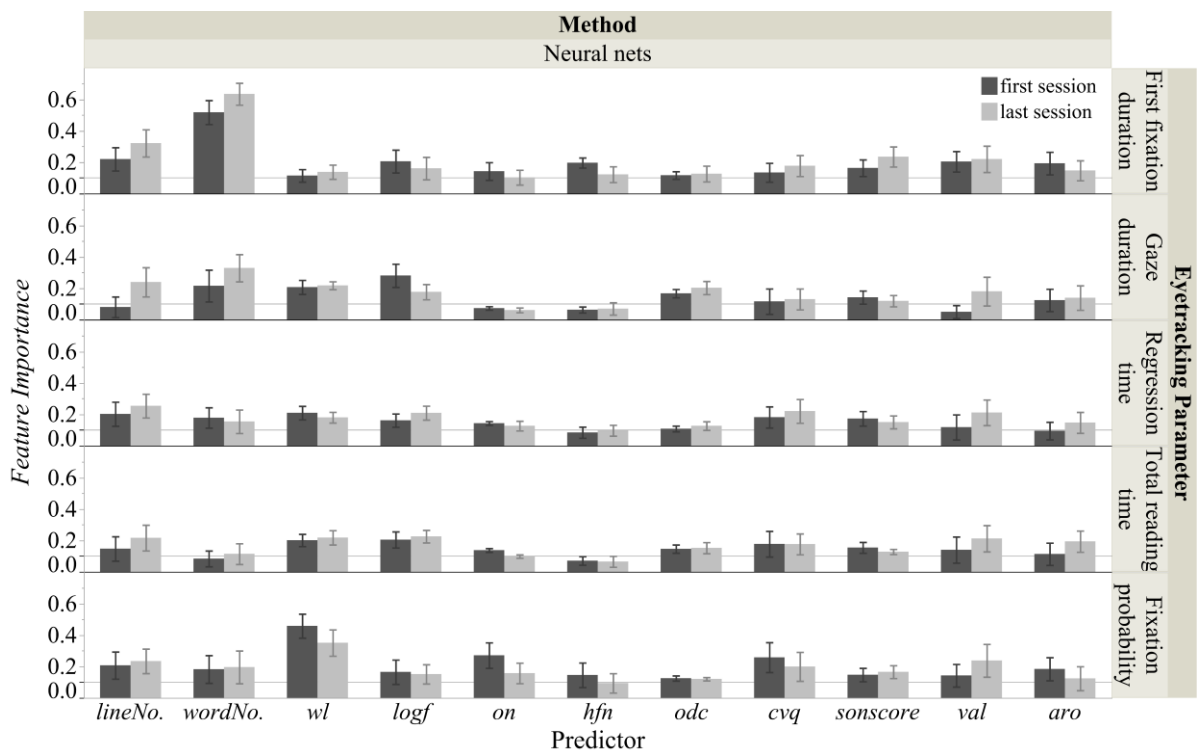


Figure 7.3 Feature Importance for the Five Eye Tracking Parameters

First fixation duration.

As shown in Figure 7.2, in the first session, neural nets produced good fits for both the training and test sets (mean $R^2_{train} = .84$, $SD R^2_{train} = .02$; mean $R^2_{test} = .81$, $SD R^2_{test} = .09$). The same was true for the last session (mean $R^2_{train} = .78$, $SD R^2_{train} = .03$; mean $R^2_{test} = .76$, $SD R^2_{test} = .11$). For both training and test sets, the model fits in the last session were lower than those of the first session.

The *FI* analysis of neural nets approach in Figure 7.3 suggested that in the first session all the predictors were important for predicting first fixation duration (*wordNo.* [.52], *lineNo.* [.22], *logf* [.21], *val* [.21], *hfn* [.20], *aro* [.19], *sonscore* [.16], *on* [.14], *cvq* [.14], *odc* [.12], *wl* [.12]). Similarly, in the last session also all predictors were important (*wordNo.* [.64], *lineNo.* [.32], *sonscore* [.24], *val* [.22], *cvq* [.18], *logf* [.16], *aro* [.15], *wl* [.14], *odc* [.13], *hfn* [.12], *on* [.10]). While *wordNo* coding the position of a word on a line had the highest feature importance in the first and last reading, the feature importance for the affective-semantic feature *val* was on the same level as feature importance for the most important surface features *logf* (for first reading) and *sonscore* (for last reading). The feature importance for *aro* was slightly lower compared to the values for *val*, but always above the critical value of .01.

Gaze duration.

Figure 7.2 also shows that for both sessions, neural nets produced acceptable fits (first session: mean $R^2_{train} = .87$, $SD R^2_{train} = .03$; mean $R^2_{test} = .87$, $SD R^2_{test} = .10$; last session: mean $R^2_{train} = .85$, $SD R^2_{train} = .02$; mean $R^2_{test} = .83$, $SD R^2_{test} = .09$). Again, for both training and test sets, the model fits in the last session were lower than those of the first session.

The *FI* analysis of neural nets approach shown in Figure 7.3 suggested that in the first session, seven predictors were important for predicting gaze duration (*logf* [.28], *wordNo.* [.22], *wl* [.21], *odc* [.17], *sonscore* [.14], *aro* [.12], *cvq* [.12]), while *lineNo.* (.08), *on* (.07), *hfn* (.06) and *val* (.05) were less important. For the last session, there were nine important predictors (*wordNo.* [.33], *lineNo.* [.24], *wl* [.22], *odc* [.20], *val* [.18], *logf* [.18], *aro* [.14], *cvq* [.13], *sonscore* [.12]), while this time the less important ones were *hfn* (.07) and *on* (.06). While the feature importance for *val* was below the critical value of 0.1 in the first reading, *val* turned out to be an important factor in predicting gaze duration in the last reading. Here, the feature importance for *val* was as high as the importance for *logf*, one of the most

established predictors for word-based eye tracking measures. The values for *aro* were always above .1, they also increased slightly from first to last reading.

Regression time.

As illustrated in Figure 7.2, similarly, neural nets produced acceptable fits for both sessions (first session: mean $R^2_{train} = .85$, $SD R^2_{train} = .03$; mean $R^2_{test} = .83$, $SD R^2_{test} = .10$; last session: mean $R^2_{train} = .86$, $SD R^2_{train} = .02$; mean $R^2_{test} = .83$, $SD R^2_{test} = .09$). For the training sets, the model fits increased a little bit in the last session.

Figure 7.3 shows the *FI* analysis of the optimal neural nets approach suggesting that in the first session, ten predictors were important for regression time (*wl* [.21], *lineNo.* [.20], *cvq* [.18], *wordNo.* [.18], *sonscore* [.17], *logf* [.16], *on* [.14], *val* [.12], *odc* [.11], *aro* [.10]), while *hfn* (.09) were less important. For the last session, all predictors were important (*lineNo.* [.26], *cvq* [.22], *val* [.21], *logf* [.21], *wl* [.18], *wordNo.* [.16], *sonscore* [.15], *aro* [.15], *odc* [.13], *on* [.13], *hfn* [.10]). All values for the *val* and *aro* were always above the critical value of .1. For both affective-semantic features, we observed an increase in feature importance from first to last reading which was more pronounced for *val* compared to *aro*. In the last reading, the importance for *val* was at the same level as the value of the most important surface feature *cvq*.

Total reading time.

Likewise, Figure 7.2 shows results for neural nets during the first session (mean $R^2_{train} = .84$, $SD R^2_{train} = .03$; mean $R^2_{test} = .81$, $SD R^2_{test} = .11$) and the last session (mean $R^2_{train} = .87$, $SD R^2_{train} = .02$; mean $R^2_{test} = .80$, $SD R^2_{test} = .12$). For the training sets, model fits increased in the last session, while there was no such an increase for the test sets.

The *FI* analysis of neural nets approach shown in Figure 7.3 suggested that in the first session, nine predictors were important for total reading time (*logf* [.21], *wl* [.20], *cvq* [.18], *sonscore* [.15], *lineNo.* [.15], *odc* [.15], *val* [.14], *on* [.14], *aro* [.11]), while *wordNo.* (.08) and *hfn* (.07) were less important. For the last session, there were ten important predictors (*logf* [.23], *wl* [.22], *lineNo.* [.22], *val* [.21], *aro* [.19], *cvq* [.18], *odc* [.15], *sonscore* [.13], *wordNo.* [.11], *on* [.10]), and the less important one was *hfn* (.07). Again, all feature importance values for *val* and *aro* were above .01 and again we observed an increase in the importance values from first to last reading especially for *val*. And again, in last reading the

feature importance value for *val* was at the same level than the importance values for the most important surface features *logf* and *wl*.

Fixation probability.

Finally, Figure 7.2 also gives results for the two session using neural nets (first session: mean $R^2_{train} = .93$, $SD R^2_{train} = .02$; mean $R^2_{test} = .91$, $SD R^2_{test} = .07$; last session: mean $R^2_{train} = .92$, $SD R^2_{train} = .01$; mean $R^2_{test} = .91$, $SD R^2_{test} = .06$). For the training sets, the model fits decreased in the last session.

The *FI* analysis of the optimal neural nets approach in Figure 7.3 suggested that in the first session, all predictors were important for fixation probability (*wl* [.46], *on* [.27], *cvq* [.26], *lineNo.* [.21], *aro* [.18], *wordNo.* [.18], *logf* [.17], *sonscore* [.15], *hfn* [.15], *val* [.14], *odc* [.13]). For the last session, ten predictors were important (*wl* [.35], *val* [.24], *lineNo.* [.24], *cvq* [.20], *wordNo.* [.20], *sonscore* [.17], *on* [.16], *logf* [.15], *aro* [.12], *odc* [.12]), as the less important one was *hfn* (.09). Also, for fixation probability all *FI* values for *val* and *aro* were above .1. While the *FI* for *val* increased from first to last reading, the *FI* value for *aro* slightly decreased. In the last reading, *val* was one of the important predictors after *wl*.

7.5 Discussion

Reading enables ideas to be conveyed and emotions to be transferred. In the process of reading, a lot of psycholinguistic features come into play to guide the understanding of the meanings and feelings, leaving it is necessary to investigate the underlying mechanism. There are two routes of these investigations: one is the theoretical route in which researchers concentrate on qualitatively summarizing and classifying the natural texts, e.g., narratives or poetries (e.g., Jakobson & Jones, 1970; Simonto, 1989; Vendler, 1997); the other one is an empirical, in which researchers are good at manipulating several psycholinguistic features with experimental designed materials (e.g., Just & Carpenter, 1980; Inhoff & Rayner, 1986; Raney & Rayner, 1995; Pynte, New, & Kennedy, 2008) to examine their interplay or the time course of the effects (e.g., Lee, Rayner, & Pollatsek, 1999; Knickerbocker et al., 2019). However, seldom do this kind of empirical study use natural texts, such as poems. It seems that there is a gap between the two routes.

The working group of Jacobs has been working for quite a long time on creating a pathway between both routes of investigation (e.g., Willems, 2015; Willems & Jacobs, 2016; Jacobs & Willems, 2018). Recently, an approach combining QNA with machine-learning-

based predictive modeling has been introduced to the field of literary reading. Some studies have already proved the utility of this approach (e.g., Xue et al., 2019; Xue, Jacobs, & Lüdtke, 2020). With the predictive models and computational means now available, we can analyze human cognition, emotion and behavior, such as eye movements, in naturally rich settings (Lappi, 2015) such as literature (e.g., Willems, 2015; Willems & Jacobs, 2016; Jacobs & Willems, 2018). However, the two studies using this approach to analyze poetry reading only investigated the effects of the “cold” psycholinguistic features (e.g., *word length*, *word frequency*). Up to now, there is no empirical study has ever investigated both the multiple “cold” surface features and the “hot” affective-semantic features in one study.

For the present study, we used besides positional information, seven surface, and two affective-semantic psycholinguistic features at the lexical level to predict eye movements of readers while they (re-)reread Shakespeare’s sonnets. This study is a reanalysis of one study from our group (Xue, Jacobs, & Lüdtke, 2020). By applying the QNA combined with a machine learning-based predictive modeling approach, the importance of the psycholinguistic features was calculated for reading in a first and a last reading session.

Like in the original analysis (Xue, Jacobs, & Lüdtke, 2020), the results again highlighted the good reputation of the machine-learning-based predictive modeling approach. Although in the present study only content words were included, the neural nets still built satisfactory models for all eye tracking parameters. We again confirmed that surface features play important roles in poetry reading and rereading. The most important surface features were still *word length*, *word frequency*, *orthographic neighborhood*, *orthographic dissimilarity*, and *sonority* (Xue et al., 2019; Xue, Jacobs, & Lüdtke, 2020), even when affective-semantic features were put into the analyses. Moreover, for the prediction of first fixation duration, it seems that positional information (*lineNo.* and *wordNo.*) explained most of the variances, which again proved that first fixation duration was due more to fast and automatic reading behavior rather than to lexical parameters (Hyönä & Hujanen, 1997; Clifton, Staub, & Rayner, 2007).

