What's in the brain that ink may character ….

A quantitative narrative analysis of Shakespeare's 154 sonnets for use in (Neuro-)cognitive poetics

Arthur M. Jacobs[1,2,3], Sarah Schuster[1,4], Shuwei Xue[1] and Jana Lüdtke[1,2]

[1]Department of Experimental and Neurocognitive Psychology, Freie Universität, Berlin, Germany / [2]Dahlem Institute for Neuroimaging of Emotion (D.I.N.E.), Berlin, Germany / [3]Center for Cognitive Neuroscience Berlin (CCNB), Berlin, Germany / [4]Universität Salzburg, Centre for Cognitive Neuroscience, Salzburg, Austria

In this theoretical paper, we would like to pave the ground for future empirical studies in Neurocognitive Poetics by describing relevant properties of *Shakespeare's 154 sonnets* extracted via Quantitative Narrative Analysis. In the first two parts, we quantify aspects of the sonnets' cognitive and affective-aesthetic features, as well as indices of their thematic richness, symbolic imagery, and semantic association potential. In the final part, we first demonstrate how the results of these quantitative narrative analyses can be used for generating testable predictions for empirical studies of literature. Second, we feed the quantitative narrative analysis data into a machine learning algorithm which successfully classifies the 154 sonnets into two main categories, i.e. the *young man* and *dark lady* poems. This shows how quantitative narrative analysis data can be combined with computational modeling for identifying those of the many quantifiable sonnet features that may play a key role in their reception.

Keywords: neurocognitive poetics, quantitative narrative analysis, Coh-metrix, SEANCE, machine learning, digital humanities, sonnets

Although Shakespeare's works must count among the most successful and popular pieces of verbal art and have been the object of countless essays by literary critics and of theoretical – as opposed to empirical – scientific studies (e.g., Jakobson & Jones, 1970), they still seem full of surprises even for eminent experts. Thus, in the preface of his book on the language of Shakespeare "Think of my words",

Scientific Study of Literature 7:1 (2017), **4–51**. DOI 10.1075/ssol.7.1.02jac
ISSN 2210–4372 / E-ISSN 2210–4380

Crystal (2012, preface) states, "Everytime I do even the most menial search of my Shakespeare database, I discover something I have never noticed before."

Shakespeare's sonnets first appeared in 1609, a collection of about 18000 words (17515 according to our counts) that have changed the world and the way our mind-brains feel and think about it (Schrott & Jacobs, 2011). The majority of the sonnets (1–126), termed *fair youth* or *young man* sonnets, are addressed to a young man, with whom the poet is said to have had an intense relationship. In sonnets 1–17 the poet tries to convince the young man to marry and have children (e.g., beautiful children that will look just like their father, ensuring his immortality). Many of the remaining sonnets in the young man sequence focus on the power of poetry and pure love to defeat death and "all oblivious enmity" (sonnet 55.9). Sonnets 127–154, termed the *dark lady* or *mistress* sonnets, are said to speak to a promiscuous and scheming woman. Both the poet and his *fair youth* have become obsessed with the raven-haired temptress in these sonnets, and the poet's whole being is at odds with his insatiable "sickly appetite" (sonnet 147.4). The tone is distressing, with language of sensual feasting, uncontrollable urges, and sinful consumption (Shakespeare online. n.d.). This sequence is sometimes considered a proto-sketch for Shakespeare's drama Othello, although the true actors in lyric are words, not characters in conflict: The drama in the sonnets is thus produced by new linguistic strategies and internal changes in topic or syntactic structure (Vendler, 1997).

General features of sonnets

English or Shakespearean sonnets (from the Italian word *sonetto* meaning a small song or lyric) typically are decasyllabic 14-liners in iambic pentameter. Besides a clear surface structure of three (isomorphic) quatrains and one (anomalous) couplet, and a typical – but not absolute – rhyme scheme (*abab cdcd efef gg*) sonnets feature *structural coherence, logical development* and *unit of play*. According to Vendler (1997), a sonnet presents a conundrum and unfolds itself in a developing dynamic of feeling and thought marked by a unifying play of mind and language.

The sonnet's versification encourages the greater use of monosyllabic words (of which English is much richer than, e.g., German) and it allows metrical variation to be introduced more easily. Sonnets can be said to have a comparatively volatile thought structure with changes in rhyme-sound from quatrain to quatrain encouraging new turns of thought, and a step-by-step movement towards the definite closure provided by the *couplet* (Wainwright, 2011). Thus, a sonnet's structure appears good for argument (e.g., Shakespeare's sonnet 138) or polemics (e.g., Milton), sequences exploring different aspects of a single motif, e.g. love. A sonnet also is *a system in motion*: Its four parts can be set in a number of logical relations

(e.g., successive and equal, hierarchical, contrastive or contradictory, successively "louder" or "softer") and play with changes of agency or speech act, rhetorical address, grammatical form, discursive texture each producing its own emotional dynamic moves – within the speaker's mind and heart – and poetic effects in readers mindbrains (Vendler, 1997). Following Vendler, the dynamic can be assimilated to a narrowing down (funnel-shape) movement from *quatrain 1* (e.g., wide epistemological field) to *quatrain 2* (e.g., queries, contradicts, subverts position in quatrain 1) to *quatrain 3* (e.g., subtlest, most comprehensive/truthful position and solution) to the final couplet (summarizing, ironic or expansive coda – restating semantically the body of the sonnet, i.e., quatrain 1 to 3 – with a crucial tonal difference and an often a self-ironizing turn to the proverbial or idiomatic, e.g., sonnet 94). The so-called *couplet tie* are the significant, usually thematically central words from the body (quatrain 1–3) repeated in the couplet. In addition, many of the sonnets also exhibit the two-part (*octave-sestet*) structure of Petrarchan sonnets, i.e. the first eight lines logically or metaphorically stand against the last six, e.g. as a problem-solution, question-answer or generalization-application dynamic.

In a way this dynamic parallels the narrowing of the *text world* of a reader during the incremental reading act in the sense that the number of potential events, characters or new text world referents (e.g., entities, attributes, relations) decreases towards the end of a text. As argued by Steen (2004), this can have notable effects on the way readers process poetic text elements and metaphors in particular.

On the one hand, sonnets are comparable to narrative in that practically each sonnet "tells a little story". This is an advantage, since it allows to supplement qualitative and typological text analyses (e.g., Jakobson & Lévi-Strauss, 1962; Jakobson & Jones, 1970; Meireles, 2005) by *quantitative narrative analysis* (i.e., turning words into numbers) which requires a minimum of text length and structural variability to provide reliable results. On the other hand, sonnets differ from narratives in form and content structure, since they exhibit the potential for lots of new beginnings, fresh angles, different tones (intimate, meditative, comic, polemical) and do not need a narrative's thread (Wainwright, 2011). Their nice juxtaposition of the strict regularity and continuity of form against the other likely changes in subject, mood or style make them ideal candidates for evoking affective and aesthetic reader responses, and thus for scientific studies of literary experience (Delmonte, 2016; Jacobs, 2015c, 2016a). If the poet himself indeed learned to find strategies to *enact feeling in form* and replicate human (affective) responses in a unique richness and virtuosity of linguistic forms throughout the composition of the 154 sonnets (Vendler, 1997), then readers may well sequentially and incrementally acquire new insights into their own feelings. Thus, the principle of interest to sustain rereadings of sonnets are believed to be their discourse variety and fertility in structural complexity (cf. Simonton, 1989).

In sum, sonnets offer a rich structure at all textual processing levels and thus a great potential for multilevel poetic effects (cf. Jacobs, Lüdtke, Aryani, Meyer-Sickendiek, & Conrad, 2016a). Due to their short length, sonnets are relatively easy to manage for both the writer and the reader. On the other hand, they also are long enough to contain and induce alternations in moods (e.g., mood empathy changes; Jacobs et al., 2016a; Lüdtke, Meyer-Sickendiek, & Jacobs, 2014) and can be considered a great *repository of moods* induced by treating the plot elegiacally, sardonically, ironically and tragically (Vendler, 1997). In sum, they seem to be ideal candidates for empirical studies in neurocognitive poetics (Jacobs, 2015a,b).

Cognitive, evolutionary, and quantitative narrative analyses of Shakespeare's dramas and sonnets

Shakespeare's works were among the first to have been analysed in terms of cognitive and evolutionary approaches and through the use of computer tools. The latter empower researchers to quickly analyze large bodies of literary texts on many characteristics of language and discourse, thus offering predictions about their aesthetic success, artistic worth or comprehensibility (e.g., Anderson & Crossley, 2011; Delmonte, 2014; Graesser et al., 2004, 2010, 2011; Simonton, 1989, 1990). Word lists like the *General Inquirer* have been used for computerized text analyses from the sixties on (Stone, Bales, Namenwirth, & Ogilvie, 1962). As impressively illustrated by Hope and Witmore (2004) such word lists allow readers educated in a certain literary tradition to experience the information texts contain in different ways. Computational text analyses complement evolutionary and cognitive approaches to Shakespeare's texts (e.g., Boyd, 2012; Cook, 2010; Crane, 2001) by revealing a web of structures and categories through which meaning is created and tracing the complex interactions of cultural and cognitive determinants of meaning as they play themselves out in Shakespeare's texts. However, to which extent these approaches help to predict reader responses during or after reading Shakespeare's works – or other literary texts – is an open empirical question.

In his seminal quantitative narrative analysis study of Shakespeare's sonnets, Simonton (1989) discovered that the sonnets with superior aesthetic success (as assessed by an archival popularity measure) had the following distinctive features: (a) treat specific themes; (b) display considerable *thematic richness* in the number of issues discussed; (c) exhibit greater *linguistic complexity* as gauged by such objective measures as the type-token ratio (i.e., the ratio of different words to total words as an index of lexical variability/verbal complexity) and adjective-verb quotient (i.e., the proportion of adjectives to verbs as an alternative gauge of linguistic complexity); and (d) feature more *primary process language* and imagery (as assessed by Martindale's, 1975, Regressive Imagery Dictionary).

Simonton identified a few supremely popular sonnets standing out from the universe of 154 (sonnets 29, 30, 73, and 116). Based on the topical index of the *Great Books of the Western World* (Hutchins, 1952), he also identified 24 topics or specific themes, such as art, beauty and love, with variable frequencies of occurrence in the 154 sonnets: Love in its various facettes (e.g., the intensity and power of love, its increase or decrease, its constructive or destructive force, friendly, tender, or altruistic love, fraternal love, love in relation to virtue and happiness, the sacrifices of love) was by far the dominant topic occurring in more than 100 sonnets. Simonton assumed that the quantifiable features type-token ratio, number of unique words, adjective-verb quotient, broken lines and run-on lines – and perhaps thematic richness – determine a sonnet's arousal potential and thus its aesthetic value (via complexity, novelty, surprise, and other collative properties, Berlyne, 1971; Cupchik, 1986; Marin, Lampatz, Wandl, & Leder, 2016).

In a more recent quantitative narrative analysis of Shakespeare's sonnets using a novel tool called SPARSAR that allows both a broader and deeper form and content analysis than Simonton's, Delmonte (2016) challenges Simonton's claims that unique words and type-token ratio characterize better sonnets or that the most popular sonnets have a majority of concrete or primary process related concepts. Using semantic classes from WordNet (Fellbaum, 1988), Delmonte claims that the superior sonnets according to his own web-based search all contain a majority of abstract concepts, as opposed to primary process concepts (unopportunely, Delmonte's paper does not list the specific concepts).