Most importantly, the focus of the present study is on the affective-semantic features. For almost all eye tracking measures, the feature importance values for word-based valence and arousal were above .1 indicating the importance of the two affective-semantic features in reading poetry. For first fixation duration, especially valence was as important as other well-established surface features word length and word frequency, which is well in line with

results from EEG studies about single word comprehension indicating an influence of valence starting relatively early around 200 *ms* after word onset (e.g., Hofmann et al., 2009; Kissler, Assadollahi, & Herbert, 2006; Schacht & Sommer, 2009). For all other eye tracking measures, the observed features importance values for valence reached also at the same level as the feature importance values observed for the other most important, but in contrast to first fixation duration, this pattern could be observed only in last reading.

Taken together our results showed that affective-semantic features became more important in the last reading session, especially for eye tracking measures associated with higher level comprehension processes. The importance of *valence* and *arousal* in predicting the gaze duration, regression time, total reading time, and fixation probability was higher in the last session compared to the first session, except for a little decrease of the importance for *arousal* in predicting fixation probability. For gaze duration, *valence* turned out to be an important feature only in the last reading session. These results indicate that readers began to pay more attention to the affective-semantic aspects of the poem in rereading compared to the first reading. These findings are in line with the resource allocation theory (Millis & Simon, 1994; Millis, Simon, & TenBroek, 1998). It is assumed that surface features could be processed automatically and obligatorily, and the resources distributed to this kind of features were the same across reading sessions. Whereas, high-level features like affective-semantic features, might need extra resources and thus could be processed well in a later reading session. After a first reading, readers might have spare vigor to process the “hot” affective-semantic aspects of the poem. However, on the poem level, the paper-pencil task showed that the two poems (sonnet 27 and sonnet 66) chosen in the present study expressed relatively neutral emotions. There was no difference in the overall evaluation of supralexicvalence and arousal, which may be a reason for the relatively small changes in the importance of the two affective-semantic features. In the future, different kinds of poetic texts with stronger supralexicvalence emotions need to be included to further study the effect of lexical affective-semantic features.

In conclusion, this study reanalyzed the data of the study on poetry (re-)reading (Xue, Jacobs, & Lüdtkke, 2020) by combining the “cold” surface features and the “hot” affective-semantic features, to check their roles in determining the eye movements of readers. We found that: neural nets could build satisfactory models in studies containing both surface and affective-semantic features; surface features play fundamental roles in both the first reading and the rereading; in rereading affective-semantic features and especially word-based valence

are as important as other well-established surface features, and affective-semantic features become more important in rereading compared to the first reading, suggesting that readers started to pay more attention to the affective-semantic aspects when they had already known the poem.

7.6 Limitations and Outlook

The first limitation is that for the present study only two affective-semantic features were added to predict the eye movements of readers in poetry reading. However, there are also other affective-semantic features which may also play roles in reading. For instance, the *imageability* might also be important, especially for literary reading (cf. Magyari et al., 2020). *Imageability* refers to the extent to which a word evokes a tangible sensation (Westbury et al., 2013). Words with higher imageability have processing advantages over those with lower imageability in remembrance and naming (e.g., Hamilton & Rajaram, 2001; de Groot, 1989). But building mental images during reading might be also associated with longer fixation duration on words (Magyari et al., 2020)

The second limitation is that we only used two sonnets rated as relatively neutral. It could be assumed that the small emotion potential indicated by neutral supralexical valence ratings and small supralexical arousal ratings diminished the overall importance of the used lexical affective-semantic features and also possible importance differences between first and rereading. Nevertheless, we observed an increase in feature importance from first to last reading for both lexical valence and arousal. Our conclusion that affective-semantic features are more important in rereading compared to the first reading needs to be checked with other poetic texts with a wider range of supralexical valence and arousal, and also with other kinds of challenging literary texts with polysemic meaning like short stories or entire novels.

Author Contributions

All authors contributed to the design of the experiment. Xue S. carried out the experiment, analyzed the data, and wrote the first draft of the manuscript; Lüdtke J. modified the manuscript; Jacobs A. M. improved the manuscript. All authors have contributed to and approved the final manuscript.

III

General

Discussion

Chapter 8: General Discussion and Outlook

The goal of my dissertation was to investigate how psycholinguistic features influence the reading behavior during poem perception. Since previous research mostly focused on prose (e.g., Carrol et al., 2015), textoids or investigated reading on single word level only (e.g., Clifton, Staub, & Rayner, 2007; Radach & Kennedy, 2013; Rayner, 2009), different issues had to be solved. First, appropriate (natural text) stimulus material had to be found with the goal that it has a clear comparable structure. The choice fell on Shakespeare's sonnets due to its standardized structure. These sonnets were broken up into measurable and testable features by applying QNA (e.g., Jacobs et al., 2017; Jacobs, 2017, 2018a, b). The great advantage of QNA is that word and text properties can be defined, and thus literary text can be translated into statistical analysis right in the sense of Franzosi (2010) as "turning words into numbers". As such, the QNA bridges the gap between qualitative poetry research in the field of literary studies and quantitative approaches of neurocognitive and psycholinguistic research. Second, I faced the issue that previous research mainly used general linear models for analysis, although previously criticized as it disregards intercorrelation and non-linear relationships (e.g., Kliegl, Olson, & Davidson, 1982). Due to the high complexity of language, especially when it comes to poetry with its literary figures and similes, linear models are doomed to fail. Recent strategies to solve this issue were mainly based on interim solutions, such as introducing intercorrelation thresholds in these linear models (e.g., Balota & Chumbley, 1984). Thus, the aim of my dissertation also was to seek out statistical ways, to deal with the non-linear patterns of correlations affecting poetry reception (e.g., Willems, 2015; Willems & Jacobs, 2016; Jacobs & Willems, 2018). I applied machine-learning tools to disentangle and identify complex relationships in and between the data (e.g., Coit, Jackson, & Smith, 1998; Francis, 2001; Breiman, 2001; Tagliamonte & Baayen, 2012; Yarkoni & Westfall, 2017; LeCun, Bengio, & Hinton, 2015). These machine learning algorithms, however, have only seldom been applied to natural language stimuli (Jacobs et al., 2017; Jacobs & Kinder, 2017, 2018). Consequently, a further challenge of the dissertation project was to identify appropriate machine learning tools, which are applicable to eye tracking data.

To face these open questions, I conducted two eye tracking experiments using Shakespeare's sonnets (single reading study 1, rereading study 2 and 3). To get a deeper insight into the interplay of linguistic features and cognitive processing during poetry perception, I implemented machine learning algorithms to predict eye movement parameters with QNA features. Since over 100 features in Shakespeare's sonnets were reported by Jacobs et al. (2017), my dissertation project can be seen as a starting point in investigating the plethora of features and their role in literary perception. My focus was on the so-called surface features. In study 1 (Xue et al., 2019), we addressed the question of the general role of psycholinguistic surface features in Shakespeare's sonnets reading. In study 2 (Xue, Jacobs, & Lüdtke, 2020), we aimed to find out whether the roles of these psycholinguistic features may be changed in rereading. In study 3 (Xue, Jacobs, & Lüdtke, in preparation), we used the data of study 2, extracted the content words to add two affective-semantic features, *valence* and *arousal*, as these cannot be computed for grammatical words. Thus study 3 goes one step further by combining surface and affective-semantic features (for an overview, please see Table 8.1). In the following sections, I discuss the most important findings and implications of my empirical work in a broader context.

Table 8.1 Overview of all Three Studies Regarding Their Experimental Details and Main Outcomes

	Study 1	Study 2	Study 3
Material	3 Sonnets	2 Sonnets	Content words of 2 Sonnets
Paradigm	Reading	Rereading	Reading
Analysis	Reading	Reading + Rereading	Reading + Rereading
Features	Seven surface	Two positional Seven surface	Two positional Seven surface Two affective-semantic
Tools	General linear model (standard least squares) Non-linear model (bootstrap forest, neural nets)	General linear model (standard least squares) Non-linear model (neural nets)	Non-linear model (neural nets)
Important Lexical predictors	Total reading time: <i>wl</i> , <i>logf</i> , <i>on</i> , <i>sonscore</i> , <i>odc</i> Fixation probability: <i>wl</i> , <i>logf</i> , <i>on</i> , <i>sonscore</i> Simple linear regression results indicate that: Words with longer <i>wl</i> , lower <i>logf</i> , smaller <i>on</i> , higher <i>sonscore</i> had longer total reading time and a higher fixation probability Words with higher <i>odc</i> , attracted longer total reading time	First session: Gaze duration: <i>wl</i> , <i>logf</i> , <i>on</i> , <i>sonscore</i> , <i>odc</i> , <i>cvq</i> Regression time: <i>wl</i> , <i>logf</i> , <i>on</i> , <i>sonscore</i> , <i>cvq</i> Total reading time: <i>wl</i> , <i>logf</i> , <i>on</i> , <i>sonscore</i> , <i>odc</i> , <i>cvq</i> Fixation probability: <i>wl</i> , <i>logf</i> , <i>on</i> , <i>sonscore</i> , <i>cvq</i> Last session: Gaze duration: <i>wl</i> , <i>logf</i> , <i>on</i> , <i>sonscore</i> , <i>odc</i> Regression time: <i>wl</i> , <i>logf</i> , <i>on</i> , <i>sonscore</i> , <i>cvq</i> Total reading time: <i>wl</i> , <i>logf</i> , <i>on</i> , <i>sonscore</i> , <i>odc</i> Fixation probability: <i>wl</i> , <i>logf</i> , <i>on</i> , <i>sonscore</i> , <i>cvq</i>	First session: Gaze duration: <i>wl</i> , <i>logf</i> , <i>sonscore</i> , <i>odc</i> , <i>cvq</i> , <i>aro</i> Regression time: <i>wl</i> , <i>logf</i> , <i>on</i> , <i>sonscore</i> , <i>cvq</i> , <i>val</i> , <i>aro</i> Total reading time: <i>wl</i> , <i>logf</i> , <i>on</i> , <i>sonscore</i> , <i>odc</i> , <i>cvq</i> , <i>val</i> , <i>aro</i> Fixation probability: <i>wl</i> , <i>logf</i> , <i>on</i> , <i>sonscore</i> , <i>odc</i> , <i>cvq</i> , <i>hfn</i> , <i>val</i> , <i>aro</i> Last session: Gaze duration: <i>wl</i> , <i>logf</i> , <i>sonscore</i> , <i>odc</i> , <i>cvq</i> , <i>val</i> , <i>aro</i> Regression time: <i>wl</i> , <i>logf</i> , <i>on</i> , <i>sonscore</i> , <i>cvq</i> , <i>hfn</i> , <i>val</i> , <i>aro</i> Total reading time: <i>wl</i> , <i>logf</i> , <i>on</i> , <i>sonscore</i> , <i>odc</i> , <i>cvq</i> , <i>val</i> , <i>aro</i> Fixation probability: <i>wl</i> , <i>logf</i> , <i>on</i> , <i>sonscore</i> , <i>odc</i> , <i>cvq</i> , <i>val</i> , <i>aro</i>

New insights	<p><i>wl</i>, <i>logf</i>, <i>on</i>, and <i>sonscore</i> are basic features for both eye tracking parameters</p> <p>smaller <i>on</i> → increased total reading time and fixation probability</p> <p>higher <i>sonscore</i> → increased total reading time and fixation probability</p> <p>higher <i>odc</i> → longer total reading time</p>	<p><i>wl</i>, <i>logf</i>, <i>on</i>, and <i>sonscore</i> are basic features, unaffected by repetition</p>	<p>The importance of <i>val</i> and <i>aro</i> in predicting gaze duration, regression time, total reading time and fixation probability was higher in the last session compared to the first session, except for a little decrease of the importance for <i>aro</i> in predicting fixation probability</p>
Model winner	Neural nets	Neural nets	—

8.1 Methodological point of view

Machine learning in poetry research. I investigated whether machine learning tools could be successfully used in literary reading research and predict eye tracking parameters. The simple answer is: yes. All three studies, comprising two different data samples, showed the reliability of this approach in poetry reading. In study 1, both machine learning tools outperformed the general linear model with higher model fits (mean R^2) in the training and test sets. The poor performance of the general linear models suggests that there are relatively large low-order (e.g., two-way or higher ranked) interactions or other nonlinearities that the machine learning tools captured but the regression did not (cf. Breiman, 2001; Yarkoni & Westfall, 2017). This finding confirms the previously described issue of the insufficiency of general linear model-based analyses (e.g., van Halteren et al., 2005; Yarkoni & Westfall, 2017). However, there were fine performance differences among the two non-linear interactive models: neural nets got good cross-validated performance; although bootstrap forests produced higher mean R^2 in the training sets, they could not generalize well to the test sets, as indicated by large standard deviations. In study 2, I only compared neural nets to linear models as the neural nets outperformed bootstrap forests. This could be due to the different conceptualizations of the two machine learning tools examined. The bootstrap forests are based on a single distribution, whereas neural nets are based on interactive and multiple distributions. It seems that interactively calculating algorithms better reflect interactive psycholinguistic features. In study 2, I could replicate the predictive power of

neural nets in poetry as they produced acceptable models for all five eye tracking parameters (mean $R^2 > .30$). Likewise, they produced much higher model fits than standard least squares regression. In study 3, I again focused on neural nets. Indeed, the neural nets also showed good performance in predicting all five eye tracking parameters when focusing on the content words only. Neural nets might therefore be particularly suitable to examine the complex relationship of linguistic features and cognitive processing during poetry perception (as measured by eye tracking). Likewise, the neural net approach appears advantageous for studies on natural reading in which multiple psycholinguistic and context features may play a role (Jacobs, 2015a, 2018a). However, since those tools are continuously developed and modified, also newer techniques, such as various support vector space models (Jacobs, 2019; Jacobs & Kinder, 2019) need to be tested regarding their sufficient application to eye tracking data on poem reception.