The present study

The aim of our study was to continue the quantitative narrative analysis efforts of helping readers, critics and empirical researchers of Shakespeare's sonnets in their private or public analyses of why and how these brilliant pieces of verbal art can induce significant cognitive, affective and aesthetic responses. More particularly, we aim at providing further quantitative narrative analysis-based hypotheses and predictions for empirical studies in the emerging field of neurocognitive poetics concerning the readability, comprehensibility and affective-aesthetic potential of literary texts (Jacobs, 2015b; cf. also Burke, 2015; Nicklas & Jacobs, 2017). As argued recently by Jacobs (2015b) and Willems and Jacobs (2016), neurocognitive poetics studies using natural and ecologically valid materials – like the sonnets or poetic metaphors (Jacobs & Kinder, 2017) – can usefully inform and constrain models and theories in a number of domains, like emotion and language (e.g., Koelsch et al., 2015; Lindquist, MacCormack, & Shablack, 2015; Panksepp, 2008), emotion and literature (Miall, 1989; Oatley, 1994), affective word recognition and

reading (Bestgen, 1994; Briesemeister, Kuchinke, & Jacobs, 2014; Briesemeister, Kuchinke, Jacobs, & Braun, 2015; Jacobs et al., 2015, Jacobs, Hofmann, & Kinder, 2016b; Hofmann & Jacobs, 2014; Hsu, Jacobs, Citron, & Conrad, 2015; Kuhlmann, Hofmann, Briesemeister, & Jacobs, 2016; Lüdtke & Jacobs, 2015), empathy and mental simulation (e.g., Goldman, 2006; Oatley, 2016), immersion and transportation (Green & Brock, 2000; Hsu, Conrad, & Jacobs, 2014; Jacobs & Schrott, 2015; Ryan, 2001; Schrott & Jacobs, 2011), literary imagery (Kuzmičová, 2014), foregrounding (Miall & Kuiken, 1994; Van Peer, 1986), emotion and language development (Jacobs & Kinder, 2015; Miall & Dissanayake, 2003, Sylvester, Braun, Schmidtke, & Jacobs, 2016), self-construction and life narrativity (e.g., Habermas & de Silveira, 2008; Pleh, 2003), general aesthetics (e.g., Chatterjee & Vartanian, 2014; Jacobsen, 2006; Kintsch, 2012; Leder et al., 2004, 2015, Leder & Nadal, 2014; Marin, 2015; Pelowski, Markey, Lauring, & Leder, 2016), cultural adaptation (Hutcheon, 2012; Nicklas & Jacobs, 2017), creativity (Beaty, Benedek, Silvia, & Schacter, 2016; De Beaugrande, 1979), cognitive poetics (e.g., Stockwell, 2009; Tsur, 1998, Turner & Pöppel, 1983), or literary reading (Burke, 2011, 2015; Jacobs, 2011, 2015a,b; Schrott & Jacobs, 2011) and its effects on well-being (e.g., O'Sullivan, Davis, Billington, Gonzalez-Diaz, & Corcoran, 2015).

Our quantitative narrative analyses were based on tools like Coh-metrix (Graesser et al., 2004), TAACO (Crossley, Kyle, & McNamara, 2016), SEANCE (Crossley, Kyle, & McNamara, 2017), or the Regressive Imagery Dictionary (Martindale, 1975), as well as on a number of other measures detailed below. The paper is structured into four parts presenting cognitive (Part I) and affective-aesthetic quantitative narrative analyses of the sonnets (Part II) before discussing how the data of these analyses can be used in empirical investigations of neurocognitive poetics (Part III) and computational modeling (Part IV). Please note that the use of tools like Coh-metrix or SEANCE is tentative because (a) their components are not totally transparent as their exact factor structure is – as far as we can tell – not publically available, and (b) the hit rate, i.e. the degree to which the word lists underlying the Coh-metrix or SEANCE indices match with the words of a given analyzed text, are not provided by the software. Evidently, the higher the hit rate, the more reliable is the quantitative narrative analysis and thus providing such relevant information may push and improve future applications of such tools.

Part I. Cognitive quantitative narrative analyses: Readability and easability analyses

At the most general level, our quantitative narrative analysis results show that the vocabulary of the sonnets comprises roughly 4000 different words (3957; the exact

numbers can vary slightly according to how spelling corrections were applied for better application of the tools), of which 2480 words (about 14%) occur only once in the sonnets. Only 262 words are used 10 times or more. Important key words among the 100 most frequently occurring words are (in order of frequency): *love, beauty, sweet, eye(s), heart*, *time*, and *world*. At the poem level, the sonnet with the highest number of words (130) is the "betrayal" sonnet 42 (*That thou hast her, it is not all my grief,*.), while the "downhearted" sonnet 66 (*Tired with all these, for restful death I cry,*.) achieves its poetic effects with as little as 89 words arranged in the most repetitive fashion of all 154 sonnets. Sonnet 148 (*O me, what eyes hath Love put in my head,*.) with its many "O's" and "I's" – and which can be considered a rewrite of sonnet 137 (*Thou blind fool, Love, what dost thou to mine eyes,*.) – features the highest number of content words (86/123). In contrast, sonnet 62 (*Sin of self-love possesseth all mine eye.*) not only features an odd dramatic scenario (Vendler, 1997, chapter 62), but also has the smallest number of content words (56/107).

Put simply, two main factors determine the comprehensibility of a text: lexis and grammar (Miller, 1993). Both factors enter into the so-called *Coh-metrix L2 readability index* (CML2; McNamara, Graesser, McCarthy, & Cai, 2014) which we computed for all sonnets as a tentative simple composite measure providing a single number to assess text difficulty:

(1) CML2 readability index = −45.032 + (52.230 * Content word overlap) + (61.306 * Sentence syntax similarity) + (22.205 * Mean log minimum frequency for content words)

The universe of 154 sonnets provides approximately normally distributed data on this measure of readability allowing to identify the two sonnets with extreme CML2 values, i.e. which theoretically are hardest or easiest to read. According to the Coh-metrix analyses these are sonnets 1 (and 107) and 138, respectively.

Sonnet 1

From fairest creatures we desire increase,
That thereby beauty's rose might never die,
But as the riper should by time decease,
His tender heir might bear his memory:
But thou, contracted to thine own bright eyes,
Feed'st thy light's flame with self-substantial fuel,
Making a famine where abundance lies,
Thyself thy foe, to thy sweet self too cruel.
Thou that art now the world's fresh ornament
And only herald to the gaudy spring,
Within thine own bud buriest thy content
And, tender churl, makest waste in niggarding.
Pity the world, or else this glutton be,
To eat the world's due, by the grave and thee.

Sonnet 138

1 When my love swears that she is made of truth
2 I do believe her, though I know she lies,
3 That she might think me some untutor'd youth,
4 Unlearned in the world's false subtleties.
5 Thus vainly thinking that she thinks me young,
6 Although she knows my days are past the best,
7 Simply I credit her false speaking tongue:
8 On both sides thus is simple truth suppress'd.
9 But wherefore says she not she is unjust?
10 And wherefore say not I that I am old?
11 O, love's best habit is in seeming trust,
12 And age in love loves not to have years told:
13 Therefore I lie with her and she with me,
14 And in our faults by lies we flatter'd be.

According to Vendler (1997), sonnet 1 may have been deliberately composed late, as a "preface" or index to the others, standing out from the rest by two features: (a) its sheer abundance of values, images, and concepts important in the sequence which are called into play and (b) the number of significant words brought to our attention. Self-evidently good values and salient images enumerated by Vendler include: beauty or sweetness, and rose or famine. As evidence for her view that Shakespeare's mind works by contrastive taxonomy, Vendler cites the pairs of opposite concepts in sonnet 1: increase vs. decrease, ripening vs. dying, or immortality vs. memory. "Making an aesthetic investment in profusion" (p. 47), sonnet 1 also introduces *catachresis*, that is metaphors from incompatible categories applied to the same object (i.e., the young man as "a candle which refuses to bud forth", p. 48), which according to Vendler should vigorously call attention to itself (at least, if detected by the expert reader's mind) and, by the cognitive dissonance it produces, should press readers into reflection. Another outstanding feature Vendler mentions is the greater than norm number of speech-acts in sonnet 1, especially in the vocative quatrain 2 with its many direct addresses (e.g., thou, thyself, thy). Sonnet 138 is said to depend wholly on reported discourse and to either represent a "depraved picture of cynical partners" or a "sophisticated rendition of the way all lovers flatter each other" (Vendler, 1997, p. 138). It thus could be assimilated to an optical illusion like the Necker cube – a bistable figure –, a stylistic device often used in Petrarchan love sonnets (Schrott & Jacobs, 2011).

However, text difficulty can be assessed at both a broader and deeper level than in equation (1) by using the eight text easability principal component z-scores of Coh-metrix (cf. Graesser & McNamara, 2010; Graesser et al., 2011; McNamara, Louwerse, McCarthy, & Graesser, 2010). Figure 1 further illustrates the contrast

between the presumably easiest to read sonnet 138 vs. the most difficult sonnet 1 based on these eight scores.

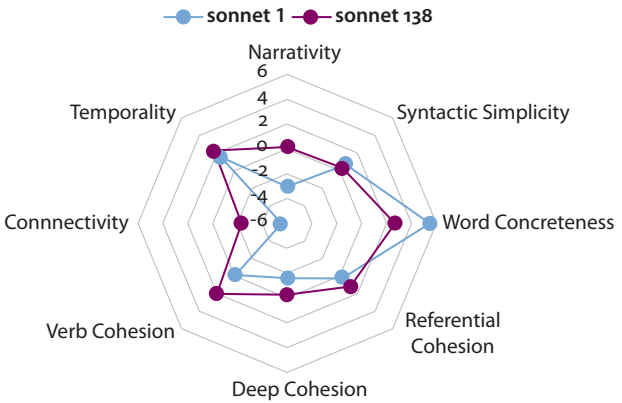


Figure 1. Eight CM easability indices (z scores) for sonnets 1 and 138

Descriptively, the differences between the two sonnets are:

1. Sonnet 138 has a higher narrativity score (1.35) and thus is theoretically closer to everyday oral conversation than sonnet 1 (−2.15), which may have some face validity (narrative text tells a story, with characters, events, places, and things that are familiar to the reader and is closely affiliated with everyday, oral conversation. The robust Narrativity component is highly affiliated with word familiarity, world knowledge, and oral language. Non-narrative texts on less familiar topics lie at the opposite end of the continuum).
2. Sonnet 138 is syntactically a bit less simple than sonnet 1 (0.03 vs. 0.41; the Syntactic simplicity component reflects the degree to which the sentences in the text contain fewer words and use simpler, familiar syntactic structures, which are less challenging to process).
3. Sonnet 138 possesses less concrete words (−0.59; texts that contain content words that are concrete, meaningful, and evoke mental images are usually easier to process and understand. Abstract words represent concepts that are difficult to represent visually. Texts that contain more abstract words are more challenging).
4. Sonnet 138 features a higher referential cohesion score than sonnet 1 (1.96 vs. −.08), i.e. it should typically be easier to process because there are more connections that tie the ideas evoked in the poem together for the reader (A text with high referential cohesion contains words and ideas that overlap across sentences and the entire text, forming explicit threads that connect the text for the reader. Low cohesion text is typically more difficult to process because there are fewer connections that tie the ideas together for the reader).

5. Sonnet's 138 higher deep cohesion score (1.65 vs. 0.48) suggests that it helps the reader to form a more coherent and deeper understanding of the causal events, processes, and actions in the poem (This score reflects the degree to which the text contains causal and intentional connectives when there are causal and logical relationships within the text. These connectives help the reader to form a more coherent and deeper understanding of the causal events, processes, and actions in the text).
6. Sonnet's 138 higher verbal cohesion score (2.01 vs. −0.11) suggests that it enhances situation model building as compared to sonnet 1 (This score reflects the degree to which there are overlapping verbs in the text. When there are repeated verbs, the text likely includes a more coherent event structure that will facilitate and enhance situation model understanding. Note that this component score is likely to be more relevant for texts intended for younger readers and for narrative texts; McNamara, Graesser, & Louwerse, 2012).
7. Sonnet 138 also has a relatively greater connectivity score than sonnet 1 (−2.33 vs. −5.56), and thus is likely to facilitate readers' deeper understanding of the relations in the poem (This component score reflects the degree to which a text contains explicit adversative, additive, and comparative connectives to express relations in the text. Thus, it reflects the number of logical relations in the text that are explicitly conveyed. This score is likely to be related to the reader's deeper understanding of the various relations in the text).
8. The slightly higher temporality score of sonnet 138 (2.30 vs. 1.55) theoretically also would facilitate its comprehension as compared to sonnet 1 (Texts that contain more cues about temporality and that have more consistent temporality like tense or aspect are easier to process and understand. In addition, temporal cohesion contributes to the reader's situation model level understanding of the events in the text).

In sum, descriptive differences in 6/8 indices are consistent in suggesting that sonnet 138 is easier to read and understand than sonnet 1. However, sonnet 1 may have a slightly simpler syntax and feature some words with a higher concreteness value than sonnet 138 which hypothetically makes it easier to understand than sonnet 138 on 2/8 dimensions. Lacking any inference statistics, these descriptive analyses allow no conclusions but serve an illustrative and heuristic, hypothesis-generating purpose demonstrating how Coh-metrix can be used to compare the readability of two or more poems at a more sophisticated level than the traditional readability scores.

Even though Coh-metrix was designed to primarily analyze longer text book materials rather than short poems, it already was succesfully applied to the language of Shakespeare (Graesser et al., 2011) and – given the paucity of specialized quantitative narrative analysis alternatives for the structural description of poetry

(Jacobs, 2015b) – the data in Figure 1 can be submitted to empirical testing in rating, eye movement or neuroimaging studies (see Part III).