To my knowledge, my dissertation project provides the first results on natural reading combining eye tracking on natural reading with predictive modeling. With my three studies, I can repeatedly show that neural nets are a highly appropriate tool to investigate the highly complex composition of the readers behavior (eye tracking) during poetry perception. I hope that my results might inspire future researchers to finally leave over-simplistic linear models in the past for the sake of predictive modeling approaches.

Pitfalls and Promises of eye tracking in poetry research. Eye tracking experiments investigate reading on the single word level or sentence level and recently also whole texts by using e.g., newspaper articles (Jarodzka & Brand-Gruwel, 2017). Although the patterns of eye movements during reading are purely behavioral data, they provide first insights into the cognitive processes during reading (e.g., Just & Carpenter, 1980; Hyönä & Hujanen, 1997; Rayner, 1998; Rayner et al., 2006; Clifton, Staub, & Rayner, 2007). As proved, first fixation duration and gaze duration reflect the first pass reading of the word, which are related to the fast and automatic initial word recognition processes; gaze duration, regression time, total reading time and fixation probability are measures reflecting revisits to a word, which are relevant to the delayed lexical access or the integrative processing of the word (Rayner, 1998). However, literary reception fundamentally differs from expository text reading. On the reader side, this holds true for the overall aim of the reading activity (leisure time versus information acquisition) that might influence our reading focus. From a thematic perspective, literary texts aim to meet figurative and aesthetic standards while expository texts need to be informative. Consequently, the linguistic composition differs substantially, and the

information eye tracking can give might differ accordingly. Here I applied eye tracking to poetry reading to investigate whether the well-established eye tracking parameters (e.g., first fixation duration) can also help to better understand how we read literary texts. In study 1, by using the neural nets approach the seven psycholinguistic features could satisfactorily predict the eye tracking measures related to later processing stages (i.e., total reading time, fixation probability), but not the first fixation duration. In study 2 and 3, by adding the positional information as predictors, first fixation duration could also be well predicted. Moreover, in study 2 there were rereading benefits for regression time, total reading time and fixation probability, but not for first fixation duration and gaze duration. All these findings prove that eye tracking measures related to early processes are less influenced by lexical parameters (cf. Hyönä & Hujanen, 1997; Clifton, Staub, & Rayner, 2007).

8.2 Theoretical point of view

Many studies showed the important role of surface features and affective-semantic features (e.g., Just & Carpenter, 1980; Inhoff & Rayner, 1986; Raney & Rayner, 1995; Pynte, New, & Kennedy, 2008; Scott, O'Donnell, & Sereno, 2012). However, nearly all these studies were confined to examine one or two features and consequently return a very simplified insight into reading. To gain a deeper and more comprehensive idea of literary reading, I considered several features that might influence the reading behavior of readers, measured by indirect online measurements (Dixon & Bortolussi, 2015). Given that I had to pare down the number of features to a computable and interpretable amount, I can only surmise the complex composition of linguistic features guiding the reading process and thus poetry perception.

In study 1, our findings confirm those of previous studies in that long and low-frequency words tend to be fixated more often and longer (e.g., Just & Carpenter, 1980; Inhoff & Rayner, 1986; Raney & Rayner, 1995; Pynte, New, & Kennedy, 2008). However, as an additional influence on total reading time and fixation probability, we found that words with a low *orthographic neighborhood density* attract longer fixations and higher fixation probability. Additionally, words which were orthographically dissimilar to other words in the corpus (i.e., more salient/ higher *orthographic dissimilarity*) also attracted longer total reading time. These findings support the facilitative effect hypothesis of Andrews (1989, 1992). He assumed that a larger neighborhood benefits lexical access, which means the partial activation of neighbors in some way speeds up access to the target representation.

However, inconsistent to one of the premises of the multiple read-out model (MROM, Jacobs et al., 1998)—namely, words with higher frequency neighbors will be processed more slowly than words without higher frequency neighbors, I did not detect the effect of higher frequency neighbors in study 1, which suggests the MROM may overestimate the role of inhibition in the orthographic processing of English words. Sears, Campbell, and Lupker (2006) also found that higher frequency neighbors have little effect on the identification of English words. He explained that most English words three to five letters in length do have higher frequency neighbors (Andrews, 1997; Siakaluk, Sears, & Lupker, 2002). Unlike other languages, this neighborhood structure for English words (i.e., larger neighborhoods and many higher frequency neighbors) may necessitate a lexical processor with weaker inhibitory connections.

Besides, I found that a higher *sonority* of a word increased both its total reading time and fixation probability, which suggests that readers tend to have a more intensive phonological recoding during poetry reading (e.g., Kraxenberger, 2017). The impact of the sonority score in literary reading was also investigated by Aryani et al. (2013; 2016, 2018; Aryani, Hsu, & Jacobs, 2018), who found that almost 15% of word rating variance can be derived from sonority score. Thus, our predictive modeling results can bridge the gap between computational models (calculating the QNA feature sonority score), subjective behavioral ratings (Aryani et al., 2013; 2016, 2018; Aryani, Hsu, & Jacobs, 2018) and cognitive reading processing measured by eye tracking. The results concerning the feature *higher frequent neighbors* are inconclusive across the three models which may be because 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.

In study 2, rereading improved the fluency of reading on poem level (shorter total reading times, shorter regression times, and lower fixation probability) and the depth of comprehension (e.g., Hakemulder, 2004; Kuijpers & Hakemulder, 2018). Contrary to the other rereading studies using literary texts (e.g., Dixon et al., 1993; Millis, 1995; Kuijpers & Hakemulder, 2018), no increase in appreciation was apparent. However, when checking the appreciation rating for each sonnet separately, the effect reappeared for sonnet 66 but not for sonnet 27, as readers liked sonnet 66 slightly more after the last session than after the first. Whether this difference is the result of a ceiling effect (sonnet 27 was already well appreciated after the first session) or the result of different levels of general

comprehensibility (sonnet 66 has longer and less frequent words than sonnet 27) needs to be tested in future with larger sample sizes and different literary material.

Most importantly, we successfully replicated the findings of Xue et al. (2019, study 1) about reading Shakespeare's sonnets and extend our knowledge to rereading behavior. Again, *word length*, *word frequency*, *orthographic neighborhood density*, and *sonority score* were most important in predicting total reading time and fixation probability. Also, *orthographic dissimilarity* proved to be important for total reading time. Since two out of the three sonnets of study 1 were retested with a new bigger sample of participants, we can validate the findings of study 1 and show that the same poems elicited similar eye movement behavior in new readers. Interestingly, the basic features, which showed higher feature importance during the first reading were also the most important ones during the last reading session. Surface features like *word length*, *word frequency*, *orthographic neighborhood density*, and *sonority* seem to be basic to eye movement behavior in the first reading and also the last reading. Thus, these surface features are quite stable regarding their influence on eye tracking parameters. However, in this rereading study also the *consonant vowel quotient* was indicated as a potentially important feature for total reading time (first reading session) and fixation probability (first and last reading session). The emergence of the *consonant vowel quotient* as an additional important feature influencing reading can be explained by the larger sample size. The bigger dataset might have improved the sensitivity of the predictive model. Apparently, two important phonological features, *sonority score* and *consonant vowel quotient*, have a decisive impact during poetry perception (Kraxenberger, 2017). This is in line with the assumption that consonant status and sonority also play a role in silent reading (Maionchi-Pino et al., 2008; Berent, 2013).

In study 3, we reanalyzed the data of study 2 with the difference, that we focused on content words (121 words). Using QNA, two lexical affective-semantic features, *valence* and *arousal*, were additionally calculated and added to predict the eye movements in (re-)reading. Using neural nets, I replicated my former results (in the present study with fewer words and thus, fewer time points) that the surface features, *word length*, *word frequency*, *orthographic neighborhood*, *orthographic dissimilarity*, and *sonority* stand out when predicting aggregated measures of first fixation duration, gaze duration, total reading time and fixation probability. However, the analysis showed that word-based *valence* and *arousal* also played an important role in predicting the eye tracking data. For both features, I observed an increase in the importance from first to last reading. Especially for *valence*, the values of the feature

importance observed in the last reading were as high as the values for the most important surface features. Based on the Resource Allocation Theory (Millis & Simon, 1994), I assume that surface features lay the foundation for poetry reception, and once a first understanding on the (sub)lexical level is established, readers have more free-up resources to pay more attention to the affective-semantic aspects (i.e., in the last reading session). A different interpretation would be according to the building of a poetry specific situation model by the reader (e.g., Schmidt, 1989; Fechino, Jacobs, & Lüdtkke, 2020). Schmidt (1982, 1989) defined two most important conventions for literary reading, the *aesthetic convention* and the *polyvalence convention*, which rely highly on the processing of affective-semantic features. Since the feature importance of surface variables is stable in both first and last reading, the reader may ‘use’ them as landmarks while perusing the sonnet. These are still needed within the last reading even if then affective-semantic features are more considered. It would mean surface variables form the basic structure of the readers’ situation model, which is then added further details such as affective-semantic information and so on. This hypothesis needs further investigation by further poems and features. However, it could give further insight into reading models particularly poetry reading and possible relation to the contribution of other features, such as the beauty of words (Jacobs, 2017).

In summary, due to replication, my three studies show highly reliable data for high feature importance of surface variables in reading Shakespeare’s sonnets, and in rereading an increasing impact of affective-semantic features. From a methodological viewpoint, all three studies show a much better sufficiency of neural net approach than the classical general linear model approach in psycholinguistic eye tracking research.

With my dissertation project, I can show that the application of predictive modeling to investigate poetry might be far more suitable to capture the highly interactive, non-linear composition of linguistic features in natural texts that guide reading behavior and reception. My results seem to be stable and valid as I could replicate these novel findings using machine learning algorithms within my dissertation project. Besides, the choice of a larger set of linguistic features allows disentangling their contribution during different stages of the reading process. Surface features seem to influence reading during all different stages, while affective-semantic features seem to increase their importance in line with processing depth as indicated by higher influence during rereading. In sum, my dissertation project is a first step towards a more differentiated picture of the guiding factors of poetry reception and a poetry specific reading model.

8.3 Limitation and outlook

Besides highly interesting results of all three studies on sonnet reading, regarding the successful application of neural nets to eye tracking data and influences of seven surface and two affective-semantic features in the literary reading process, there also some limitations. In fact, there are over 100 features computed for the corpus of Shakespeare's sonnets (Jacobs et al., 2017). According to the multilevel hypothesis of the NCPM (e.g., Hsu et al., 2015a; Jacobs et al., 2016), many fore- and backgrounding features additionally contribute to the highly complex literary reading process. The 4×4 feature matrix has been developed for the application of the Neurocognitive Poetics Model (Jacobs, 2011, 2015a, b; Nicklas & Jacobs, 2017; Willems & Jacobs, 2016) in systematic (neuro)cognitive experiments. The matrix is based on the hypothesis, that processing of poetry is influenced by a whole set of sublexical, lexical, interlexical, and supralelexical features at the metric, phonological, morpho-syntactic and semantic levels, which has been proved by several studies from single words to proverbs and whole poems (Aryani et al., 2016; Jacobs & Lüdtke, 2017; Jacobs et al., 2015, 2016; Jacobs, Hofmann, & Kinder, 2016; Ullrich et al., 2017). To gain further insight into the interplay of this high number of features, we may start by adding more affective-semantic features to the model, and then sort to identify, define and classify interlexical and supralelexical features more reliably, so as to efficiently include them in empirical eye tracking studies. The present dissertation focusses on sublexical and lexical features. However, also on these two levels there are still uninvestigated features, which also have to be taken into account.

How superficial are surface variables?

The term 'surface variables', which I chose for my studies may seem confusing. Of course, there are many more features affecting the reader's behavior at first glance. For example, Niikuni, Iwasaki, and Muramoto (2015) manipulated Japanese interpunctuation, through which a Japanese written text seems more difficult (by inserting more commas). Immediately the participants' eye movement behavior change compared to the same text with fewer punctuations. In the same vein, Fechino, Jacobs, and Lüdtke (2020) showed that the reading behavior of a poem significantly changes, when a poem is presented as a prose text. Thus, the readers' expectancy about reading literary influences their eye movements. When I talk about surface, I would like to ensure that the readers of the dissertation recognize that I am aware of the fact that many more aspects and features could count under the term 'surface

feature'. Since the whole methodology, using QNA and machine learning tools on eye tracking data was never done before this dissertation, this term was used to give the article's readers a quick entrance to this approach. Also, it is a first step in systematizing the high number of features in more psychological terms, beside the 4x4 matrix, which is more systematized from linguistic perspective.

Affective-semantic features.

The first limitation is that only two affective-semantic features were added to predict the eye movements of readers in poetry reading. However, there are also other lexical- or affective-semantic features which may also play roles in reading. As a starting point, it may be also necessary to take additional affective-semantic features into account. Prominent examples are the *imageability* and the *aesthetic potential*, which are highly interesting regarding poem reception.

Imageability refers to the extent to which a word evokes a tangible sensation (Westbury et al., 2013) and words with higher imageability have processing advantages over those with lower imageability in remembrance and naming (e.g., Hamilton & Rajaram, 2001; de Groot, 1989). *Imageability* is a highly interesting candidate for further eye tracking analyses since poems often include many semantic features, especially metaphors. These again comprise a density of various lexical and sublexical features (Jacobs & Kinder, 2018), particularly features also investigated in the present dissertation. However, the relation of *imageability* to the surface variable is quite interesting, because of its close relation regarding word frequency, neighborhood density, and consonant vowel quotient (Westbury et al., 2013). Thus, such research would give a further insight into the interplay between surface and semantic features.

Aesthetic potential is a novel feature firstly introduced by Jacobs (2017). He found that this feature could be used as an important predictor of the word beauty, although the effect of this feature is comparatively small. Furthermore, Jacobs and Kinder (2019) provided detailed material regarding the SentiArt tool and the calculation of aesthetic potential. The feature is correlated with subjective liking ratings, so direct and concrete hypotheses regarding eye tracking parameter distributions can be made: Since the *aesthetic potential* is similar to valence and is deviated from semantic similarity, the current data (study 3) should also provide evidence for higher importance for aesthetic potential while rereading compared

to first pass reading. Thus, participants' liking ratings should also increase from first to last reading.

Like *valence* and *arousal*, based on the resource allocation theory (Millis & Simon, 1994; Millis, Simon, & TenBroek, 1998), we also assume that after a prior reading, readers might have more free-up resources which could be redistributed to high-level processes. Therefore, readers might be more affectively involved, and the importance of these features might be increased after the first reading session. However, if these hypothetically free-up resources are also used for further still not investigated features, is a question for future research.

Beyond lexical level.

As shown by the results of the rereading studies, on word level already about 60% variances of the five eye tracking parameters were explained by surface features. However, that does not mean that surface features are more important than affective-semantic features. It means more, that affective-semantic features may also influence reading beyond the lexical level, such as sonority score (Aryani et al., 2013; 2016, 2018; Aryani, Hsu, & Jacobs, 2018). For instance, using data from a recent study by Lehne et al. (2015), Jacobs (2015c) showed that arousal-span can account for about 25% of the variance in suspense ratings from readers of E.T.A. Hoffmann's black romantic story.

However, extending the present research to other inter/supra-lexical features requires extending sample sizes (i.e., more/longer texts and more participants). In the future, we need to check the validity of our findings with larger samples and the generalizability to other literary works.

References

- Ahmed, Y., Francis, D. J., York, M., Fletcher, J. M., Barnes, M., & Kulesz, P. (2016). Validation of the direct and inferential mediation (DIME) model of reading comprehension in grades 7 through 12. *Contemporary Educational Psychology, 44–45*, 68–82. <https://doi.org/10.1016/j.cedpsych.2016.02.002>
- Amlund, J. T., Kardash, A. C. M., & Kulhavy, R. W. (1986). *Repetitive Reading and Recall of Expository Text. Source: Reading Research Quarterly* (Vol. 21). Winter.
- Andrews, M., Frank, S., & Vigliocco, G. (2014). Reconciling Embodied and Distributional Accounts of Meaning in Language. *Topics in Cognitive Science, 6*(3), 359–370. <https://doi.org/10.1111/tops.12096>
- Andrews, S. (1989). Frequency and Neighborhood Effects on Lexical Access: Activation or Search? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(5), 802–814. <https://doi.org/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), 234–254. <https://doi.org/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. <https://doi.org/10.3758/BF03214334>
- Aryani, A., Conrad, M., Schmidtke, D., & Jacobs, A. (2018). Why 'piss' is ruder than 'pee'? The role of sound in affective meaning making. *PLoS ONE, 13*(6), 1–27. <https://doi.org/10.1371/journal.pone.0198430>
- Aryani, A., Hsu, C. T., & Jacobs, A. M. (2018). The sound of words evokes affective brain responses. *Brain Sciences, 8*(6), 94. <https://doi.org/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*(OCT), 1–15. <https://doi.org/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), 191–204. <https://doi.org/10.1037/aca0000033>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Bailey, H., & Zacks, J. M. (2011). Literature and event understanding. *Scientific Study of Literature*, 1, 1. <https://doi.org/10.1075/ssol.1.1.07bai>
- Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, 10(3), 340–357. <https://doi.org/10.1037/0096-1523.10.3.340>
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577–600.
- Barton, J. J. S., Hanif, H. M., Eklinder Björnström, L., & Hills, C. (2014). The word-length effect in reading: A review. *Cognitive Neuropsychology*, 31(5–6), 378–412. <https://doi.org/10.1080/02643294.2014.895314>
- Berent, I. (2013). The phonological mind. *Trends in Cognitive Sciences*, 17(7), 319–327. <https://doi.org/10.1016/J.TICS.2013.05.004>
- Berkenkotter, C., & Huckin, T. N. (1993). Rethinking Genre from a Sociocognitive Perspective. *Written Communication*, 10(4), 475–509. <https://doi.org/10.1177/0741088393010004001>
- 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. <https://doi.org/10.1002/hbm.20643>
- Breiman, L. E. O. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breland, H. M. (1996). Word Frequency and Word Difficulty: A Comparison of Counts in Four Corpora. *Psychological Science*, 7(2), 96–99. <https://doi.org/10.1111/j.1467-9280.1996.tb00336.x>
- Brewer, W. F., & Lichtenstein, E. H. (1982). Stories are to entertain: A structural-affect

- theory of stories. *Journal of Pragmatics*, 6(5–6), 473–486. [https://doi.org/10.1016/0378-2166\(82\)90021-2](https://doi.org/10.1016/0378-2166(82)90021-2)
- Briesemeister, B. B., Kuchinke, L., & Jacobs, A. M. (2011). Discrete emotion effects on lexical decision response times. *PLoS ONE*, 6(8). <https://doi.org/10.1371/journal.pone.0023743>
- Britton, B. K., Graesser, A. C., Glynn, S. M., Hamilton, T., & Penland, M. (1983). Use of cognitive capacity in reading: Effects of some content features of text. *Discourse Processes*, 6(1), 39–57. <https://doi.org/10.1080/01638538309544553>
- Brown, S. (2002). Is Rereading more Effective than Prereading? *RELC Journal*, 33(1), 91–100. <https://doi.org/10.1177/003368820203300105>
- Brysbaert, M., Mander, P., & Keuleers, E. (2018). The Word Frequency Effect in Word Processing: An Updated Review. *Current Directions in Psychological Science*, 27(1), 45–50. <https://doi.org/10.1177/0963721417727521>
- Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research*, 24(3), 355–363. <https://doi.org/10.1016/j.cogbrainres.2005.02.020>
- Carminati, M. N., Stabler, J., Roberts, A. M., & Fischer, M. H. (2006). Readers' responses to sub-genre and rhyme scheme in poetry. *Poetics*, 34(3), 204–218. <https://doi.org/10.1016/J.POETIC.2006.05.001>
- Carrol, G., & Conklin, K. (2014). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism*, 17(4), 784–797. <https://doi.org/10.1017/S1366728913000795>
- Carrol, G., Conklin, K., Guy, J., & Scott, R. (2015). Processing punctuation and word changes in different editions of prose fiction. *Scientific Study of Literature*, 5(2), 200–228. <https://doi.org/10.1075/ssol.5.2.05con>
- Chamberland, C., Saint-Aubin, J., & Légère, M.-A. (2013). The impact of text repetition on content and function words during reading: Further evidence from eye movements. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 67(2), 94–99. <https://doi.org/10.1037/a0028288>
- Cichy, R. M., & Kaiser, D. (2019). Deep Neural Networks as Scientific Models. *Trends in*

- Cognitive Sciences*, 1–13. <https://doi.org/10.1016/j.tics.2019.01.009>
- Clements, G. N. (1990). *The role of the sonority cycle in core syllabification*. *Papers in Laboratory Phonology*. <https://doi.org/10.1017/cbo9780511627736.017>
- Clifton, C., Staub, A., & Rayner, K. (2007). *Eye movements in reading words and sentences*. *Eye Movements: A Window on Mind and Brain*. Amsterdam, Netherlands: Elsevier. <https://doi.org/10.1016/B978-008044980-7/50017-3>
- Coit, D. W., Jackson, B. T., & Smith, A. E. (1998). Static neural network process models: considerations and case studies. *International Journal of Production Research*, 36(11), 2953–2967. <https://doi.org/10.1080/002075498192229>
- Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D. (1977). Access to the internal lexicon. *Attention and Performance VI*. <https://doi.org/10.1006/brln.2001.2475>
- de Groot, A. M. B. (1989). Representational Aspects of Word Imageability and Word Frequency as Assessed Through Word Association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5), 824–845. <https://doi.org/10.1037/0278-7393.15.5.824>
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391–407. [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6<391::AID-ASII>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASII>3.0.CO;2-9)
- Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*. Cell Press. <https://doi.org/10.1016/j.neuron.2007.10.004>
- Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: Behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience*. Nature Publishing Group. <https://doi.org/10.1038/nrn3924>
- Dixon, P., & Bortolussi, M. (2015). Measuring Literary Experience: Comment on Jacobs. *Scientific Study of Literature*, 5(2), 178–182. <https://doi.org/10.1075/ssol.5.2.03dix>
- Dixon, P., Bortolussi, M., Twilley, L. C., & Leung, A. (1993). Literary processing and interpretation: Towards empirical foundations. *Poetics*, 22(1–2), 5–33. [https://doi.org/10.1016/0304-422X\(93\)90018-C](https://doi.org/10.1016/0304-422X(93)90018-C)
- Durgunoğlu, A. Y., Mir, M., & Ariño-Martí, S. (1993). Effects of repeated readings on

- bilingual and monolingual memory for text. *Contemporary Educational Psychology*, 18(3), 294–317. <https://doi.org/10.1006/ceps.1993.1022>
- Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, 6, 621. <https://doi.org/10.3389/fpsyg.2015.00621>
- Ekman, P. (2005). Basic Emotions. In *Handbook of Cognition and Emotion* (pp. 45–60). Chichester, UK: John Wiley & Sons, Ltd. <https://doi.org/10.1002/0470013494.ch3>
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A Dynamical Model of Saccade Generation During Reading. *Psychological Review*, 112(4), 777–813. <https://doi.org/10.1037/0033-295X.112.4.777>
- Fairley, I. R. (1986). The Reader's Need For Conventions: When is a Mushroom Not a Mushroom? *Style*. Penn State University Press. <https://doi.org/10.2307/42945578>
- Fechino, M., Jacobs, A. M., & Lüdtke, J. (2020). Following in Jakobson and Lévi-Strauss' footsteps: A neurocognitive poetics investigation of eye movements during the reading of Baudelaire's "Les Chats." *Journal of Eye Movement Research*, 13(3), 4. <https://doi.org/10.16910/jemr.13.3.4>
- Francis, L. (2001). Neural Networks Demystified. In *Casualty Actuarial Society Forum* (pp. 253–320).
- Franzosi, R. (2010). *Quantitative Narrative Analysis*. Sage.
- Freedman, A. (2003). *Genre In The New Rhetoric*. *Genre In The New Rhetoric*. Routledge. <https://doi.org/10.4324/9780203393277>
- Gambino, R., & Pulvirenti, G. (2018). *Storie menti mondi. Approccio neuroermeneutico alla letteratura*. Milano: Mimesis Edizioni.
- Gao, C., & Guo, C. (2018). The Experience of Beauty of Chinese Poetry and Its Neural Substrates. *Frontiers in Psychology*, 9, 1540. <https://doi.org/10.3389/fpsyg.2018.01540>
- Gernsbacher, M. A., & Kaschak, M. (2013). Text comprehension. In D. Reisberg (Ed.), *The Oxford Handbook of Cognitive Psychology* (pp. 462–474). Oxford University Press.
- Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., & Buccino, G. (2008). Processing abstract language modulates motor system activity. *Quarterly Journal of Experimental Psychology*, 61(6), 905–919. <https://doi.org/10.1080/17470210701625550>

- Gough, P. B., & Tunmer, W. E. (1986). Decoding, Reading, and Reading Disability. *Remedial and Special Education, 7*(1), 6–10.
<https://doi.org/10.1177/074193258600700104>
- Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). DISCOURSE COMPREHENSION. *Annual Review of Psychology, 48*(1), 163–189.
<https://doi.org/10.1146/annurev.psych.48.1.163>
- 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. <https://doi.org/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), 518–565.
<https://doi.org/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 & Psychophysics, 45*(3), 189–195. <https://doi.org/10.3758/BF03210696>
- Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review, 114*(2), 211–244. <https://doi.org/10.1037/0033-295X.114.2.211>
- Hakemulder, J. F. (2004). Foregrounding and Its Effect on Readers’ Perception. *Discourse Processes, 38*(2), 193–218. https://doi.org/10.1207/s15326950dp3802_3
- Hamilton, M., & Rajaram, S. (2001). The Concreteness Effect in Implicit and Explicit Memory Tests. *Journal of Memory and Language, 44*(1), 96–117.
<https://doi.org/10.1006/jmla.2000.2749>
- Hanauer, D. (1995). Literary and poetic text categorization judgementsin. *Journal of Literary Semantics, 24*(3), 187–210.
- Hanauer, D. (1996). Academic Literary Competence Testing. *Journal of Literary Semantics, 25*(2), 142–153. <https://doi.org/10.1515/jlse.1996.25.2.142>
- Hanauer, D. (1998a). Reading Poetry: An Empirical Investigation of Formalist, Stylistic, and Conventionalist Claims. *Poetics Today, 19*(4), 565. <https://doi.org/10.2307/1773260>
- Hanauer, D. (1998b). The genre-specific hypothesis of reading: Reading poetry and encyclopedic items. *Poetics, 26*(2), 63–80. <https://doi.org/10.1016/S0304->