Figure 2 shows the distributions of the five most commonly used standardized easability scores for all 154 sonnets (i.e., Narrativity, Syntactic simplicity, Word concreteness, and Referential and Deep cohesion; Graesser et al., 2011). The approximately normally distributed data (two distributions are normal, three fail the

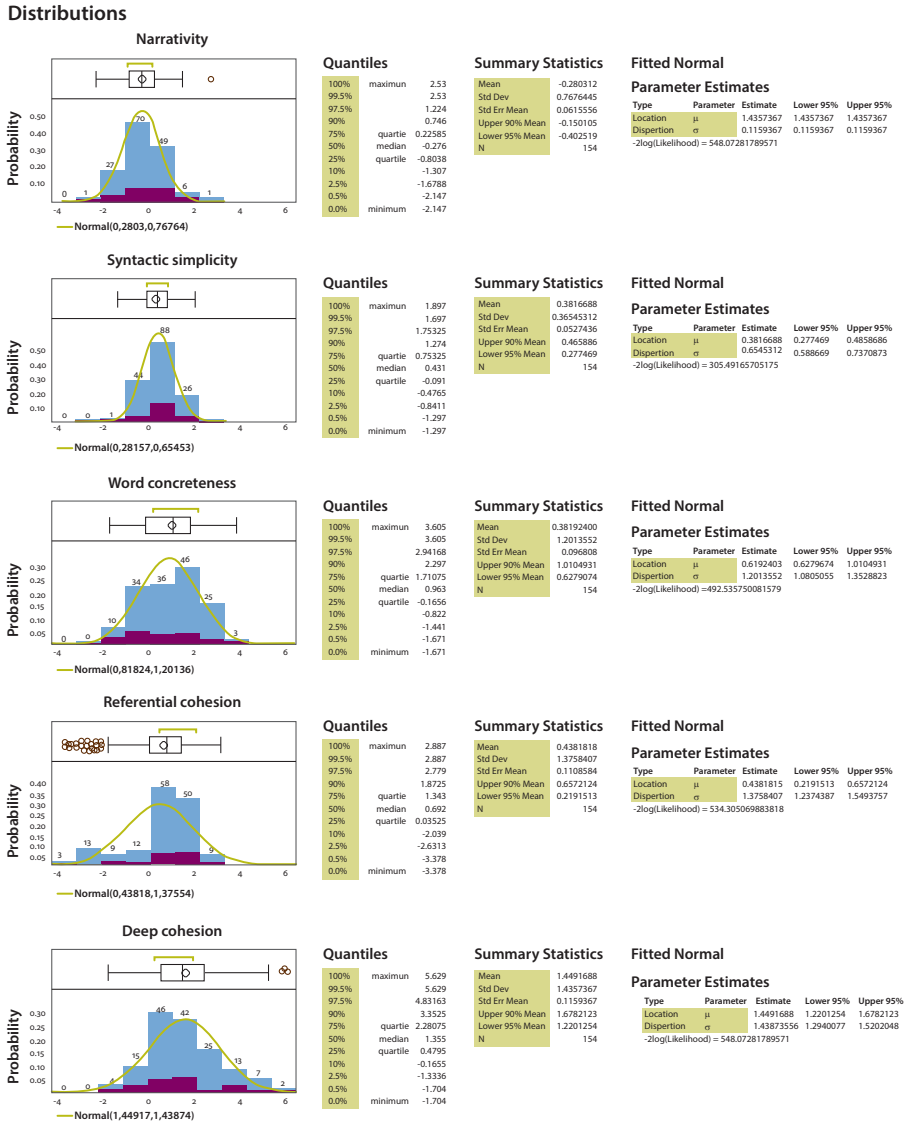


Figure 2. Distribution of five CM easability indices (z scores) for all sonnets (magenta area = "dark lady" sonnets: 127–154)

W-test) facilitate the use of these scores in statistical effect analyses such as linear mixed or regression models and demonstrate the potential of this sonnet corpus for empirical studies in neurocognitive poetics (Jacobs, 2015b).

Table 1 gives an overview of the three sonnets theoretically easiest vs. hardest to process, respectively, for each of the five dimensions. According to Table 1 then, sonnet 138 mainly is easier to read because it has a considerably higher narrativity score than sonnet 1.

Table 1. Three easiest and hardest to read sonnets according to five coh-metrix easability scores

	Narrativity	Syntactic simplicity	Word concreteness	Referential cohesion	Deep cohesion
easiest	42, 138, 149	145, 125, 10	63, 153, 20	47, 134, 136	51, 52, 22
hardest	1, 77, 95	80, 67, 73	105, 90, 115	125, 65, 85	31, 144, 35

Note: The easiest and hardest sonnet, respectively, is mentioned first.

Surprisal

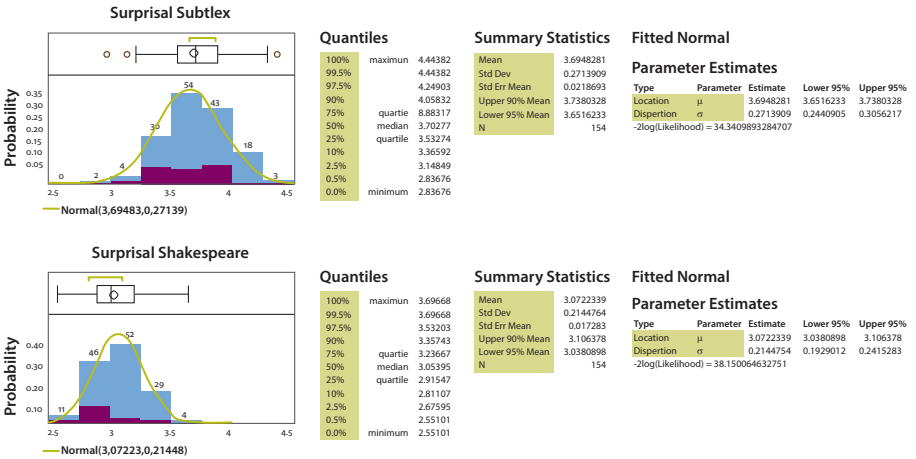
Surprisal, the most common quantification of words' information content (Frank, 2013) is known to be a co-determinant of reading speed and eye movement parameters correlating positively with reading time (e.g., Boston, Hale, Kliegl, Patil, & Vasishth, 2008; Frank, 2013; Smith & Levy, 2013). Moreover, the amplitude of the N400 event-related potential (ERP) component was found to correlate with word surprisal values (Frank, Leun, Giulia, & Vigliocco, 2015). When words come unexpected to the reader – which is part of the attraction of poetry – their surprisal value is higher than when they can be anticipated by context and skilled knowledge of lexis and grammar. For each of the 154 sonnets we computed two indices of their surprisal value.[1]

Figure 3a shows the distributions of these two different, normally distributed mean surprisal values for the sonnet corpus: The upper panel shows surprisal values based on the Subtlex database (Brysbaert & New, 2009), the lower panel shows surprisal values based on a Shakespeare corpus assembled from "The Complete Works of William Shakespeare", n.d.

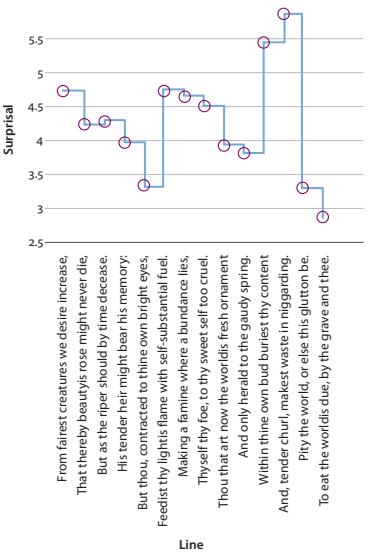
1. The surprisal values were estimated by means of a corpus-based trigram model – the values for each word being computed by the SRILM package (see Willems et al. 2015): (1) a corpus consisting of the works of Shakespeare (http://shakespeare.mit.edu/) excluding his sonnets (encompassing 1433958 sentences), and (2) a contemporary corpus of spoken sentences (SUBTLEX; encompassing 6043188 sentences). The trigram model already was successfully applied to experimental data (EEG: Frank et al. 2015; reading time: Smith & Levy 2013).

The context effect of Shakespeare's verbal art can easily be seen in the difference between the two means (3.7 vs. 3.1): Not surprisingly, Shakespeare's words are notably more surprising when matched against a modern database than when taxed within their own verbal neighborhood.

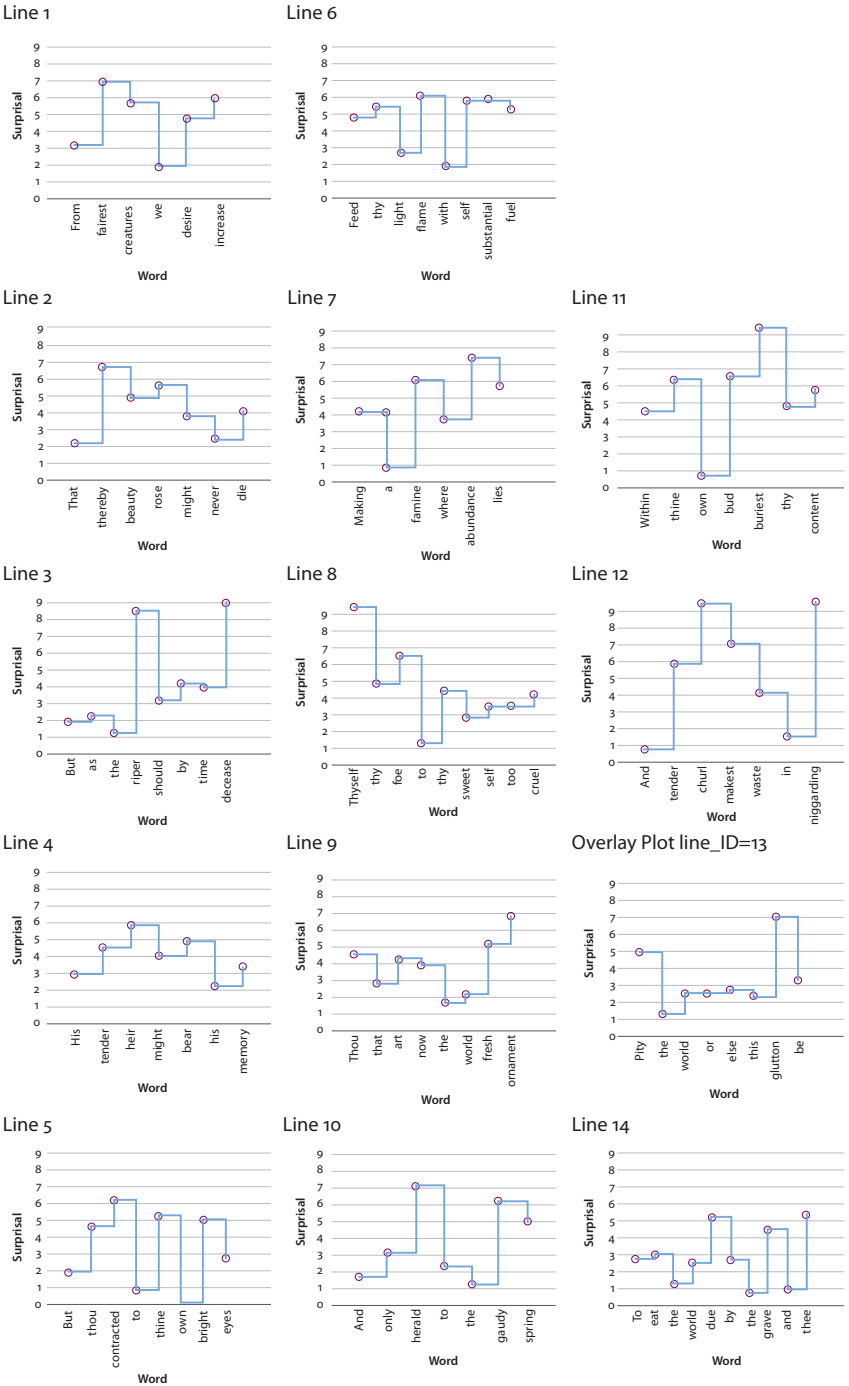
Distributions



a. Distribution of two Surprisal indices (means) for all sonnets (Magenta area = "dark lady" sonnets: 127–154)



b. Linewise Surprisal for sonnet 1



c. Wordwise Surprisal for sonnet 1

Figure 3a–c.

It was interesting to see whether our Shakespeare corpus surprisal measure correlates with coh-metrix's readability and easability scores: All correlations were significant but one (Deep cohesion). Thus, CML2 readability, Narrativity and Referential cohesion all significantly decreased with increasing Surprisal: $F(1, 152) = 63.42$, $p < .0001$, $R^2 = .29$; $F(1, 152) = 42.87$, $p < .0001$, $R^2 = .22$; $F(1, 152) = 49.77$, $p < .0001$, $R^2 = .25$, respectively. Interestingly, Syntactic simplicity and Word concreteness increased with increasing Surprisal: $F(1, 152) = 11.52$, $p < .0009$, $R^2 = .07$; $F(1, 152) = 15.31$, $p < .0001$, $R^2 = .091$, respectively. The higher the surprisal value of a sonnet, the less easy to read, the less "narrative" and the less (referentially) "coherent" it seems. However, at least for this corpus, sonnets with a higher surprisal also tend to feature a simpler syntax and more concrete words (or words with high concreteness values). Even if the effects were small, this finding seems interesting material for further research: Perhaps, in some cases the poet chose to trade-off a critical amount of *poetic surprisal* for fewer or less abstract words and simpler grammar in order not to make the poem too hard to comprehend. Examples for poems that are high on both Syntactic simplicity and Surprisal are sonnets 60, 66 and 125. Examples for poems that are high on both Word concreteness and Surprisal are sonnets 66, 153 and 154.

The surprisal value of 4.25 (Subtlex) for the theoretically hardest to read sonnet 1 is significantly higher than that of sonnet 138 (3.4; $F(1, 26) = 8.9$, $p < .006$, $R^2 = .26$) thus confirming the results summarized in Figure 4. Overall, the line with highest surprisal is line 12 in sonnet 1 (5.85: *And, tender churl, makest waste in niggarding*), the one with the lowest is line two from sonnet 138 (2.43: *I do believe her, though I know she lies*). Figure 3b and c zoom into the individual lines and words of sonnet 1 to reveal their line- and wordwise surprisal values for even more fine-grained hypotheses, e.g., concerning eye movement or ERP parameters (see Part III).

Before we take an "emotional turn" in our analyses, a short summary of the cognitive quantitative narrative analyses of Part I seems in order. Thanks to their approximately normally distributed features describing the readability, comprehensibility and surprisal, the 154 sonnets offer a rich playground for generating and testing hypotheses about reader responses in neurocognitive poetics studies (Jacobs, 2015b) at different levels of inquiry: the metalevel of poem category, i.e. young man vs. dark lady poems, the poem level, e.g. poems with low vs. high comprehensibility, and even the line- or wordwise levels, e.g., lines/words with low vs. high surprisal. Together with the abundance of qualitative (or quasi-quantitative) content analyses and hermeneutic interpretations of Shakespeare's sonnets by critics and scholars (e.g., De Beaugrande, 1979; Jakobson & Jones, 1970; Vendler, 1997) the present quantitative narrative analysis results should motivate a series of empirical studies using combined qualitative-quantitative, multimethod,

multilevel designs (Jacobs et al., 2016a) providing new insights into the complexities of "brain and poetry" (Schrott & Jacobs, 2011) by allowing to (i) better manage (i.e., manipulate, match or control) a great number of potentially relevant stimulus variables for more natural and ecologically valid experiments (Willems, 2015; Willems & Jacobs, 2016); ii) disentangle effects of surface and form vs. deep structure and content features, or iii) cognitive vs. affective variables (e.g., Jacobs et al., 2016a,b; Menninghaus et al., 2014, 2015).

Complementing the previous one, Part II looks at quantifiable affective-aesthetic variables of sonnets that theoretically and empirically are at least as relevant for reader responses to poetry as the cognitive ones (e.g., Jacobs, 2015a,b, 2016a,b; Lüdtke et al., 2014; Schrott & Jacobs, 2011).

Part II. Affective-aesthetic quantitative narrative analyses

Emotion and mood potential

The recently growing interest in emotional word and text processing (see Citron, 2012; Jacobs et al., 2015, 2016a, for review) has been made possible by databases allowing to quantitatively estimate affective word features, such as the Berlin Affective Word List (BAWL, Briesemeister et al., 2011; Võ et al., 2006, 2009; Jacobs et al., 2015), the Affective Norms for English Words (ANEW; Bradley & Lang, 1999), the Affective Norms for German Sentiment Terms (ANGST; Schmidtke, Schröder, Jacobs, & Conrad, 2014), or by computational algorithms (Westbury, Keith, Briesemeister, Hofmann, & Jacobs et al., 2014). Several recent studies demonstrate how such tools can be used to predict and interpret reader responses to poetry (e.g., Aryani, Kraxenberger, Ullrich, Jacobs, & Conrad, 2016; Jacobs et al., 2016b; Ullrich, Aryani, Kraxenberger, Jacobs, & Conrad, 2016) and prose (Altmann, Bohrn, Lubrich, Menninghaus, & Jacobs., 2012, 2014; Hsu et al., 2015), which are theoretically predicted by the *Neurocognitive Poetics Model* of literary reading (Jacobs, 2011, 2015a,b).

A sophisticated comprehensive tool for English texts that complements cognitive quantitative narrative analysis tools like Coh-metrix is SEANCE (Crossley et al., 2017). Here we computed the 20 SEANCE component scores for each sonnet and summed them to determine the sonnets with the theoretically highest *emotion potential*. Taking the sum of all 20 components is only a first step using the simplest possible and most general explorative model supposedly covering the greatest amount of words (i.e., the maximum hit rate; see Introduction). Before testing more specific models such as, say, the action component, a theoretical motivation for the selection should be proposed, such as for the mood scores computed below (Figure 4 and Table 3). Moreover, detailed data on the hit rates of each of the indices underlying a component like Action, e.g., General Inquirer ought verbs,

try verbs, or descriptive action verbs (Stone et al., 1962) would motivate more specific tests. The 20 scores summed here for an index of the Emotion potential are: Negative adjectives, Social order, Action, Positive adjectives, Joy, Affect for friends and family, Fear and disgust, Politeness, Polarity nouns, Polarity verbs, Virtue adverbs, Positive nouns, Respect, Trust verbs, Failure, Well-being, Economy, Certainty, Positive verbs, and Objects (for details, see Crossley et al., 2017).

The three sonnets with the highest and lowest Emotion potentials according to our SEANCE composite score, respectively, are: 140, 151, 144 and 3, 90, 53.

Sonnet 140

1 Be wise as thou art cruel; do not press
2 My tongue-tied patience with too much disdain;
3 Lest sorrow lend me words and words express
4 The manner of my pity-wanting pain.
5 If I might teach thee wit, better it were,
6 Though not to love, yet, love, to tell me so;
7 As testy sick men, when their deaths be near,
8 No news but health from their physicians know;
9 For if I should despair, I should grow mad,
10 And in my madness might speak ill of thee:
11 Now this ill-wresting world is grown so bad,
12 Mad slanderers by mad ears believed be,
13 That I may not be so, nor thou belied,
14 Bear thine eyes straight, though thy proud heart go wide.

Sonnet 3

1 Look in thy glass, and tell the face thou viewest
2 Now is the time that face should form another;
3 Whose fresh repair if now thou not renewest,
4 Thou dost beguile the world, unbless some mother.
5 For where is she so fair whose unear'd womb
6 Disdains the tillage of thy husbandry?
7 Or who is he so fond will be the tomb
8 Of his self-love, to stop posterity?
9 Thou art thy mother's glass, and she in thee
10 Calls back the lovely April of her prime:
11 So thou through windows of thine age shall see
12 Despite of wrinkles this thy golden time.
13 But if thou live, remember'd not to be,
14 Die single, and thine image dies with thee.

Sonnet 140 indeed features many emotion-laden words like *cruel, pity* or *despair* and in quatrain 3 a "pathological picture of the world in which both speaker and audience are conceded to be mad" (Vendler, 1997, chapter 140) is drawn. The three top emotion potential sonnets all belong to the dark lady category.

As a first cross-validation check, we computed the correlation across all 154 sonnets between the simple emotion potential measure proposed in Jacobs (2015b), i.e. the product between the absolute mean values for valence and arousal of each word in a text (as computed from the database of Warriner, Kuperman, & Brysbaert, 2013) and the SEANCE composite score. The correlation was small but significant: $F(1, 152) = 14.75$, $p < .0002$, $R^2 = .09$. In a second cross-validation check, we computed the correlation of the SEANCE score with Martindale's Regressive Imagery Dictionary measure Emotion (Martindale, 1975; cf. Simonton, 1989). The correlation was smaller than for the previous measure but still significant: $F(1, 152) = 4.9$, $p < .028$, $R^2 = .03$.

Given Simonton's claim that sonnets with superior aesthetic success feature more primary process imagery, in Table 2 we give an overview of the three sonnets with the highest and lowest values, respectively, for each of the three Regressive Imagery Dictionary indices and their sum total. Interestingly, the theoretically easiest-to-read sonnet 138 also features a high secondary process score.

Table 2. Sonnets with lowest/highest regressive imagery dictionary values

	Primary process	**Secondary process**	**Emotion**	**Sum Regressive Imagery Dictionary**
highest	153, 154, 73	57, 49, 138	40, 25, 30	153, 2, 43
lowest	36, 149, 4	99, 7, 94	12, 81, 43	74, 134, 94

Note: The sonnet with the highest and lowest value is mentioned first, respectively.

Three indices can be used in a straightforward way to predict the potential of a sonnet to induce either a positive or negative mood (Aryani et al., 2016): Mean word valence, Valence span, and Word valence sum. Since these measures are only moderately correlated in our data (all $r < .44$), they can lead to divergent predictions. In a second analysis, we computed two component scores of SEANCE to estimate the potential of a sonnet to induce either a negative or positive mood: A negative mood score was obtained by summing up the scores for components one and seven (Negative adjectives and Fear and disgust), and a positive mood score by summing up the scores for components Four, Five, 12 and 19 (Positive adjectives, Positive nouns, Positive verbs and Joy). Figure 4 shows the distribution of these five indices.

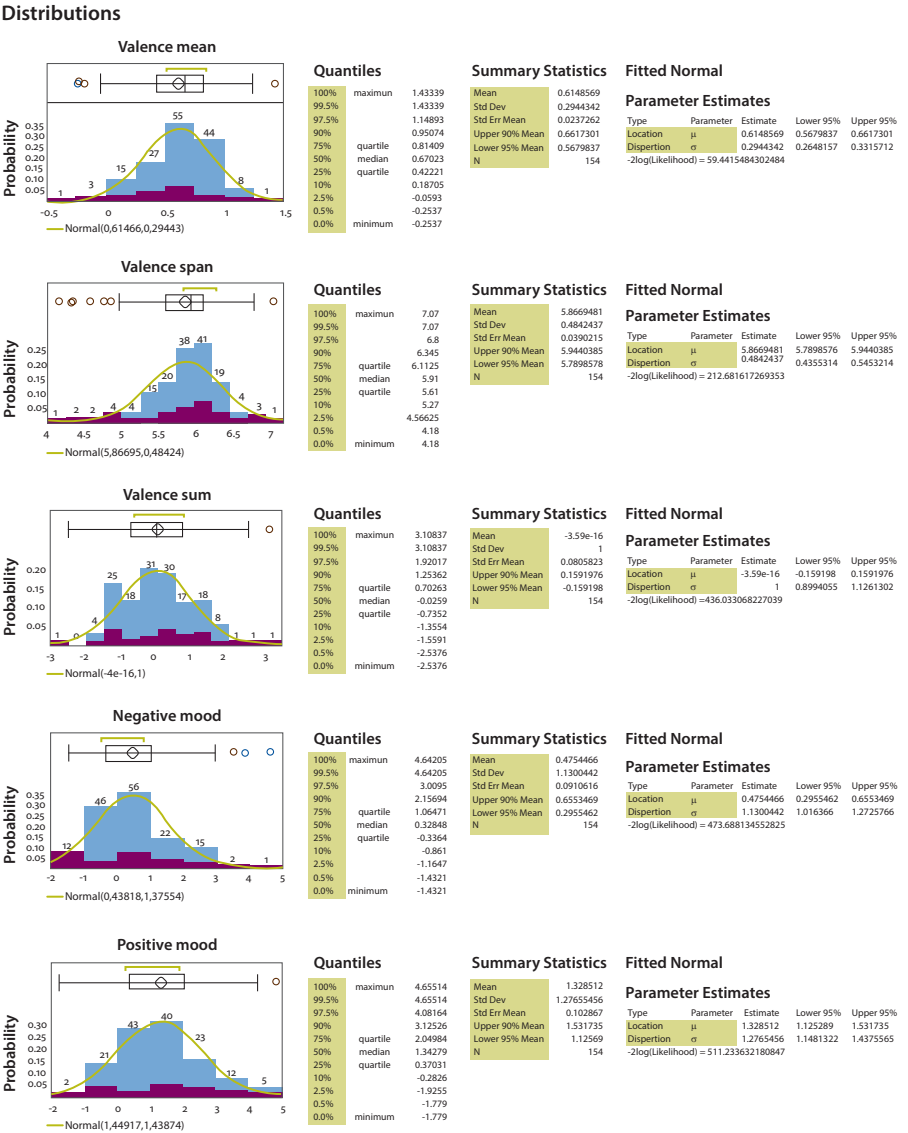


Figure 4. Distribution of five "mood" indices for all sonnets (magenta area = "dark lady" sonnets: 127–154; see text for details)

Table 3 gives an overview of the three sonnets with the highest and lowest values, respectively, for each of the five mood indices. The results provide a heterogeneous picture calling for empirical investigation, since each index makes different predictions as to which three sonnets induce a positive vs. negative mood. As has been shown empirically by Lüdtke et al. (2014), and Jacobs et al. (2016a), also other factors than these five indices play a role in mood induction through poetry,

but the results of Aryani et al. (2016) suggest that whether a poem is rated as "sad" or "friendly" is clearly affected by word valence. To what extent such ratings reflect the perception (in the poem) and/or genuine feeling of a sad vs. joyful mood is an open issue for future research that can use the quantitative narrative analysis data produced in this paper. In music psychology there is a debate between a cognitivist and emotivist position accounting for the "sad music paradoxon," i.e. the phenomenon that people like sad music (Taruffi, 2016). The first position states that people do not experience genuine sadness at all, but merely recognize the sadness depicted by the music (e.g., Kivy, 1991), while the second claims that sad music induces an emotion similar to "real" sadness, although it is not clear to what extent they overlap (Levinson, 1997). Regarding sad (vs. joyful) poetry one can argue in a similar vein. The observation that similarly to music (Krumhansl, 1997) poetry also shows measurable peripheral-physiological effects (Jacobs et al., 2016a) can be taken to suggest that it evokes "real" feelings at least to some extent.