422X(98)00011-4

- Hastie, T., Tibshirani, R., & Friedman, J. (2009). Unsupervised Learning. In *The Elements of Statistical Learning* (pp. 485–585). Springer, New York, NY.
https://doi.org/10.1007/978-0-387-84858-7_14
- Hickok, G., & Poeppel, D. (2007, May 13). The cortical organization of speech processing. *Nature Reviews Neuroscience*. Nature Publishing Group.
<https://doi.org/10.1038/nrn2113>
- Hoffstaedter, P. (1987). Poetic text processing and its empirical investigation. *Poetics*, *16*(1), 75–91. [https://doi.org/10.1016/0304-422X\(87\)90037-4](https://doi.org/10.1016/0304-422X(87)90037-4)
- Hofmann, M. J., & Jacobs, A. M. (2014). Interactive activation and competition models and semantic context: From behavioral to brain data. *Neuroscience & Biobehavioral Reviews*, *46*, 85–104. <https://doi.org/10.1016/J.NEUBIOREV.2014.06.011>
- Hofmann, M. J., Kuchinke, L., Tamm, S., Võ, M. L. H., & Jacobs, A. M. (2009). Affective processing within 1/10th of a second: High arousal is necessary for early facilitative processing of negative but not positive words. *Cognitive, Affective and Behavioral Neuroscience*, *9*(4), 389–397. <https://doi.org/10.3758/9.4.389>
- Hoorn, J. (1996). Psychophysiology and literary processing: ERPs to semantic and phonological deviations in reading small verses. *Empirical Approaches to Literature and Aesthetics*, *52*, 339–358.
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127–160. <https://doi.org/10.1007/BF00401799>
- Hsu, C.-T., Conrad, M., & Jacobs, A. M. (2014). Fiction feelings in Harry Potter. *NeuroReport*, *25*(17), 1356–1361. <https://doi.org/10.1097/WNR.0000000000000272>
- Hsu, C.-T., Jacobs, A. M., Altmann, U., & Conrad, M. (2015). The Magical Activation of Left Amygdala when Reading Harry Potter: An fMRI Study on How Descriptions of Supra-Natural Events Entertain and Enchant. *PloS One*, *10*(2), e0118179.
<https://doi.org/10.1371/journal.pone.0118179>
- Hsu, C.-T., Jacobs, A. M., Citron, F. M. M., & Conrad, M. (2015). The emotion potential of words and passages in reading Harry Potter - An fMRI study. *Brain and Language*, *142*, 96–114. <https://doi.org/10.1016/j.bandl.2015.01.011>

- Hsu, C.-T., Jacobs, A. M., & Conrad, M. (2015). Can Harry Potter still put a spell on us in a second language? An fMRI study on reading emotion-laden literature in late bilinguals. *Cortex*, 63, 282–295. <https://doi.org/10.1016/j.cortex.2014.09.002>
- 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. <https://doi.org/10.1080/713755738>
- Hyönä, J., & Niemi, P. (1990). Eye movements during repeated reading of a text. *Acta Psychologica*, 73(3), 259–280. [https://doi.org/10.1016/0001-6918\(90\)90026-C](https://doi.org/10.1016/0001-6918(90)90026-C)
- Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, 40(6), 431–439. <https://doi.org/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 R. Schrott & A. M. Jacobs (Eds.), *Gehirn und Gedicht: Wie wir unsere Wirklichkeiten konstruieren [Brain and Poetry: How We Construct Our Realities]* (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 in Human Neuroscience*, 9, 186. <https://doi.org/10.3389/fnhum.2015.00186>
- Jacobs, A. M. (2015b). The scientific study of literary experience: sampling the state of the art. *Scientific Study of Literature*, 5(2), 139–170. <https://doi.org/10.1075/ssol.5.2.01jac>
- Jacobs, A. M. (2015c). 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.
- Jacobs, A. M. (2017). Quantifying the Beauty of Words: A Neurocognitive Poetics Perspective. *Frontiers in Human Neuroscience*, 11, 622. <https://doi.org/10.3389/fnhum.2017.00622>
- Jacobs, A. M. (2018a). (Neuro-)Cognitive poetics and computational stylistics. *Scientific Study of Literature*, 8(1), 165–208. <https://doi.org/10.1075/ssol.18002.jac>
- Jacobs, A. M. (2018b). The Gutenberg English Poetry Corpus: Exemplary Quantitative Narrative Analyses. *Frontiers in Digital Humanities*, 5, 5.

<https://doi.org/10.3389/fdigh.2018.00005>

- Jacobs, A. M. (2019). Sentiment Analysis for Words and Fiction Characters From The Perspective of Computational (Neuro-)Poetics. *Frontiers in Robotics and AI*, 6(53). <https://doi.org/10.3389/FROBT.2019.00053>
- Jacobs, A. M., Graf, R., & Kinder, A. (2003). Receiver Operating Characteristics in the Lexical Decision Task: Evidence for a Simple Signal-Detection Process Simulated by the Multiple Read-Out Model. *Journal of Experimental Psychology: Learning Memory and Cognition*, 29(3), 481–488. <https://doi.org/10.1037/0278-7393.29.3.481>
- Jacobs, A. M., Hofmann, M. J., & Kinder, A. (2016). On elementary affective decisions: to like or not to like, that is the question. *Frontiers in Psychology*, 7, 1836. <https://doi.org/10.3389/fpsyg.2016.01836>
- 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(3), 139–160. <https://doi.org/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. <https://doi.org/10.1080/10926488.2018.1434943>
- Jacobs, A. M., & Kinder, A. (2019). Computing the Affective-Aesthetic Potential of Literary Texts. *AI*, 1(1), 11–27. <https://doi.org/10.3390/ai1010002>
- Jacobs, A. M., & Lüdtke, J. (2017). Immersion into narrative and poetic worlds. *Narrative Absorption*, 27, 69–96. <https://doi.org/10.1075/lal.27.05jac>
- Jacobs, A. M., Lüdtke, J., Aryani, A., Meyer-Sickendieck, B., & Conrad, M. (2016). Mood-empathic and aesthetic responses in poetry reception. *Scientific Study of Literature*, 6(1), 87–130. <https://doi.org/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 *Localist connectionist approaches to human cognition*. (pp. 147–188). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- 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(1), 4–51.
<https://doi.org/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. (2015). 10 years of BAWLing into affective and aesthetic processes in reading: what are the echoes? *Frontiers in Psychology*, 6, 714.
<https://doi.org/10.3389/fpsyg.2015.00714>
- Jacobs, A. M., & Willems, R. M. (2018). The fictive brain: Neurocognitive correlates of engagement in literature. *Review of General Psychology*, 22(2), 147–160.
<https://doi.org/10.1037/gpr0000106>
- Jacobsen, T. (2006). Bridging the arts and sciences: A framework for the psychology of aesthetics. *Leonardo*, 39(2), 155–162. <https://doi.org/10.1162/leon.2006.39.2.155>
- Jakobson, R. (1960a). Closing statement: Linguistics and poetics. In T. A. Sebeok (Ed.), *Style in language* (pp. 350–377). Cambridge: MIT Press.
- Jakobson, R. (1960b). Linguistics and poetics. In T. Sebeok (Ed.), *Style in language* (pp. 350–377). John Wiley.
- Jakobson, Roman, & Jones, L. G. (1970). *Shakespeare's verbal art in Th'expeñce of spirit*. Walter de Gruyter.
- Jakobson, Roman, & Lévi-Strauss, C. (1962). "Les Chats" de Charles Baudelaire. *L'Homme*, 2(1), 5–21. <https://doi.org/10.3406/hom.1962.366446>
- Jarodzka, H., & Brand-Gruwel, S. (2017). Tracking the reading eye: towards a model of real-world reading. *Journal of Computer Assisted Learning*, 33(3), 193–201.
<https://doi.org/10.1111/jcal.12189>
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness* (6th ed.). Harvard University Press.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329–354. <https://doi.org/10.1037/0033-295X.87.4.329>
- Kaakinen, J. K., & Hyönä, J. (2007). Perspective effects in repeated reading: An eye movement study. *Memory and Cognition*, 35(6), 1323–1336.
<https://doi.org/10.3758/BF03193604>

- Kamienkowski, J. E., Carbajal, M. J., Bianchi, B., Sigman, M., & Shalom, D. E. (2018). Cumulative Repetition Effects Across Multiple Readings of a Word: Evidence From Eye Movements. *Discourse Processes*, 55(3), 256–271.
<https://doi.org/10.1080/0163853X.2016.1234872>
- Kendeou, P., McMaster, K. L., & Christ, T. J. (2016). Reading comprehension: Core components and processes. *Policy Insights from the Behavioral and Brain Sciences*, 3(1), 62–69. <https://doi.org/10.1177/2372732215624707>
- Kidd, D. C., & Castano, E. (2013). Reading Literary Fiction Improves Theory of Mind. *Science*, 342(6156), 377 LP – 380. <https://doi.org/10.1126/science.1239918>
- Kinoshita, S. (2006). Additive and interactive effects of word frequency and masked repetition in the lexical decision task. *Psychonomic Bulletin and Review*, 13(4), 668–673. <https://doi.org/10.3758/BF03193979>
- Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 209–226). Oxford, UK: Blackwell.
- Kintsch, Walter. (1988). The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model. *Psychological Review*, 95(2), 163–182.
<https://doi.org/10.1037/0033-295X.95.2.163>
- Kintsch, Walter. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Kintsch, Walter, & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363–394. <https://doi.org/10.1037/0033-295X.85.5.363>
- Kissler, J., Assadollahi, R., & Herbert, C. (2006). Emotional and semantic networks in visual word processing: insights from ERP studies. In *Progress in Brain Research* (Vol. 156, pp. 147–183). Elsevier. [https://doi.org/10.1016/S0079-6123\(06\)56008-X](https://doi.org/10.1016/S0079-6123(06)56008-X)
- Kissler, J., Herbert, C., Peyk, P., & Junghofer, M. (2007). Buzzwords: Early Cortical Responses to Emotional Words During Reading. *Psychological Science*, 18(6), 475–480. <https://doi.org/10.1111/j.1467-9280.2007.01924.x>
- 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 & Cognition*, 10(3), 287–296. <https://doi.org/10.3758/BF03197640>

- Klin, C. M., Ralano, A. S., & Weingartner, K. M. (2007). Repeating phrases across unrelated narratives: Evidence of text repetition effects. *Memory & Cognition*, *35*(7), 1588–1599. <https://doi.org/10.3758/BF03193493>
- Klitz, T. S., Legge, G. E., & Tjan, B. S. (2000). *Saccade Planning in Reading With Central Scotomas: Comparison of Human and Ideal Performance. Reading as a Perceptual Process*. Amsterdam, Netherlands: North-Holland/Elsevier Science Publishers. <https://doi.org/10.1016/B978-008043642-5/50031-0>
- Knickerbocker, F., Johnson, R. L., Starr, E. L., Hall, A. M., Preti, D. M., Slate, S. R., & Altarriba, J. (2019). The time course of processing emotion-laden words during sentence reading: Evidence from eye movements. *Acta Psychologica*, *192*, 1–10. <https://doi.org/10.1016/j.actpsy.2018.10.008>
- Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In *Appears in the International Joint Conference on Artificial Intelligence (IJCAI)* (pp. 1137–1145).
- Koopman, E. M. (Emy). (2016). Effects of “literariness” on emotions and on empathy and reflection after reading. *Psychology of Aesthetics, Creativity, and the Arts*, *10*(1), 82–98. <https://doi.org/10.1037/aca0000041>
- Koops van 't Jagt, R., Hoeks, J. C. J., Dorleijn, G. J., & Hendriks, P. (2014). Look before you leap: How enjambment affects the processing of poetry. *Scientific Study of Literature*, *4*(1), 3–24. <https://doi.org/10.1075/ssol.4.1.01jag>
- Kraxenberger, M. (2017). *On Sound-Emotion Associations in Poetry*. Freie Universität Berlin.
- Kraxenberger, M., & Menninghaus, W. (2017). Affinity for poetry and aesthetic appreciation of joyful and sad poems. *Frontiers in Psychology*, *7*, 1–6. <https://doi.org/10.3389/fpsyg.2016.02051>
- Kuijpers, M. M., & Hakemulder, F. (2018). Understanding and Appreciating Literary Texts Through Rereading. *Discourse Processes*, *55*(7), 619–641. <https://doi.org/10.1080/0163853X.2017.1390352>
- Kuiken, D., Campbell, P., & Sopčák, P. (2012). The Experiencing Questionnaire: Locating exceptional reading moments. *Scientific Study of Literature*, *2*(2), 243–272. <https://doi.org/10.1075/ssol.2.2.04kui>
- Kuperman, V., Dambacher, M., Nuthmann, A., & Kliegl, R. (2010). The effect of word