We checked whether the young man sonnets differed significantly from the dark lady sonnets in their positive vs. negative mood potential. There was no difference for the former, but the latter indeed was significantly greater for the dark lady sonnets: 1.1 > 0.33, $F(1, 152) = 11.15$, $p < .001$, $R^2 = .07$. Thus whether *perceived* and/or *felt* mood, in an empirical study the dark lady sonnets should produce higher response measures of negative mood than the young man sonnets.

Table 3. Sonnets with lowest/highest scores for various estimates of mood potential

	Valence mean	Valence span	Valence sum	SEANCE negative	SEANCE positive
highest	153, 154, 73	57, 49, 138	40, 25, 30	129, 140, 120	128, 26, 136
lowest	36, 149, 4	99, 7, 94	12, 81, 43	74, 134, 94	44, 90, 133

Note: The sonnet with the highest and lowest value is mentioned first, respectively.

Thematic richness

As outlined above, thematic richness is one of Simonton's (1989) key sonnet features for superior aesthetic success. Moreover, recent empirical research on poetry reception supports the notion that the motif or topic of a poem is important for reader responses (e.g., Lüdtke et al., 2014; Jacobs et al., 2016a). As stated by the latter authors (p. 97): "Knowing, inferring, or guessing the overall motif of a poem might therefore be especially important – as a kind of orienting *metameaning* active in working memory – for interpreting hidden multiple meanings and unexpected meaning twists typical for abstract or obscure poetic texts (Shimron, 1980; Yaron, 2002, 2008)", or for the couplet at the end of Shakespearean sonnets. Here we used Simonton's (1989) 24 different topics to compute a Thematic richness index (= sum of all of topics per sonnet).

According to this analysis, most sonnets (68) have two motifs or topics, combining, e.g., love and poetry. Nineteen sonnets highlight none of the topics listed by Simonton, whereas six sonnets (14, 15, 25, 65, 76, and 82) play with as many as six different motifs. Thus, in Simonton's terms (1989, Table 1, p. 704) sonnet 14 features the topics (i) beauty 1b, i.e. beauty and truth, the beautiful as an object of contemplation; (ii) family 6a, i.e. the desire for offspring; (iii) immortality 6a, i.e. immortality through offspring, the perpetuation of the species; (iv) love 1e, i.e. the intensity and power of love, its increase or decrease, its constructive or destructive force; (v) love 2b, i.e. friendly, tender, or altruistic love; fraternal love, and (vi) time 7, i.e. the temporal course of the passions emotional attitudes toward time and mutability.

Which of the other variables of the present analyses significantly correlate with the thematic richness index? Among the more than 40 variables, for which this was the case ($p < .05$), for 17 at least 5% of the variance was accounted for by the Thematic richness index. Among those, the most relevant for the present purposes were: Mean and summed word valence, Number of positive words, and Positive noun component (all SEANCE) which all correlated positively with the thematic richness index, and Number of negative words, Negative adjectives, and Negative mood score which correlated negatively (all $R^2 > .05$). These results allow to state the following theoretical claim to be empirically tested in future studies: The higher a sonnet's thematic richness index, the higher the likelihood that readers will like it and rate it as perceiving/inducing a positive mood. Thus, perhaps, the aesthetic success of sonnets is mediated by their mood perception/induction potential?

Symbolic imagery

The sonnets contain a wealth of recurrent images, archetypes, archetypal patterns and personal myths "through which the imaginary of the writer and that of the reader bind, generating meaning" (Meireles, 2005, p. 5). Music (sound) and painting (imagery) have been characterized as perhaps the most distinctive features of poetry (Schrott & Jacobs, 2011), but little is known about which kind of imagery prevails in poetry, what its neural correlates are, and how it can reliably and validly be measured (Jacobs, 2016a). Determining the type and occurrence of imagery in poetry thus may constitute a first step towards tackling these issues. Based on the work of Meireles, here we computed a typological Symbolic imagery index by coding each sonnet for the occurrence of the following eight types of recurrent symbolic/archetypical images: *Time* (i.e., words expressing symbols for time like *Clock, Moment* or *hours*), *Solar* (i.e., images conveying solar symbols as seen in words like *day, sun*, or *stars*), *Water* (e.g., words like *liquid, tears*), *Nocturnal* (e.g., *darkness, moon*), *Season* (e.g., *summer*, *april*APRIL), *Nature* (i.e., only the word *nature* itself), *Immortality* (e.g., *eternal, soul, body*), and *Color* (e.g., *black, scarlet*). Much as for the thematic richness index analyses, we use this list in a heuristic

fashion – without any claims regarding its completeness, validity or poetic effectiveness – for purposes of comparison and cross-validation with the other tools used here (for a discussion of the symbolic imagery in the sonnets see Vendler, 1997, or Meireles, 2005).

For all sonnets, we computed the Symbolic imagery index as the sum of the eight indices listed above. The results indicate that 15 sonnets have a Symbolic imagery index of 0 being void of the symbols in our list. At the other extreme, sonnet 65 features 75% (6/8) of the eight symbols (Time, Solar, Water, Season, Immortality, Color) and 23 sonnets have more than two. The great majority (105), however, focuses on one or two symbols or archetypes according to Meireles' (2005) typology.

Sonnet 65

1 Since brass, nor stone, nor earth, nor boundless sea,
2 But sad mortality o'er-sways their power,
3 How with this rage shall beauty hold a plea,
4 Whose action is no stronger than a flower?
5 O, how shall summer's honey breath hold out
6 Against the wreckful siege of battering days,
7 When rocks impregnable are not so stout,
8 Nor gates of steel so strong, but Time decays?
9 fearful meditation! where, alack,
10 Shall Time's best jewel from Time's chest lie hid?
11 Or what strong hand can hold his swift foot back?
12 Or who his spoil of beauty can forbid?
13 O, none, unless this miracle have might,
14 That in black ink my love may still shine bright.

Which variables of the quantitative narrative analysis tools used here correlate with the symbolic imagery index? This was the case for seven variables. The strongest significant (positive) correlation was found for the primary process score based on Martindale's Regressive Imagery Dictionary (1975): $F(1, 152) = 22.93$, $p < .0001$, $R^2 = .13$, thus cross-validating Meireles' typology. The other significantly correlated variables that accounted for at least 5% of variance in the symbolic imagery index were (in order of R^2): Coh-metrix's Word concreteness score (positive, $R^2 = .12$), TAACO Content types (positive, $R^2 = .085$), Summed word valence (positive, $R^2 = .083$), TAACO Type-token ratio and Repeated content lemmas and pronouns (both positive, $R^2 = .06$), and the Thematic richness index (positive, $R^2 = .06$).

To summarize, sonnets that are rich in symbolic imagery like 55, 12, 14, 18, 27, 56, 61, 63, 65, 68, or 98 should feature more concrete and unique content words, as well as words associated with oral or sexual needs like *bread* or *lust* with *sensations* (e.g., *sharp*), *defensive symbols* (e.g., *pilgram*), *regressive knowledge* (e.g., *secret*), or

icarian imagery (e.g., *valley*) according to the Regressive Imagery Dictionary. They also should have an overall more positive valence, supported by words expressing *joy, anticipation* or *surprise* like *happy* or *magical*, a greater *lexical diversity* and more repetitions of content lemmas and pronouns. In line with this, their thematic richness index should be higher. On the other hand, they should feature less adjectives expressing fear or disgust, less emotion words (e.g., *afraid, harsh*), and less words affiliated with everyday, oral conversation (due to small but significant negative correlations with SEANCE's fear and disgust component, Regressive Imagery Dictionary's emotion score, and Coh-metrix's Narrativity score).

Semantic association potential

The number of semantic associates of a word is a factor that has various effects on both behavioral and neuronal measures in word and text processing, as well as in memory tasks (see Hofmann & Jacobs, 2014, for review). For example, recent computational and neurocognitive studies suggest that the affective evaluation of words and texts is co-determined by their semantic associations (Kuhlmann et al., 2016; Hofmann & Jacobs, 2014; Recchia & Louwerse, 2015; Westbury et al., 2014). Here we were interested in the variance across the 154 sonnets concerning their semantic association potential.

Our Semantic association potential index is based on the unique number of word associations according to the *Edinburgh Associative Thesaurus* (Kiss et al., 1973) – an index of the unique number of associations produced by 100 participants in a free association task –, summed across all words of a sonnet. The sonnets had a mean Semantic association potential of 4004 and the three sonnets with the highest semantic association potential were: 137, 69, and 43 (all > 4524). With a Semantic association potential of 2937, sonnet 66 was at the low end of the distribution. As expected from the above cited studies, Semantic association potential correlated significantly with the valence (sum) of the sonnets: $F(1, 152) = 11.6$, $p < .0008$, $R^2 = .07$, allowing the tentative hypothesis that sonnets rich in Semantic association potential will produce higher liking ratings. Interestingly, Semantic association potential also significantly correlated (negatively) with both Surprisal indices (Subtlex: $F[1, 152] = 74.6$, $p < .0001$, $R^2 = .33$; Shakespeare: $F[1, 152] = 65.5$, $p < .0001$, $R^2 = .30$) – indicating that a higher amount of different Semantic associations predicts lower Surprisal – and with the CML2 readability index ($F[1, 152] = 34.6$, $p < .0001$, $R^2 = .18$), indicating a positive relationship between the number of different associations and readability, at least in Shakespeare's sonnets.

Overall, both the cognitive and affective-aesthetic indices discussed above appear sensitive enough to be used for generating and testing hypotheses concerning cognitive and emotional reader responses in neurocognitive poetics studies on sonnet reception. Apart from the reported correlations, cognitive and affective

indices were only marginally related to each other. All of the reported indices can be found in Table A1 and the Supplementary Materials to test further theoretically motivated hypotheses. Next we deal with the issue whether together with the cognitive indices the affective-aesthetic indices also are sensitive to more or less subtle changes in form and content across different sonnet parts.

Can quantitative narrative analysis capture sonnet dynamics?

In the Introduction, sonnets were described as *systems in motion* with a thematic-semantic narrowing down movement from quatrain 1 to 4 and the couplet (Vendler, 1997). Moreover, according to Vendler, the sequence of images, for example, should have a notable effect on its interpretation. Simonton (1990) also found effects of sonnet part by showing that "as we ascend from the mediocre sonnets to those that have likely earned a permanent position in literary history, the probability of encountering a unique word in either the third quatrain or the final couplet decreases" (p. 261).

Here we wanted to see whether the present quantitative narrative analysis tools also can detect traces of such dynamics, e.g. can tools like Coh-metrix, the Regressive Imagery Dictionary, or SEANCE capture aspects that reveal a thematic diminution from quatrain 1 to the couplet, or changes in comprehensibility or emotion potential from the octave to the sestet? In doing so, we computed the Coh-Metrix features and factors separately for each line and aggregated them across the different sonnet parts. For word-based indices like Surprisal we calculated the average of all words belonging to the specific sonnet part; for the semantic associations we summed up the number of unique associations and divided this sum by the number of words. Table 4 summarizes the results of our analyses. Regarding composite indices of ease of comprehension, we found significant differences between the four parts of a sonnet for the CML2 readability index and several of the five Coh-metrix easability indices. Using Sonnet part as the independent variable in several one-way ANOVAs the following picture emerged: The final couplets had significantly higher CML2 readability scores than the quatrains which did not differ from each other (means: 6.7 vs. 1.4, 1.0, 2.4, respectively). The couplets also had significantly higher Narrativity, Syntactic simplicity, Referential and Deep cohesion but lower Surprisal scores and a lower number of different semantic associations than the body parts of the sonnet. The only other significant effect of sonnet part on Coh-metrix easability scores was that, on average, quatrains 2 had a higher Syntactic simplicity than quatrains 1. Quatrain 2 had a higher Surprisal index than quatrains 1 and 3.

Table 4. Various indices for sonnet parts

	Quatrain 1	Quatrain 2	Quatrain 3	Couplet	octave	sestet
CML2	1.4	1.0	2.4	6.7*	1.2	3.8*
Narrativity	– 1.6	– 1.6	– 1.4	– 0.98*	– 1.6	– 1.3*
Syntactic simplicity	0.52	0.77*	0.70	0.83*	–	–
Referential cohesion	0.65	0.69	0.66	0.77*	–	–
Deep cohesion	– 1.06	– 0.73	– 0.6	0.84*	– 0.9	– 0.12*
Surprisal (Shakespeare)	3.07	3.17*	3.08*	2.92*	3.11	3.02*
Semantic Association	38.28	38.24	37.88	37.36*	38.05	37.9
Emotion potential	5.84	5.71	5.87	6.64*	5.8	6.1*
Positive verbs	–	–	–	–	– 0.05	– 0.01*
Affect for friends & family	–	–	–	–	0.4	0.45*

Note: We only report variables and mean values, if one of the comparisons reached a *p*-value of at least <.05. (*)

Regarding the *octave-sestet contrast*, sestets were systematically easier to read than octaves (means: 3.8 vs. 1.2, respectively) and had significantly higher Narrativity and Deep cohesion scores but lower surprisal values. Complementing this cognitive quantitative narrative analysis by an affective one using the Regressive Imagery Dictionary, the simple emotion potential measure mentioned above calculated as the product of word valence and arousal ratings (Jacobs, 2015b), and the 20 SEANCE components (calculated according to the procedure described above), the only significant effects we observed were that on average the couplet has a higher Emotional potential than the body parts, and that the sestets had an overall higher Emotion potential, Affect for friends and family and Positive verbs score than the octaves.

A final analysis looked at the assumption mentioned in the Introduction that the number of new text world referents decreases towards the end of a text. As evidenced by Figure 5 this was definitely the case: $F(1, 2152) = 516.33$, $p < .0001$, $R^2 = .19$. While quatrains 1 feature 87% new words on average, this value drops to 76% for quatrains 2 and 68% for quatrains 3 reaching a plateau for couplets with 59% (all $p < .0001$). This quantitative narrative analysis discovery sheds new light on Vendler's (1997) "funnel-shape" movement assumption mentioned in the Introduction, supports Steen's (2004) view of a narrowing down of the reader's text world, and thus has interesting implications for future empirical studies, such as the hypothesis derived from Steen (2004), that style figures (e.g., metaphors) should be easier to recognize towards the end of poems.

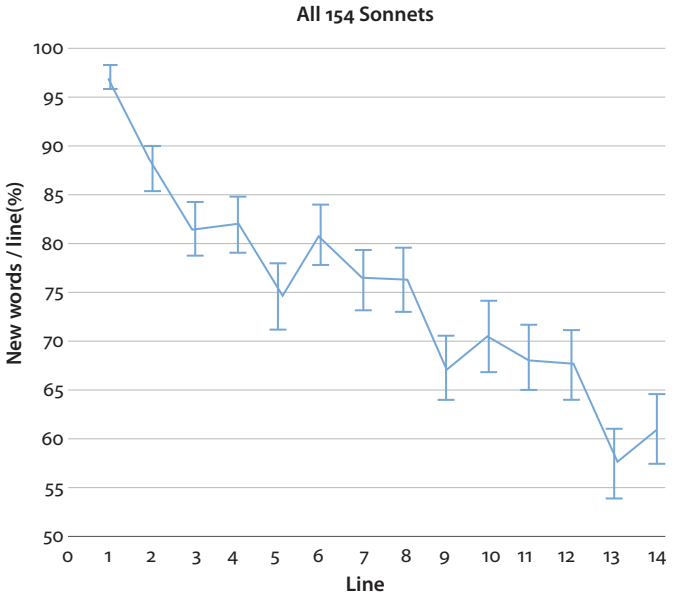


Figure 5. Percentage of new words (means with confidence intervals) as a function of position (line) for all sonnets.

In sum, the cognitive and affective indices considered here are sensitive – albeit to different degrees – to sequential changes in form and content across sonnet parts. The present results thus motivate more work testing Vendler's (1997), Steen's (2004) or further assumptions that will now be discussed in Part III.

Part III. Hypotheses for neurocognitive poetics studies

What can the present quantitative narrative analyses be used for?

What good is all the effort spent in applying quantitative narrative analysis tools to sonnets? Despite his extensive analyses including phonetic, poetic and syntactic-semantic relational levels in poems, Delmonte (2016, p. 93) concludes his work on a challenging note: "From the data reported above, it is hard to understand what criteria would be best choice for the individuation of most popular sonnets. It seems clear, however, that neither themes nor readability indices are sufficient by themselves to identify them all. Nor do evaluations based on semantic/pragmatic criteria derived from existing lexica help in the final classification. We surmise that an evaluation of how much popular a poem can be should also take into account cultural issues which have not been tackled by this study […]. In particular, the contribution

of rhetoric devices, like similes and metaphors, is hard to compute consistently for all sonnets: Shakespeare's best virtue was his subtlety in generating a great quantity of secondary meanings from simple juxtaposition of terms and images. So eventually, what SPARSAR can do is help practitioners in that direction without giving a final complete result, but leave the user to combine different schemes, graphs, tables and other data together in the puzzle constituted by poetry that aims at excellence and lasts forever, like the one we have been commenting in this article".

In a somewhat more optimistic vein, Simonton (1989, p. 703) advanced that "Our understanding of artistic creativity would be enlarged if we knew which of these four alternative measures optimally predicted aesthetic success." Similarly optimistic, Graesser et al. (2011) conclude their quantitative narrative analysis of three dramas by Shakespeare using the Coh-metrix tool with: "In closing, we believe that there is so much to be learned from computer analyses of literature. Computers may never understand and fully appreciate Shakespeare. But humans don't either. Meanwhile we can learn from computer analyses just as we learn from the insights of literary scholars. A computational science of literature is a worthy player in the interdisciplinary arena" (p. 31).

We leave it to interested readers to form their own opinion and meanwhile propose a few potentially useful applications in the following sections. For us, a first straightforward use of the present quantitative narrative analysis results is in empirical studies on neurocognitive poetics that require quantitative variables for their stimulus selection and/or statistical data analyses testing specific predictions. As argued in Jacobs (2015b), the dynamically developing but still very recent field of neurocognitive poetics needs extended and refined text-analytical tools. These are necessary for both model development and for inspiring experimental designs that use more natural and ecologically valid stimuli and tasks, as well as a combination of direct/indirect and online/offline measures aiming at a higher overall validity. All these are part and parcel of the neurocognitive poetics perspective (Bohrn et al., 2012a,b, 2013; Chen et al., 2016; Dixon & Bortolussi, 2015; Jacobs, 2011, 2015a,b,c, 2016a; Lehne et al., 2015; Liu et al., 2015; O'Sullivan, Davis, Billington, Gonzalez-Diaz, & Corcoran, 2015; Vaughan-Evans et al., 2016; Wallentin et al., 2011; Willems, Frank, Nijhof, Hagoort, & van den Bosch, 2015; Willems & Jacobs, 2016; Zeman, Milton, Smith, & Rylance, 2013). Next, we discuss some example predictions straightforwardly emerging from the above quantitative narrative analyses.

Predictions based on present results

The present quantitative narrative analysis data allow to formulate nested hypotheses at three levels of detail: poem category (young man vs. dark lady), across poem contrasts (poem X vs. poem Y), and within-poem contrasts (quatrain 1–3 vs. couplet, octave vs. sestet, line- or wordwise). We will give examples for all of

them in the hope to encourage further empirical research in line with the goals set in a recent review paper on the scientific study of literary experience and response (Jacobs, 2015c; 2016a).

Poem category

At this supra-poem level, we found several descriptive differences between the young man and dark lady poems that lead to testable hypotheses. An example of interest in the light of Simonton's (1989, 1990) analyses concerns the Thematic richness index which, on average, was significantly higher for the young man sonnets than the dark lady ones, suggesting a global thematic-semantic narrowing: means = 2.9 + −0.12 vs. 1.3 + −0.25, $F(1, 152) = 32.6$, $p < .0001$, $R^2 = .18$.[2] Thus, we can hypothesize that on average – and all other things being equal – readers will be more inclined to like one or more randomly selected young man sonnet(s) more than dark lady one(s) and rate it as perceiving/inducing a more positive mood.

Before developing and testing any specific hypotheses in this regard, we recommend to augment such quantitative narrative analysis-based statistical analyses by qualitative content analyses done by experts, e.g. literary scholars who could use the Abstractness Scale (Jacobs, 2015b) or similar tools for rating the Thematic richness index or similar features (cf. Jacobs et al., 2016a). Other indirect on- or offline measures (e.g., eye tracking, neuroimaging, free recall, response times) could be used to cross-validate the direct offline measures (rating data; Dixon & Bortolussi, 2015; Jacobs, 2016a): neuroimaging data indicating a higher activation of neural networks associated with aesthetic liking (e.g., orbitofrontal cortex; Brown et al., 2011; Jacobs et al., 2016b) for young man sonnet(s) would be a case in point.

Across-poem contrasts

A simple hypothesis based on the quantitative narrative analysis data summarized in Part I is that overall reading time is greater for sonnet 1 than for sonnet 138. More specific hypotheses regarding eye tracking studies can be derived from Figure 1, e.g., that sonnet 138 with its multiple higher Cohesion scores should produce longer mean first-pass fixation times on text parts important for coherence building (e.g., conjunctions) than sonnet 1 (cf. Louwerse, 2001). These hypotheses can be generalized, of course, by stating them "parametrically," e.g., the higher

2. We first established equal variances for the two different sample sizes (O'Brien $F(1, 152) = 1.75$, $p = .18$; Brown-Forsythe $F(1, 152) = 2.44$, $p = .12$), and also checked by bootstrapping ($N = 100$) that the confidence intervals of the mean difference (−1.6 + −0.95, $CI_u = -1.05$, $CI_l = -2.2$) did not differ from those of the bootstrapping ($CI_u = -1.04$, $CI_l = -2.16$). Finally, we computed a nonparametric test which confirmed the results of the one-way ANOVA: Wilcoxon/Kruskal-Wallis: $S = 1106.5$, $Z = -5.25$, $p < .0001$.

the CML2 readability score of a sonnet, the shorter should be its reading time, mean gaze durations etc. The data in Table 1 allow more specific hypotheses, such as that, say, narrativity ratings are higher for sonnet 42 than for sonnet 1, or that comprehensibility ratings are higher for sonnets 51 and 52 than for 31 and 144. Eye tracking experiments could also test the hypothesis that sonnets 145 and 125 produce a smaller likelihood of regressive saccades and longer gaze durations (associated with syntactic complexity, if other relevant variables are controlled for) than sonnets 80 or 67.

If reading speed or related eye movement parameters were the response measure of choice, the surprisal data of Figure 3 also are of interest. They allow to hypothesize that reading time also (co-)varies significantly with the sonnets' surprisal value, as should do the N400 amplitude, if ERP were the response measure. The German poet Durs Grünbein (1996) called for a poetry full of images rich in "factor N400", which he considered to be an index of the foregrounding potential of metaphors, speculating that such metaphors cause "neurolinguistic clashes" (cf. Jacobs, 2015b). According to the Neurocognitive Poetics Model (Jacobs, 2011, 2015a, 2015b), sonnets/lines/words with higher surprisal – and thus foregrounding – potential should more likely produce higher liking ratings, smaller saccades and longer fixation durations than sonnets low on surprisal. Data from a recent eye tracking study using short literary stories support these predictions (Van den Hoven, Hartung, Burke, & Willems, 2016) and it will be intriguing to see whether they also hold for sonnet reading. Regarding potential neuroimaging studies on sonnet reception, the surprisal data in Figure 3 can be used to predict selective activation in the left inferior temporal sulcus, bilateral superior temporal gyrus, right amygdala, bilateral anterior temporal poles, and right inferior frontal sulcus (cf. Willems et al., 2015).