- position on eye-movements in sentence and paragraph reading. *Quarterly Journal of Experimental Psychology*, 63(9), 1838–1857.
<https://doi.org/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. *Quarterly Journal of Experimental Psychology*, 66(3), 563–580. <https://doi.org/10.1080/17470218.2012.658820>
- Ladefoged, P. (1993). *A Course in Phonetics*. Fort Worth: Harcourt: Brace Jovanovich College Publishers. <https://doi.org/10.2307/412918>
- Landauer, T. K., & Dumais, S. T. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, 104(2), 211–240. <https://doi.org/10.1037/0033-295X.104.2.211>
- Lappi, O. (2015). Eye Tracking in the Wild: the Good, the Bad and the Ugly. *Journal of Eye Movement Research*, 8(5). <https://doi.org/10.16910/jemr.8.5.1>
- Lausberg, H. (1960). *Handbuch der literarischen Rhetorik : Eine Grundlegung der Literaturwissenschaft*. 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. <https://doi.org/10.1006/IJHC.1996.0039>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/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. <https://doi.org/10.1006/JMLA.2000.2725>
- Lee, H. W., Rayner, K., & Pollatsek, A. (1999). The time course of phonological, semantic, and orthographic coding in reading: Evidence from the fast-priming technique. *Psychonomic Bulletin and Review*, 6(4), 624–634. <https://doi.org/10.3758/BF03212971>
- Leech, G. N. (1969). *A linguistic guide to English poetry*. London, United Kingdom: Longmans.
- Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. Chips: An ideal-observer model of

- reading. *Psychological Review*, *104*(3), 524–553. <https://doi.org/10.1037/0033-295X.104.3.524>
- Lehne, M., Engel, P., Rohrmeier, M., Menninghaus, W., Jacobs, A. M., & Koelsch, S. (2015). Reading a suspenseful literary text activates brain areas related to social cognition and predictive inference. *PLoS One*, *10*(5). <https://doi.org/10.1371/journal.pone.0124550>
- Levy, B. A., di Persio, R., & Hollingshead, A. (1992). Fluent rereading: Repetition, automaticity, and discrepancy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 957–971. <https://doi.org/10.1037/0278-7393.18.5.957>
- Levy, B. A., Masson, M. E., & Zoubek, M. A. (1991). Rereading Text: Words and Their Context. *Canadian Journal of Psychology Revue Canadienne de Psychologie*, *45*(4), 492–506. <https://doi.org/10.1037/h0084308>
- Levy, B. A., Nicholls, A., & Kohen, D. (1993). Repeated Readings: Process Benefits for Good and Poor Readers. *Journal of Experimental Child Psychology*, *56*(3), 303–327. <https://doi.org/10.1006/JECP.1993.1037>
- Lopopolo, A., Frank, S. L., & Willems, R. M. (2019). *Dependency Parsing with your Eyes: Dependency Structure Predicts Eye Regressions During Reading*.
- Lorch, R. F., & Myers, J. L. (1990). Regression Analyses of Repeated Measures Data in Cognitive Research. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 149–157. <https://doi.org/10.1037/0278-7393.16.1.149>
- 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. <https://doi.org/10.3758/s13428-016-0748-7>
- Lüdtke, J., & Jacobs, A. M. (2015). The emotion potential of simple sentences: additive or interactive effects of nouns and adjectives? *Frontiers in Psychology*, *6*. <https://doi.org/10.3389/fpsyg.2015.01137>
- Lüdtke, J., Meyer-Sickendieck, B., & Jacobs, A. M. (2014). Immersing in the stillness of an early morning: Testing the mood empathy hypothesis of poetry reception. *Psychology of Aesthetics, Creativity, and the Arts*, *8*(3), 363–377. <https://doi.org/10.1037/a0036826>
- Magyari, L., Mangen, A., Kuzmičová, A., Jacobs, A. M., & Lüdtke, J. (2020). Eye movements and mental imagery during reading of literary texts with different narrative styles. *Journal of Eye Movement Research*. <https://doi.org/10.16910/jemr.13.3.3>

- Maïonchi-Pino, N., Cara, B. De, 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).
- Maïonchi-Pino, N., de Cara, B., Écalle, 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. <https://doi.org/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. <https://doi.org/10.1046/j.1365-2664.1999.00440.x>
- Margolin, S. J., & Snyder, N. (2018). It may not be that difficult the second time around: the effects of rereading on the comprehension and metacomprehension of negated text. *Journal of Research in Reading*, 41(2), 392–402. <https://doi.org/10.1111/1467-9817.12114>
- Marr, M. J. (2018). Overtaking the Creases: Introduction to Francis Mechner’s “A Behavioral and Biological Analysis of Aesthetics.” *Psychological Record*, 68(3), 285–286. <https://doi.org/10.1007/s40732-018-0307-y>
- Martin Cichy, R., 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. <https://doi.org/10.1016/j.neuroimage.2016.03.063>
- 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. <https://doi.org/10.1080/10888438.2015.1107073>
- Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist*, 70(6), 487–498. <https://doi.org/10.1037/a0039400>
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, 88(5), 375–407. <https://doi.org/10.1037/0033-295X.88.5.375>

- McNamara, D. S., & Magliano, J. (2009). Toward a Comprehensive Model of Comprehension. *Psychology of Learning and Motivation, 51*, 297–384. [https://doi.org/10.1016/S0079-7421\(09\)51009-2](https://doi.org/10.1016/S0079-7421(09)51009-2)
- Megalakaki, O., Ballenghein, U., & Baccino, T. (2019). Effects of Valence and Emotional Intensity on the Comprehension and Memorization of Texts. *Frontiers in Psychology, 10*, 179. <https://doi.org/10.3389/fpsyg.2019.00179>
- Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs: Language affects low-level visual perception. *Psychological Science, 18*(11), 1007–1013. <https://doi.org/10.1111/j.1467-9280.2007.02016.x>
- Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex, 48*(7), 788–804. <https://doi.org/10.1016/j.cortex.2010.11.002>
- Miall, D. S., & Kuiken, D. (1994). Foregrounding, defamiliarization, and affect: Response to literary stories. *Poetics, 22*(5), 389–407. [https://doi.org/10.1016/0304-422X\(94\)00011-5](https://doi.org/10.1016/0304-422X(94)00011-5)
- Miall, D. S., & Kuiken, D. (2002). A feeling for fiction: Becoming what we behold. *Poetics, 30*(4), 221–241. [https://doi.org/10.1016/S0304-422X\(02\)00011-6](https://doi.org/10.1016/S0304-422X(02)00011-6)
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*. International Conference on Learning Representations, ICLR.
- Millis, K. K. (1995). Encoding discourse perspective during the reading of a literary text. *Poetics, 23*(3), 235–253. [https://doi.org/10.1016/0304-422X\(94\)00028-5](https://doi.org/10.1016/0304-422X(94)00028-5)
- Millis, K. K., & King, A. (2001). Rereading strategically: The influences of comprehension ability and a prior reading on the memory for expository text. *Reading Psychology, 22*(1), 41–65. <https://doi.org/10.1080/02702710151130217>
- Millis, K. K., & Simon, S. (1994). Rereading scientific texts: Changes in resource allocation. *Advances in Discourse Processes, 53*, 115–134.
- Millis, K. K., Simon, S., & TenBroek, N. S. (1998). Resource allocation during the rereading of scientific texts. *Memory and Cognition, 26*(2), 232–246. <https://doi.org/10.3758/BF03201136>

- Mukarovský, J. (1964). Standard Language and Poetic Language. In P. Garvin (Ed.), *A Prague school reader on esthetics, literary structure, and style* (pp. 31–69). Washington, DC: Georgetown University Press.
- Müller, H., 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). <https://doi.org/10.16910/10.1.4>
- 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. <https://doi.org/10.1016/J.ENGGEOL.2008.01.004>
- Nell, V. (1988). *Lost in a book: The psychology of reading for pleasure*. New Haven/London: Yale University Press.
- Nicklas, P., & Jacobs, A. M. (2017). Rhetoric, Neurocognitive Poetics, and the Aesthetics of Adaptation. *Poetics Today*, *38*(2), 393–412. <https://doi.org/10.1215/03335372-3869311>
- Niikuni, K., Iwasaki, S., & Muramoto, T. (2015). *The Role of Punctuation in Processing Relative-Clause Sentence Constructions in Japanese*. *undefined*.
- Oatley, K. (1995). A taxonomy of the emotions of literary response and a theory of identification in fictional narrative. *Poetics*, *23*(1–2), 53–74. [https://doi.org/10.1016/0304-422X\(94\)P4296-S](https://doi.org/10.1016/0304-422X(94)P4296-S)
- Oslund, E. L., Clemens, N. H., Simmons, D. C., & Simmons, L. E. (2018). The direct and indirect effects of word reading and vocabulary on adolescents' reading comprehension: Comparing struggling and adequate comprehenders. *Reading and Writing*, *31*(2), 355–379. <https://doi.org/10.1007/s11145-017-9788-3>
- Oslund, E. L., Clemens, N. H., Simmons, D. C., Smith, S. L., & Simmons, L. E. (2016). How vocabulary knowledge of middle-school students from low socioeconomic backgrounds influences comprehension processes and outcomes. *Learning and Individual Differences*, *45*, 159–165. <https://doi.org/10.1016/j.lindif.2015.11.013>
- 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), 767–779. <https://doi.org/10.1037/0096-1523.24.3.767>
- Perfetti, C. A., Landi, N., & Oakhill, J. (2008). The Acquisition of Reading Comprehension

- Skill. In *The Science of Reading: A Handbook* (pp. 227–247). John Wiley and Sons.
<https://doi.org/10.1002/9780470757642.ch13>
- Perfetti, C., & Stafura, J. (2014). Word Knowledge in a Theory of Reading Comprehension. *Scientific Studies of Reading, 18*(1), 22–37.
<https://doi.org/10.1080/10888438.2013.827687>
- Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114*(2), 273–315. <https://doi.org/10.1037/0033-295X.114.2.273>
- Pickering, M. J., Frisson, S., McElree, B., & Traxler, M. J. (2004). Eye movements and semantic composition. In *The On-line Study of Sentence Comprehension: Eyetracking, ERPs and Beyond* (pp. 33–50). Psychology Press.
- Ponz, A., Liégeois-Chauvel, C., Silva, C., & Braun, M. (2013). Emotion processing in words: A test of the neural re-use hypothesis using surface and intracranial EEG. *Article in Social Cognitive and Affective Neuroscience*. <https://doi.org/10.1093/scan/nst034>
- Price, C. J. (2012, August 15). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*. Academic Press.
<https://doi.org/10.1016/j.neuroimage.2012.04.062>
- Price, C. J., & Devlin, J. T. (2011, June 1). The Interactive Account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences*. Elsevier Current Trends. <https://doi.org/10.1016/j.tics.2011.04.001>
- Pulvermüller, F. (2005, June 15). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*. Nature Publishing Group. <https://doi.org/10.1038/nrn1706>
- 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), 1–11. <https://doi.org/10.16910/jemr.2.1.4>
- Pynte, J., New, B., & Kennedy, A. (2009). On-line contextual influences during reading normal text: The role of nouns, verbs and adjectives. *Vision Research, 49*(5), 544–552.
<https://doi.org/10.1016/J.VISRES.2008.12.016>
- Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological Research, 72*(6), 675–688.
<https://doi.org/10.1007/s00426-008-0173-3>

- Radach, R., & Kennedy, A. (2013). Eye movements in reading: Some theoretical context. *Quarterly Journal of Experimental Psychology*, *66*(3), 429–452. <https://doi.org/10.1080/17470218.2012.750676>
- Raney, G. E. (2003). A context-dependent representation model for explaining text repetition effects. *Psychonomic Bulletin & Review*, *10*(1), 15–28. <https://doi.org/10.3758/BF03196466>
- 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), 151–173. <https://doi.org/10.1037/1196-1961.49.2.151>
- Raney, G. E., Theriault, D. J., & Minkoff, S. R. B. (2000). Repetition Effects From Paraphrased Text: Evidence for an Integrated Representation Model of Text Representation. *Discourse Processes*, *29*(1), 61–81. https://doi.org/10.1207/S15326950dp2901_4
- Rawson, K. A., Dunlosky, J., & Thiede, K. W. (2000). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition*, *28*(6), 1004–1010. <https://doi.org/10.3758/BF03209348>
- Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. <https://doi.org/10.1080/17470210902816461>
- 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. [https://doi.org/10.1016/S0042-6989\(00\)00310-2](https://doi.org/10.1016/S0042-6989(00)00310-2)
- 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. https://doi.org/10.1207/s1532799xssr1003_3
- Rayner, K., & Pollatsek, A. (2006). *Eye-Movement Control in Reading. Handbook of Psycholinguistics* (Second). Academic Press. <https://doi.org/10.1016/B978-012369374-7/50017-1>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. <https://doi.org/10.1037/0033->