Turning to the affective-aesthetic aspects, the results in Part II (Figure 4 and Tables 2 and 3) allow a number of predictions concerning a variety of response measures. At the level of direct offline measures (e.g., questionnaires, scales) sonnets with high or optimal values for emotion potential (140, 151, or 144) and/or Regressive Imagery Dictionary/Primary process language (153, 154, 73) should produce significantly higher liking ratings, for example, than sonnets scoring low (or too low/ too high) on this composite dimension. Activation of the reward networks involved in aesthetic liking of literature (Jacobs et al., 2016b) should correlate with such ratings, as could electrodermal activity (Jacobs et al., 2016a). The data in Figure 4 and Table 3 are of special interest for empirical investigations because they raise the issue which of the five indices associated with mood perception and/or induction in poetry reception is the most valid and reliable. While the first three indices (Valence mean, Span, and Sum) already have been shown to affect mood-related reader responses to some degree (Aryani et al., 2016; Lüdtke

et al., 2014; Jacobs et al., 2016b), the other two are novel (SEANCE Negative, Positive mood) and still await empirical validation.

Part II also introduces three other novel variables (Thematic richness index, Symbolic imagery index, and Semantic association potential) that have not yet been used – as far as we know – in empirical studies on poetry reception. Tentatively, all three can be expected to correlate positively with liking and (positive) mood ratings, as well as other response measures (of the indirect type, e.g., electrodermal or neuronal activity) that are associated with liking. Following Hofmann and Jacobs's (2014) and Kuchinke et al.'s (2013) results, a neuroimaging study on sonnet reception should – all other things being equal – also find increased activity in hippocampus, left inferior frontal gyrus, or the temporal pole for sonnets high on Semantic association potential, reflecting larger semantic competition as a function of more active representations (cf. also Forgács et al., 2012).

Within-poem contrasts (e.g., quatrain 1–3 vs. couplet, octave vs. sestet, or linewise)

A straightforward prediction derived from the quantitative narrative analyses regarding the sonnet dynamics (cf. Table 4) is that generally the couplets and the sestets should be easier to process, read and comprehend than the sonnets' bodies and octaves, respectively. This effect could be captured by a variety of measures including ratings, eye tracking or brain-electrical and neuroimaging measures. Since couplets and sestets also appear to have a higher Emotion potential than the bodies or octaves, response measures sensitive to affective-aesthetic variables also should produce significant differences for these within-poem contrasts.

The line- and wordwise quantitative narrative analysis results shown in Figure 3 encourage even more fine-grained hypotheses concerning measures related to the comprehensibility and/or affective-aesthetic responses, but will depend on the exact research question at hand. A straightforward example is to test the prediction of the Neurocognitive Poetics Model that – again, all other things being equal – higher surprisal values more likely produce higher liking ratings, smaller saccades and longer fixation durations on a line- or even wordwise basis, e.g., with lines like line 12 from sonnet 1 (surprisal value = 5.85: *And, tender churl, makest waste in niggarding*), or line two from sonnet 138 (2.43: *I do believe her, though I know she lies*). Finally, in line with Steen's (2004) proposals, Figure 5 offers a wide field of interesting hypotheses concerning the processing of foregrounding elements. If everything else was controlled for, the likelihood of recognizing and/or appreciating stylistic devices, e.g. as assessed by a marking test, should increase quasi-linearly towards the end of sonnets. For metaphorically used words contained in Bob Dylan's lyrics of *Hurricane*, Steen (2004) indeed confirmed this and we can only speculate that the same should hold for the present sonnets. In

addition, the incremental nature of text comprehension – a reader's knowledge of the text world becoming progressively larger, more specific, and more concrete – coupled with the decreasing number of new words could have measurable effects on a number of mental processes, e.g. attentional, mnestic, or emotional. To what extent on-line measures of sonnet reception such as eye tracking can capture such effects is an issue we cannot develop here, but a very general prediction is that overall reading speed and its multiple correlates should increase (linearly or non-linearly) with increasing line number.

Part IV. Machine-learning-based computational modeling

In this section, we aim to show how quantitative narrative analyses can be usefully combined with machine-learning-based computational modeling for identifying those of the many quantifiable sonnet features that play a potential key role, as well as generating refined hypotheses for future empirical investigations. We basically follow the approach adopted by Jacobs et al. (2016b) and Jacobs and Kinder (2017). While the former computationally modeled elementary affective decisions (i.e., dis-liking) to words, the latter demonstrated that such machine-learning assisted quantitative narrative analysis approaches can predict the period of origin, authorship and *aptness* of poetic metaphors, encouraging the present approach. However, in contrast to these studies, in the present case we lack empirical response data. The modeling goal thus was to classify the 154 sonnets into two groups which can be interpreted as the result of a hypothetical empirical study assessing binary expert ratings on the sonnets' two meta-motifs (young man vs dark lady).[3]

Classifying the sonnets via machine-learning assisted quantitative narrative analysis

Similar to Jacobs and Kinder (2017), here we adopt an exemplary formal modeling approach to illustrate how quantitative narrative analysis data can be used to predict the topic of texts, in our case the binary decision concerning the above mentioned young man vs dark lady motifs said to divide the 154 sonnets into two categories. We tested several competing models using different text features extracted via quantitative narrative analysis (Jacobs et al., 2016b). The results of these tests are summarized in Table 5. The decision tree models (also called recursive partioning models) used here are a nonparametric method successfully used for

3. We are aware of the fact that the sonnets can be categorized into more than these two groups, e.g. into the "young man 1" group (1–17) and the "young man 2" group (18–126), or the "greek" group (153–154), but decided to keep things simple here.

exploring relationships without having a good prior theoretical model: they can handle even large data problems, i.e., large numbers of predictor variables even in the presence of complex higher-order interactions and nonlinearity efficiently allowing to test clear hypotheses, and the results are usually transparent and easily interpretable (e.g., Loh, 2011). As a simpler alternative suggested by an anonymous reviewer, we also tested Linear (parametric) Discriminant Analysis models. Since we were sceptical that in all their complexity sonnets are linearly separable stimuli we also tested nonlinear (quadratic) discriminant analysis models. Following Jacobs et al. (2016b) we used a stepwise modeling approach going from simple to complex models (i.e., few vs. many input variables) to see how much complexity in the input space is necessary to obtain an adequate model performance.

The stimuli were the 154 sonnets (126 young man and 28 dark lady). We created two sets of models tentatively termed *cognitive/C* and *affective/A*. Model C1 comprises the eight Coh-metrix easability scores shown in Figure 1 (i.e., Narrativity, Syntactic simplicity, Word concreteness, Referential, Deep and Verb cohesion, Connectivity, and Temporality) and Model C2 accumulates all 76 Coh-metrix descriptors we computed for the sonnets (see http://cohmetrix.com/), excluding the eight easability scores of model C1. The *affective set* contained Model A1 with the three Regressive Imagery Dictionary indices, model A2 with the five mood indices, and model A3 with the 20 SEANCE component scores. The final "supermodel" combined models C2 and A3 launching 96 cognitive and affective variables into the race to classify 154 stimuli.

The models were implemented using the PARTITION and MULTIVARIATE tools of the JMP Pro11 software (SAS Institute Inc., Cary, NC, 1989–2007) and model performance was gauged by the number of correct decisions, i.e., whether the model classified a sonnet correctly as belonging either to category one or two. Descriptively, decision tree model performance is expressed by the number of partitions, i.e., how many decisions are required to obtain maximum accuracy, entropy R^2 and the rate of misclassifications, i.e. how often the model classified a sonnet incorrectly. Table 5 summarizes the results. Each model in the table implements and tests a different hypothesis concerning the factors determining sonnet classification, e.g., model A1 tests to what extent the three Regressive Imagery Dictionary scores predict correct classification.

The results summarized in Table 5 offer several take-home messages. First, if a great number of input variables from quantitative narrative analyses is used, both decision tree and quadratic discriminant analysis models can almost perfectly classify the sonnets into the two standard categories. Using 96 input variables for classifying 154 stimuli might seem close to *overfitting* (using too many and thus potentially irrelevant predictors/parameters and thus picking up noise in the data together with the signal). However, both the decision tree and discriminant

Table 5. Input variables, characteristics and performance evaluation for 12 models

Model	Nbr. of input variables	Model characteristics (nbr of splits, two strongest predictors)	Model performance (entropy R^2, misclassification rate in%; [k-fold cross-validation R^2, $k = 5$])
C1a. Decision Tree	Coh-metrix eight easability scores	24, word concreteness, temporality	.88, 0%, [.68]
C1b. Linear/quadratic Discriminant Analysis	Coh-metrix eight easability scores	–	linear: −0.12,** 25% quadratic: .65, 7.1%
C2a. Decision Tree	Coh-metrix 76 features	14, adverbial phrase density incidence, lexical diversity	.92, 0% [.91]
C2b. Linear/quadratic Discriminant Analysis	Coh-metrix 76 features	–	linear: .73, 5.9% quadratic: .86, 0.7%
A1a. Decision Tree	Regressive Imagery Dictionary 3 features	42, emotions, secondary process	.79, 1%, [−.27]
A1b. Linear/quadratic Discriminant Analysis	Regressive Imagery Dictionary 3 features	–	linear: −.39, 39% quadratic: −.33, 37%
A2a. Decision Tree	Five mood indices: Val mean, span, sum, SEANCE positive & negative mood	26, valence sum, valence mean	.87, 0% [.52]
A2b. Linear/quadratic Discriminant Analysis	Five mood indices: Val mean, span, sum, SEANCE positive & negative mood	–	linear: −.2, 29% quadratic: −.16, 28%
A3a. Decision Tree	SEANCE 20	15, affect friends & family, trust verbs	.91, 0% [.8]
A3b. Linear/quadratic Discriminant Analysis	SEANCE 20	–	linear: .03, 20% quadratic: .81, 3.2%
C2 + A3a. Decision Tree	C2 + A3 96	12, lexical diversity, affect friends & family	.93, 0% [.92]
C2 + A3b. Linear/quadratic Discriminant Analysis	C2 + A3 96	–	linear: .98, 0% quadratic: 1, 0%

Notes: The minimum split for all decision tree models is $N = 1$.

* This procedure for decision tree models randomly divides the original data into K subsets. In turn, each of the K sets is used to validate the model fit on the rest of the data, fitting a total of K models. The final model is selected based on the cross-validation R^2, where a constraint is imposed to avoid overfitting the model.

** A negative entropy R2 indicates an inadequate model fit.

analysis models offer means to select relevant or important variables from the total set of 96 predictors. Thus, a closer look at the results produced by the C2 + A3 decision tree model revealed that 86 input variables accounted for 0% variance, thus leaving only 10 *relevant* input variables (rank order: Coh-metrix Lexical diversity, SEANCE Affect friends & family, SEANCE Negative adjectives, Coh-metrix Pronoun incidence, Coh-metrix Hypernymy nouns and verbs, Coh-metrix Age of acquisiton content words, Coh-metrix Noun phrase density incidence, Coh-metrix Minimal edit distance part of speech, SEANCE Positive verbs, Coh-metrix Negation density incidence; for details on these variables please refer to the homepages given in the original articles). The discriminant analysis tool offers a stepwise selection procedure also allowing to select the most important variables. Second, *decision tree* models (C1a, A1a, A2a) using only three to eight ‚cognitive' or ‚affective' input variables already do quite a good job at classifying the sonnets, as evidenced by entropy R^2s between .79 and .92 and maximum misclassification rates of 1%. Third, in order to allow reliable predictions on test data (e.g., 30% of the sonnets) obtained from training data (e.g., 70% of the sonnets), decision tree models seem to require a greater number of input variables (see k-fold cross-validation R^2 for models C2a, A3a, and C2 + A3a). Fourth, it should be noted that the results of Table 5 are purely descriptive and exploratory and thus allow no conclusions with regard to the relative validity or utility of the computational models used. At this stage of research, we do not intend to show that a particular quantitative narrative or machine learning model (with a particular parameter set) is an *optimal classifier* of the 154 Shakespeare sonnets and make no claims that the present approach will generalize to other corpora. Rather, we aim at showing how in principle such a computational modeling approach can help classify even complex poetic materials. Such classifications are useful for stimulus selection, response prediction/analysis and hypothesis or design generation in (neuro-)cognitive poetics studies. Thus, Table 5 can be used in a heuristic fashion for generating or testing hypotheses. For example, the fact that valence features among the strongest predictors in model A2a strengthens the hypothesis that basic affective features are part-and-parcel of poetic texts which can significantly co-determine liking and poeticity ratings (e.g., Aryani et al., 2016; Jacobs et al., 2016a, 2016b; Jacobs & Kinder, 2017; Ullrich et al., 2017).