2909.124.3.372

- Rayner, Keith, & Liversedge, S. P. (2011). Linguistic and cognitive influences on eye movements during reading. In P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 751–766). Oxford University Press.
<https://doi.org/10.1093/OXFORDHB/9780199539789.013.0041>
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). *Cross-Validation. Encyclopedia of Database Systems*. New York: Springer. https://doi.org/10.1007/978-1-4899-7993-3_602-2
- Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26(4), 445–476. <https://doi.org/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.
<https://doi.org/10.1016/J.COGSYS.2005.07.006>
- Roland, D., Yun, H., Koenig, J. P., & Mauener, G. (2012). Semantic similarity, predictability, and models of sentence processing. *Cognition*, 122(3), 267–279.
<https://doi.org/10.1016/j.cognition.2011.11.011>
- Saltelli, A. (2002). Sensitivity Analysis for Importance Assessment. *Risk Analysis*, 22(3), 579–590. <https://doi.org/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.
<https://doi.org/10.1080/02699931.2017.1279591>
- Schacht, A., & Sommer, W. (2009). Time course and task dependence of emotion effects in word processing. *Cognitive, Affective and Behavioral Neuroscience*, 9(1), 28–43.
<https://doi.org/10.3758/CABN.9.1.28>
- Schmidt, S. J. (1982). *Foundation for the empirical study of literature : the components of a basic theory*. (R. de Beaugrande, Ed.). Hamburg: Buske.
- Schmidt, S. J. (1989). On the construction of fiction and the invention of facts. *Poetics*, 18(4–5), 319–335. [https://doi.org/10.1016/0304-422X\(89\)90035-1](https://doi.org/10.1016/0304-422X(89)90035-1)
- Schmidtke, D. S., Conrad, M., & Jacobs, A. M. (2014). Phonological iconicity. *Frontiers in*

- Psychology*, 5, 80. <https://doi.org/10.3389/fpsyg.2014.00080>
- Schmidtke, D. S., Schröder, T., Jacobs, A. M., & Conrad, M. (2014). ANGST: affective norms for German sentiment terms, derived from the affective norms for English words. *Behavior Research Methods*, 46(4), 1108–1118. <https://doi.org/10.3758/s13428-013-0426-y>
- Schnitzer, B. S., & Kowler, E. (2006). Eye movements during multiple readings of the same text. *Vision Research*, 46(10), 1611–1632. <https://doi.org/10.1016/j.visres.2005.09.023>
- 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.
- Scott, G. G., O'Donnell, P. J., Leuthold, H., & Sereno, S. C. (2009). Early emotion word processing: Evidence from event-related potentials. *Biological Psychology*, 80(1), 95–104. <https://doi.org/10.1016/j.biopsycho.2008.03.010>
- Scott, G. G., O'Donnell, P. J., & Sereno, S. C. (2012). Emotion words affect eye fixations during reading. *Journal of Experimental Psychology: Learning Memory and Cognition*, 38(3), 783–792. <https://doi.org/10.1037/a0027209>
- Sears, C. R., Campbell, C. R., & Lupker, S. J. (2006). Is there a neighborhood frequency effect in english? evidence from reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, 32(4), 1040–1062. <https://doi.org/10.1037/0096-1523.32.4.1040>
- Seidenberg, M. S. (2005). Connectionist Models of Word Reading. *Current Directions in Psychological Science*, 14(5), 238–242. <https://doi.org/10.1111/j.0963-7214.2005.00372.x>
- Seidenberg, M. S. (2007). Connectionist models of reading. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 235–250). New York: Oxford University Press.
- Shrout, P. E., & Rodgers, J. L. (2018). Psychology, Science, and Knowledge Construction: Broadening Perspectives from the Replication Crisis. *Annual Review of Psychology*, 69, 487–510. <https://doi.org/10.1146/annurev-psych-122216>
- Siakaluk, P. D., Sears, C. R., & Lupker, S. J. (2002). Orthographic neighborhood effects in lexical decision: The effects of nonword orthographic neighborhood size. *Journal of*

- Experimental Psychology: Human Perception and Performance*, 28(3), 661–681.
<https://doi.org/10.1037/0096-1523.28.3.661>
- Simonto, D. K. (1989). Shakespeare's Sonnets: A Case of and for Single-Case Historiometry. *Journal of Personality*, 57(3), 695–721. <https://doi.org/10.1111/j.1467-6494.1989.tb00568.x>
- Speer, N. K., Reynolds, J. R., Swallow, K. M., & Zacks, J. M. (2009). Reading stories activates neural representations of visual and motor experiences. *Psychological Science*, 20(8), 989–999. <https://doi.org/10.1111/j.1467-9280.2009.02397.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(2), 280–292. <https://doi.org/10.1016/J.BANDL.2005.01.013>
- Steyer, R., Schwenkmezger, P., Notz, P., & Eid, M. (1997). *Der Mehrdimensionale Befindlichkeitsfragebogen (MDBF)*. Göttingen: Hogrefe.
- Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic Effects in Single-Word Naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(5), 1140–1154. <https://doi.org/10.1037/0278-7393.21.5.1140>
- 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), 323–348.
<https://doi.org/10.1037/a0016973>
- Sun, F., Morita, M., & Stark, L. W. (1985). Comparative patterns of reading eye movement in Chinese and English. *Perception & Psychophysics*, 37(6), 502–506.
<https://doi.org/10.3758/BF03204913>
- Sylvester, T., Braun, M., Schmidtke, D., & Jacobs, A. M. (2016). The Berlin Affective Word List for Children (kidBAWL): Exploring Processing of Affective Lexical Semantics in the Visual and Auditory Modalities. *Frontiers in Psychology*, 7, 969.
<https://doi.org/10.3389/fpsyg.2016.00969>
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2), 267–307.
https://doi.org/10.1162/COLI_a_00049
- 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(02), 135–178. <https://doi.org/10.1017/S0954394512000129>
- Thorne, J. P. (1989). What is a poem? In W. van Peer (Ed.), *The Taming of the Text: Explorations in Language, Literature and Culture*. London: Routledge.
- 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. [https://doi.org/10.1016/S0895-4356\(96\)00002-9](https://doi.org/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 H. M. Enzensberger. *Frontiers in Psychology*, 7, 2073. <https://doi.org/10.3389/fpsyg.2016.02073>
- van den Hoven, E., Hartung, F., Burke, M., & Willems, R. M. (2016). Individual Differences in Sensitivity to Style During Literary Reading: Insights from Eye-Tracking. *Collabra*, 2(1), 25. <https://doi.org/10.1525/collabra.39>
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York, NY, US: Academic Press. <https://doi.org/10.1.1.473.5491>
- 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. <https://doi.org/10.1080/09296170500055350>
- van Peer, W. (1989). *Stylistics and psychology: Investigations of foregrounding*. Taylor & Francis.
- van Peer, W. (1990). The measurement of metre. Its cognitive and affective functions. *Poetics*, 19(3), 259–275. [https://doi.org/10.1016/0304-422X\(90\)90023-X](https://doi.org/10.1016/0304-422X(90)90023-X)
- Vendler, H. (1997). *The art of Shakespeare's sonnets*. Belknap Press of Harvard University Press.
- Võ, M. L. H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, 41(2), 534–538. <https://doi.org/10.3758/BRM.41.2.534>
- Võ, M. L. H., Jacobs, A. M., & Conrad, M. (2006). Cross-validating the Berlin Affective

- Word List. *Behavior Research Methods*, 38(4), 606–609.
<https://doi.org/10.3758/BF03193892>
- 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.
<https://doi.org/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. <https://doi.org/10.1016/J.ECOLIND.2014.12.028>
- Westbury, C. F., Shaoul, C., Hollis, G., Smithson, L., Briesemeister, B. B., Hofmann, M. J., & Jacobs, A. M. (2013). Now you see it, now you don't: on emotion, context, and the algorithmic prediction of human imageability judgments. *Frontiers in Psychology*, 4.
<https://doi.org/10.3389/fpsyg.2013.00991>
- Westbury, C., Keith, J., Briesemeister, B. B., Hofmann, M. J., & Jacobs, A. M. (2015). Avoid violence, rioting, and outrage; approach celebration, delight, and strength: Using large text corpora to compute valence, arousal, and the basic emotions. *Quarterly Journal of Experimental Psychology*, 68(8), 1599–1622.
<https://doi.org/10.1080/17470218.2014.970204>
- Willems, Roel M. (2015). *Cognitive Neuroscience of Natural Language Use*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781107323667>
- Willems, Roel M., & Jacobs, A. M. (2016). Caring About Dostoyevsky: The Untapped Potential of Studying Literature. *Trends in Cognitive Sciences*, 20(4), 243–245.
<https://doi.org/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), 1072–1082.
<https://doi.org/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. In R. Radach, H. Deubel, C. Vorstius, & M. J. Hofmann (Eds.), *19th European Conference on Eye Movements* (pp. 115–116). Wuppertal, Germany: University of Wuppertal.

- Xue, S., Lüdtke, 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).
<https://doi.org/10.16910/jemr.12.5.2>
- Xue, S., Jacobs, A. M., & Lüdtke, J. (2020). What Is the Difference? Rereading Shakespeare's Sonnets —An Eye Tracking Study. *Frontiers in Psychology*, *11*, 421.
<https://doi.org/10.3389/fpsyg.2020.00421>
- Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. <https://doi.org/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), 567–593. <https://doi.org/10.1006/JMLA.1995.1026>
- Ziegler, Johannes C., Montant, M., Briesemeister, B. B., Brink, T. T., Wicker, B., Ponz, A., ... Braun, M. (2018). Do words stink? Neural reuse as a principle for understanding emotions in reading. *Journal of Cognitive Neuroscience*, *30*(7), 1023–1032.
https://doi.org/10.1162/jocn_a_01268
- Zwaan, R. A. (2004). The Immersed Experiencer: Toward an Embodied Theory of Language Comprehension. *Psychology of Learning and Motivation*, *44*, 35–62.
- Zwaan, R. A. (2014). Embodiment and language comprehension: Reframing the discussion. *Trends in Cognitive Sciences*. Elsevier Ltd. <https://doi.org/10.1016/j.tics.2014.02.008>
- Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of Situation Model Construction in Narrative Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 386–397. <https://doi.org/10.1037/0278-7393.21.2.386>
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation Models in Language Comprehension and Memory. *Psychological Bulletin*, *123*(2), 162–185. <https://doi.org/10.1037/0033-2909.123.2.162>
- Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1–11.
<https://doi.org/10.1037/0096-3445.135.1.1>

List of Figures

Figure 2.1.1 The Neurocognitive Poetics Model (NCPM) by Jacobs (taken from Jacobs, 2015a)

Figure 2.1.2 The 4×4 Matrix Illustrating Four Levels of Text Crossed with Four Groups of Features, with one Feature as an Example for Each Cell of the Matrix (from Jacobs, 2015a)

Figure 5.1 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 into two parts: the 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. 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.

Figure 5.2 Model Fits of Different Measure Groups via Different Modeling Methods. This figure shows the mean R^2 from 1000 iterations for three eye tracking parameters for both the training and test sets using all three modeling approaches. Each error bar is constructed using 1 standard deviation from the mean.

Figure 5.3 Feature Importances for Total Reading Time and Fixation Probability. Figure 5.3 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 Corpusas reference (GLEC; Jacobs, 2018b): *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 constructed using 1 standard deviation from the mean. (Note that, because of the bad model fits (see Figure 5.2), the *FIs* in explaining first fixation duration were excluded from this figure).

Figure 6.1 The Procedure of the Experiment. “1st” and “2nd” refer to the first and second sonnet.

Figure 6.2 Rereading Effect on Rating Data. (A) “Willingness to do any rereading”, (B) “Topic identification”, and (C) “Appreciation” were separately collected from three questions: “I would like to read this poem again”, “Which is the main topic of this poem”, and “I like this poem”. For questions related to “Willingness to do any rereading” and “Appreciation,” readers indicated their agreement with the statements on a 5-point rating scale ranging from 1 = totally disagree to 5 = totally agree. For the topic identification question, six choices were offered, but only one was right. If readers agreed with none of the choices, they could put forward another, which was later evaluated by two experts from the humanities. $*p < 0.05$. Error bar is constructed using one standard deviation from the mean.

Figure 6.3 Rereading Effect on Eye Tracking Parameters. To test for the rereading effects on word-level eye tracking parameters, linear mixed models (LMM) with one fixed effect (session) and one random effect (word nested within sonnet) were applied to the five eye tracking parameters (A) “First fixation duration”, (B) “Gaze duration”, (C) “Regression time”, (D) “Total reading time”, (E) “Fixation probability”. $**p < 0.01$. Error bar is constructed using one standard deviation from the mean.

Figure 6.4 Fit Scores for Different Models and Measures. For neural nets (A), R^2 s from 100 iterations were averaged for both the training and test sets. For standard least squares regressions (B), the R^2 for the whole data set and the mean R^2 s from 100 iterations for the test sets were calculated. Nine predictors (*lineNo.*, *wordNo.*, *wl*, *logf*, *on*, *hfn*, *odc*, *cvq*, and *sonscore*) and five response parameters (first fixation duration, gaze duration, regression time, total reading time, and fixation probability) were included in analyses. Each error bar is constructed using one standard deviation from the mean.

Figure 6.5 Feature Importance for the Five Eye Tracking Parameters

Figure 7.1 Rereading Effect on Rating Data. (A) “Supralelexical Valence”, and (B) “Supralelexical Arousal” were separately collected from two questions: “In general, poems can express positive or negative emotions. While reading this poem for the first time, did it feel negative or positive?” and “In general, poems can evoke feelings ranging from boredom to excitement. While reading this poem for the first time, did it feel calming or exciting?”. For the question about Supralelexical Valence, readers indicated their agreement with the statements on a 7-point rating scale ranging from -3 = extremely negative to 3 = extremely positive. For Supralelexical Arousal, readers indicated their agreement with the statements on a 5-point rating scale ranging from 1 = very calming to 5 = very exciting. Error bar is constructed using one standard deviation from the mean.

Figure 7.2 Fit Scores for Different Models and Measures. R^2 s from 100 iterations were averaged for both the training and test sets. Eleven predictors (*lineNo.*, *wordNo.*, *wl*, *logf*, *on*, *hfn*, *odc*, *cvq*, *sonscore*, *val*, and *aro*) and five response parameters (first fixation duration, gaze duration, regression time, total reading time, and fixation probability) were included in analyses. Each error bar is constructed using one standard deviation from the mean.

Figure 7.3 Feature Importance for the Five Eye Tracking Parameters

List of Tables

Table 5.1 Number of Words per Category within Each Sonnet and within all Three Sonnets

Table 5.2 Correlations between Seven QNA Features

Table 5.3 Correlations between Five Common Eye-movement Parameters used in Reading Research

Table 6.1 Correlations between the Five Eye Tracking Parameters

Table 6.2 Correlations between the Seven Psycholinguistic Features

Table 8.1 Overview of all Three Studies Regarding Their Experimental Details and Main Outcomes

Appendices

Appendix A.1. Sonnets used in eye tracking experiments

Sonnet 27

Weary with toil, I haste me to my bed,
 The dear repose for limbs with travel tired;
 But then begins a journey in my head
 To work my mind when body's work expired.
 For then my thoughts, from far where I abide,
 Intend a zealous pilgrimage to thee,
 And keep my drooping eyelids open wide,
 Looking on darkness which the blind do see.
 Save that my soul's imaginary sight
 Presents thy shadow to my sightless view,
 which like a jewel hung in ghastly night
 Makes black night beauteous and her old face new.]
 So, thus by day my limbs, by night my mind,
 For thee, and for myself, no quiet find.

Sonnet 60

Like as the waves make towards the pebbled shore,
 So do our minutes hasten to their end,
 Each changing place with that which goes before,
 In sequent toil all forwards do contend.
 Nativity, once in the main light,
 Crawls to maturity, wherewith being crowned,
 Crooked eclipses 'gainst his glory fight,
 And Time that gave doth now his gift confound.
 Time doth transfix the flourish set on youth,
 And delves the parallels in beauty's brow,
 Feeds on the rarities of nature's truth,
 And nothing stands but for his scythe to mow.
 And yet to times in hope my verse shall stand,
 Praising thy worth, despite his cruel hand.

Sonnet 66

Tired with all these, for restful death I cry:
As to behold desert a beggar born,
And needy nothing trimmed in jollity,
And purest faith unhappily forsworn,
And gilded honour shamefully misplaced,
And maiden virtue rudely strumpeted,
And right perfection wrongly disgraced,
And strength by limping sway disabled,
And art made tongue-tied by authority,
And folly, doctor like, controlling skill,
And simple truth miscalled simplicity,
And captive good attending captain ill.
Tired with all these, from these would I be gone,
Save that, to die, I leave my love alone.

Appendix A.2. An example of the paper-pencil tasks used in the eye tracking experiments

Please read each of the following statements carefully, and then answer the questions or rate to which extent you agree with the statement with regard to your experience while reading this poem:

	Totally disagree				Totally agree
	1	2	3	4	5
1. I would like to read this poem again.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

2. Generally, poetry contains many rhymes. In this poem, which rhyme pairs did appear, and which not?

	appear	Not appear
1. mind – find	<input type="radio"/>	<input type="radio"/>
2. sight – night	<input type="radio"/>	<input type="radio"/>
3. fade – shade	<input type="radio"/>	<input type="radio"/>
4. day – May	<input type="radio"/>	<input type="radio"/>
5. view – new	<input type="radio"/>	<input type="radio"/>
6. minds – finds	<input type="radio"/>	<input type="radio"/>

3. Which is the main topic of this poem?

Please indicate only one topic.

Eternity of poetry	<input type="radio"/>
The nature of time as it passes	<input type="radio"/>
Love as triumph of beauty	<input type="radio"/>
Beauty as expression of interior qualities	<input type="radio"/>
Love as never ending tension of soul and body	<input type="radio"/>
A desperate list of grievances of the state of the poet's society	<input type="radio"/>
Other _____	<input type="radio"/>

4. Please underline (“___”) the most expressive/ powerful SINGLE line of this poem.

Please also mark the key word(s) important for expressing the main idea of the poem using the yellow pen.

Weary with toil, I haste me to my bed,
The dear repose for limbs with travel tired;
But then begins a journey in my head
To work my mind, when body's work's expired:
For then my thoughts — from far where I abide —
Intend a zealous pilgrimage to thee,
And keep my drooping eyelids open wide,
Looking on darkness which the blind do see:
Save that my soul's imaginary sight
Presents thy shadow to my sightless view,
Which, like a jewel hung in ghastly night,
Makes black night beauteous, and her old face new.
Lo! thus, by day my limbs, by night my mind,
For thee, and for myself, no quiet find.

5. Please look through the poem again. Circle (“O”) the unknown word(s) of this poem.

6. Did the poem evoke any mental image(s)?

Yes No

If your answer is “Yes”, please write down the evoked image(s).

7. In general, poems can express positive or negative emotions. While reading this poem for the first time, did it feel negative or positive?

Extremely negative	Very negative	Negative	Neutral	Positive	Very positive	Extremely positive
-3	-2	-1	0	1	2	3
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please write down in your own words, which feelings are described in the poem.

8. In general, poems can evoke feelings ranging from boredom to excitation. While reading this poem for the first time, did it feel calming or exciting?

Very calming	Calming	Neither nor	Exciting	Very exciting
1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

9. Which color do you associate with this poem? Please choose one.

Red	<input type="radio"/>
Orange	<input type="radio"/>
Yellow	<input type="radio"/>
Green	<input type="radio"/>
Blue	<input type="radio"/>
Indigo	<input type="radio"/>
Purple	<input type="radio"/>

10. Please think about the temperature you associated with the poem. While reading this poem, did it feel rather warm or cold?

	very cold	cold	moderate	warm	very warm
Which temperature do you associate with the poem?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Totally disagree			Totally agree	
	1	2	3	4	5
11. I like this poem.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12. I was touched by this poem.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13. I was concerned by this poem.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
14. This poem is easy to understand.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
15. This poem inspired me to think.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
16. While reading this poem, I felt intense delight.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
17. While reading this poem, I felt profound wonder.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
18. While reading this poem, I felt deeply astonished.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

19. Imagine you read the poem aloud, how would the following characteristics affect the flow of reading in this poem?

	Not at all 1	Slightly 2	Moderately 3	Quite 4	Extremely 5
Rhyme (similarity of sound at the end of words or lines)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Break (simple rest periods between words or lines)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Stress (particular accentuations for syllables or words)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

20. Have your feelings changed through reading this poem for the first time?

Yes No

If your answer is “**Yes**”, please write down how your feelings have changed.

Appendix A.3. An example of the scripts used for predictive modeling

Chaper 5_Neural nets_Total reading time:

```

names default to here(1);
dt=Current Data Table();
dtb=New Table( "training_Total reading time",
  Add Rows( 0 ),
  New Column( "StringColBox",
    Character,
    "Nominal"
  )
);
dtc=New Table( "validation_Total reading time",
  Add Rows( 0 ),
  New Column( "StringColBox",
    Character,
    "Nominal"
  )
);
dtd = New Table("rank_DRI_Total reading time",
  Add Rows( 0 ),
  New Column( "Column",
    Character,
    "Nominal"
  ),
  New Column( "Main Effect",
    Numeric,
    "Continuous",
    Format( "Best", 12 ),
    Set Selected,
    Set Values( [] )
  ),
  New Column( "Total Effect",
    Numeric,
    "Continuous",
    Format( "Best", 12 ),
    Set Selected,
    Set Values( [] )
  )
);
dte = New Table("rank_IRI_Total reading time",
  Add Rows( 0 ),
  New Column( "Column",
    Character,
    "Nominal"
  ),
  New Column( "Main Effect",
    Numeric,
    "Continuous",
    Format( "Best", 12 ),
    Set Selected,
    Set Values( [] )
  ),
  New Column( "Total Effect",

```



```

        Numeric,
        "Continuous",
        Format( "Best", 12 ),
        Set Selected,
        Set Values( [] )
    )
);
dtf = New Table("rank_IUI_Total reading time",
    Add Rows( 0 ),
    New Column( "Column",
        Character,
        "Nominal"
    ),
    New Column( "Main Effect",
        Numeric,
        "Continuous",
        Format( "Best", 12 ),
        Set Selected,
        Set Values( [] )
    ),
    New Column( "Total Effect",
        Numeric,
        "Continuous",
        Format( "Best", 12 ),
        Set Selected,
        Set Values( [] )
    )
);
for (i=1, i<=1000, i++,
    dt<<Make Validation Column(
        Training Set( 0.90 ),
        Validation Set( 0.10 ),
        Test Set( 0.00 ),
        Formula Random
    );
    obj=dt<<Neural(
        Y( :Std Mean_DWELL_TIME),
        X(
            :Std wl,
            :Std cvq,
            :Std sonscore,
            :Std on,
            :Std odc,
            :Std logf,
            :Std hfn
        ),
        Informative Missing( 0 ),
        Validation( :Validation ),
        Fit(
            NTanH( 3 ),
            Number of Tours( 10 ),
            N Boost( 10 ),
            Profiler(
                1,
                Confidence Intervals( 1 ),

```

```

Dependent Resampled Inputs( 1 ),
Independent Resampled Inputs( 1 ),
Independent Uniform Inputs( 1 ),
Reorder X Variables(
  :Std wl,
  :Std on,
  :Std logf,
  :Std cvq,
  :Std odc,
  :Std sonscore,
  :Std hfn
),
Term Value(
  Std wl( 0, Lock( 0 ), Show( 1 ) ),
  Std cvq( 0, Lock( 0 ), Show( 1 ) ),
  Std sonscore( 0, Lock( 0 ), Show( 1 ) ),
  Std odc( 0, Lock( 0 ), Show( 1 ) ),
  Std on( 0, Lock( 0 ), Show( 1 ) ),
  Std logf( 0, Lock( 0 ), Show( 1 ) ),
  Std hfn( 0, Lock( 0 ), Show( 1 ) )
)
),
invisible
);
objq=report(obj);
sumFit=objq[tablebox(1)];
dt1=sumfit<<make into data table();
dt1<<set name("Summary of fit "||char(i));
dt1<<new column("Iteration",formula(i));
dtb << Concatenate( dt1, "Append to first table" );
close(dt1, nosave);
objr=report(obj);
sumFit=objr[tablebox(2)];
dt2=sumfit<<make into data table();
dt2<<set name("Summary of fit "||char(i));
dt2<<new column("Iteration",formula(i));
dte << Concatenate( dt2, "Append to first table" );
close(dt2, nosave);
objj=report(obj);
summary=objj[tablebox(3)];
dt3=summary<<make into data table();
dt3<<set name("Summary report "||char(i));
dt3<<new column("Iteration",formula(i));
dtd << Concatenate( dt3, "Append to first table" );
close(dt3, nosave);
objt=report(obj);
summary=objt[tablebox(4)];
dt4=summary<<make into data table();
dt4<<set name("Summary report "||char(i));
dt4<<new column("Iteration",formula(i));
dte << Concatenate( dt4, "Append to first table" );
close(dt4, nosave);
obju=report(obj);
summary=obju[tablebox(5)];

```

```
dt5=summary<<make into data table();
dt5<<set name("Summary report "||char(i));
dt5<<new column("Iteration",formula(i));
dtf << Concatenate( dt5, "Append to first table" );
close(dt5, nosave);
dt:Validation << Set Selected;
dt << Delete Columns();

);
```


List of Publications from this Thesis

- Xue, S.**, Lüdtke, J., & Jacobs, A. M. (in prep). Once Known, Twice Hedonic: Enjoying Shakespeare's Sonnets Through Rereading: A Reanalysis.
- Xue, S.**, Lüdtke, J., & Jacobs, A. M. (2020). What is the Difference? Rereading Shakespeare's Sonnets — an Eye Tracking Study. *Frontiers in Psychology*, *11*, 421. <https://doi.org/10.3389/fpsyg.2020.00421>
- Xue, S.**, Lüdtke, 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). <https://doi.org/10.16910/jemr.12.5.2>
- 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*(1), 4–51. <https://doi.org/10.1075/ssol.7.1.02jac>

Talks Given on Conferences

- 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*. Wuppertal, Germany.

Curriculum Vitae

Der Lebenslauf ist aus Gründen des Datenschutzes nicht enthalten.

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt,

- dass ich die vorliegende Arbeit selbständig und ohne unerlaubte Hilfe verfasst habe,
- dass ich mich nicht bereits anderwärts um einen Doktorgrad beworben habe und keinen Doktorgrad in dem Promotionsfach Psychologie besitze,
- dass ich die zugrunde liegende Promotionsordnung kenne.

Berlin, den 30.06.2020

Shuwei Xue