For illustrative purposes Figure A1 in the Appendix shows the detailed results for model A3a and helps understand the key variables that drive the correct classifications in this model.

In conclusion, the two meta-motifs of the 154 sonnets can very well be predicted by a machine learning algorithm using quantitative narrative analysis indices. More generally, our analyses suggest that this methodological combination can serve as a heuristic for classifying texts into meta categories that could help

identify authors, (sub)genres, epochs, or meta-motifs like in the present application. As mentioned above, it can even be used to predict ratings of the aptness of poetic metaphors, including a sample of Shakespearean ones (Jacobs & Kinder, 2017). This opens new perspectives for future research, e.g. predicting the aptness of the many metaphors occurring in the present sonnets. Another future application could make use of even more powerful combinations of machine learning and quantitative narrative analysis tools. Thus, following Jacobs and Kinder (2017), one could combine Latent Semantic Analysis (Landauer & Dumais, 1997) with a more complex variant of decision tree models called *Boosted Tree* (a variant that builds a large, additive decision tree by fitting a sequence of smaller trees, each of which is fit on the scaled residuals of the previous tree; cf. Dietterich, 2000) to try classify the 2155 lines of the 154 sonnets, i.e. deciding which line belongs to which sonnet.

Before discussing some obvious limitations of the present work, we would like to borrow Tsur's (2008) statement paraphrasing Miller (1993): "Our task is not to search for a unique paraphrase of the text, nor to find out how many meanings can be attributed to it, but to search for grounds that will constrain the basis of interpretations to a plausible set of alternatives" (p. 147). We believe that the approach chosen in the present paper is in the spirit of Miller. If there are at least two basic levels of understanding texts and poetry in particular – evocation and interpretation (at rereading; Rosenblatt, 1978) – then quantitative narrative analyses plausibly can help capture aspects of the first level and arguably also of the second. While the number of possible meanings a reader can (re-)construct from a given poem in multiple re-readings may be quasi-unlimited, empirical findings indicate that students often fail to engage the poems used in a study in a manner that accounts for the poems' "poetic significance" with the consequence that what were essentially "plain sense" prose translations of the poems (cf. Richards, 1929, ten major pitfalls in poetry reading) rather than "evocations" of their possible meanings resulted (e.g., Harker, 1997). To capture aspects of deep reading of poetry, we recommend augmenting such quantitative narrative analysis-based statistical analyses by qualitative content analyses done by experts, e.g., scholars of literature, poetics, or linguistics, as exemplified in Jacobs et al. (2016b).

Of course, quantitative narrative analyses applied to stimulus selection/control and response prediction can only be as good or useful as the task and the response measures – and the hypotheses meaningfully relating stimuli and responses – developed by the experimenters allow them to be. That is, the methods for measuring experience/response should fit well with the hypotheses based on quantitative narrative analysis or other tools. To what extent the direct vs. indirect on- and offline measures of poetry reception proposed in Dixon and Bortolussi (2015) and intensely debated with Kuiken (2015) and Jacobs (2016b) can capture

the "plain sense", evocation and/or interpretation aspects of any poetry reading act is an open issue that – in our opinion – can benefit from the application of quantitative narrative analysis as much as from the development of more sophisticated models and methods for the study of literary reading.

This being said, at least two obvious lacunae limit the usefulness of the present quantitative narrative analyses for studies of the dynamic sound-meaning nexus typical for poetry reception (e.g., Schrott & Jacobs, 2011; Tsur, 1998).

First, the lack of predictors at the level of implicit or mental sound (i.e. generated via phonological or prosodic recoding of the printed input), phonological iconicity, rhythm, or rhyme, which all have been shown to affect reader responses in silent lyrics or poetry processing to some extent (e.g., Aryani et al., 2016; Menninghaus et al., 2014; Tsur, 2006; Wallace & Rubin, 1991). As an example, in their ground-breaking case analysis of Baudelaire's "Les chats," Jakobson and Lévi-Strauss (1962) analyzed the *phonological texture* of the poem by quantifying the number of nasals in the poem's first quartet ("two to three per line") or the interaction between formal and semantic features (i.e., nasal vowels and the idea of light) in the last trio. It should be noted, though, that when dealing with written sonnets we know of no firm evidence that non-expert readers silently read sonnets in any way resembling theories of scansion. Even reading aloud the sonnets must not strictly follow the iambic pentameter but take into account subtler intonations observing inner antitheses and parallels (cf. Vendler, 1997, p. 37). In sum, complementing the present quantitative narrative analyses of sonnets by tools like SPARSAR (Delmonte, 2016) for quantifying structural, or EMOPHON (Aryani et al., 2013) for affective sound properties would be a good first step towards allowing predictions about potential "sound" (including rhythm) effects. Still, such efforts must be preceded or accompanied by experiments demonstrating exactly which implicit structural and/or affective sound properties affect poetic reading acts in addition to – or in interaction with – the present or other quantitative narrative analysis variables.

The second obvious lacuna is the absence of qualitative descriptors of the metaphoricity or, more generally, the foregrounding/backgrounding quotient (Jacobs, 2015b) of the sonnets (e.g., McQuarrie & Mick, 1996; McQuire et al., 2017; Pragglejazz group, 2007; Schrott & Jacobs, 2011; Steen, 1999, 2002, 2004; Stockwell 2009). For example, it can be safely assumed that a line like *Feed'st thy light's flame with self-substantial fuel,* from quatrain 2 of sonnet 1 above is not a summation of the phonological and semantic representations of its individual words but – as outlined in the introduction – a *catachresis* which likely evokes aesthetic and reflective reader responses according to the Neurocognitive Poetics Model (Jacobs, 2015a). Especially in poetry, meaning emerges dynamically out of the full context determining which semantic fields and senses of a word are heightened and which

are deactivated (cf. Millis & Larson, 2008; Schrott & Jacobs, 2011). Naturally, the "static" purely descriptive quantitative narrative analysis indices presented are insufficient (and also not meant) to explain such context-dependent, reader-specific emergent dynamics and attempts at, say, metaphoric constructions. Moreover, they also neglect potential conceptual or rhetorical effects produced at a deeper linguistic level, e.g., modifications of tense, subject-position, or clause-patterns. Finally, they offer no analysis of "what isn't printed in the text," e.g. ellipsis or allusions (cf. Jacobs, 2015b).

Still, the present tools can be augmented by qualitative, typological or taxonomic tools like the Abstractness Scale for determining foregrounding features such as meter or mimesis (Jacobs, 2015b; Meyer-Sickendieck, 2011), by metaphoricity analyses that, e.g., count and interpret antitheses or chiasma, so frequently used in the sonnets, or that evaluate the conceptual, linguistic, communicative, or affective qualities of metaphors (e.g., Jacobs & Kinder, 2017; Schrott & Jacobs, 2011; Steen, 1999; Stockwell, 2009; Vendler, 1997), as well as by computational linguistic analyses (e.g., Kintsch, 2000; Kintsch & Magalath, 2011). This should help develop full-fledged process models of the type discussed in Jacobs (2015b) which may serve at least as sophisticated null-models for predicting context- and reader-dependent effects of poetic text features on direct or indirect response measures.

In conclusion, the present quantitative narrative analysis approach to sonnets is not meant to replace deep-structure expert qualitative analyses or critical interpretations of the kind of Jakobson and Jones (1970) or Vendler (1997). Rather, it should serve as a complement or null-model against which any model of foregrounding effects due to stylistic devices can be tested regarding its account of additional variance in reader responses.

References

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

Altmann, U., Bohrn, I. C., Lubrich, O., Menninghaus, W., & Jacobs, A. M. (2014). Fact vs fiction-how paratextual information shapes our reading processes. *Social Cognitive and Affective Neuroscience*, 9(1), 22–29. doi: 10.1093/scan/nss098

Anderson, T., & Crossley, S. (2011). "Rue with a Difference: A Stylistic Analysis of the Rhetoric of Suicide in *Hamlet*" (pp. 192–214). In Jonathan Culpepper and Mireille Ravassat (Eds.), *Stylistics and Shakespeare's language: Transdisciplinary approaches*. Bloomsbury.

Aryani, A., Jacobs, A. M., & Conrad, M. (2013). Extracting salient sublexical units from written texts: "Emophon," a corpus-based approach to phonological iconicity. *Frontiers in Psychology*, 4:654. doi: 10.3389/fpsyg.2013.00654

Aryani, A., Kraxenberger, M., Ullrich, S., Jacobs, A. M., & Conrad, M. (2016). Measuring the basic a ective tone of poems via phonological saliency and iconicity. *Psychology of Aesthetics, Creativity, and the Arts*, 10, 191–204. doi: 10.1037/aca0000033

Beaty, R. E., Benedek, M., Silvia, P. J., & Schacter, D. L. (2016). Creative cognition and brain network dynamics. *Trends in Cognitive Sciences*, 20, 87–95. doi: 10.1016/j.tics.2015.10.004

Berlyne, D. E. (1971). *Aesthetics and Psychobiology*. New York: Appleton-Century-Crofts.

Bestgen, Y. (1994). Can emotional valence in stories be determined from words?. *Cognition & Emotion*, 8(1), 21–36. doi: 10.1080/02699939408408926

Bohrn, I. C., Altmann, U., & Jacobs, A. M. (2012a). Looking at the brains behind figurative language – A quantitative meta-analysis of neuroimaging studies on metaphor, idiom and irony processing. *Neuropsychologia*, 50, 2669–2683. doi: 10.1016/j.neuropsychologia.2012.07.021

Bohrn, I. C., Altmann, U., Lubrich, O., Menninghaus, W., & Jacobs, A. M. (2012b). Old proverbs in new skins – an FMRI study on defamiliarization. *Frontiers in Psychology*, 3:204. doi: 10.3389/fpsyg.2012.00204

Bohrn, I. C., Altmann, U., Lubrich, O., Menninghaus, W., & Jacobs, A. M. (2013). When we like what we know – A parametric fMRI analysis of beauty and familiarity. *Brain and Language*, 124(1), 1–8. doi: 10.1016/j.bandl.2012.10.003

Boston, M. F., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, 2(1), 1–12. doi: 10.16910/jemr.2.1.1

Boyd, B. (2012). *Why lyrics last: Evolution, Cognition, and Shakespeare's Sonnets*. Harvard University Press. doi: 10.4159/harvard.9780674064843

Bradley, M. M., & Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Stimuli Instruction and Affective Ratings Technical Report C-1*. Gainesville, FL: University of Florida.

Briesemeister, B. B., Kuchinke, L., & Jacobs, A. M. (2011). Discrete emotion norms for nouns: Berlin affective word list (DENN – BAWL). *Behavior Research Methods*, 43(2), 441–448. doi: 10.3758/s13428-011-0059-y

Briesemeister, B. B., Kuchinke, L., & Jacobs, A. M. (2014). Emotion word recognition: Discrete information effects first, continuous later?. *Brain Research*, 1564, 62–71. doi: 10.1016/j.brainres.2014.03.045

Briesemeister, B. B., Kuchinke, L., Jacobs, A. M., & Braun, M. (2015). Emotions in reading: Dissociation of happiness and positivity. *Cognitive, Affective, & Behavioral Neuroscience*, 15(2), 287–298. doi: 10.3758/s13415-014-0327-2

Brown, S., Gao, X., Tisdelle, L., Eickhoff, S. B., & Liotti, M. (2011). Naturalizing aesthetics: brain areas for aesthetic appraisal across sensory modalities. *Neuroimage*, 58(1), 250–258. doi: 10.1016/j.neuroimage.2011.06.012

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990. doi: 10.3758/BRM.41.4.977

Burke, M. (2011). *Literary reading, cognition and emotion: An exploration of the oceanic mind*. London: Routledge.

Burke, M. (2015). The neuroaesthetics of prose fiction: Pitfalls, parameters and prospects. *Frontiers in Human Neuroscience*, 9. doi: 10.3389/fnhum.2015.00442

Chatterjee, A., & Vartanian, O. (2014). Neuroaesthetics. *Trends in Cognitive Sciences*, 18(7), 370–375. doi: 10.1016/j.tics.2014.03.003

Chen, Q., Zhang, J., Xu, X., Scheepers, C., Yang, Y., & Tanenhaus, M. K. (2016). Prosodic expectations in silent reading: ERP evidence from rhyme scheme and semantic congruence in classic Chinese poems. *Cognition*, 154, 11–21. doi: 10.1016/j.cognition.2016.05.007

Citron, F. M. (2012). Neural correlates of written emotion word processing: a review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language*, 122(3), 211–226. doi: 10.1016/j.bandl.2011

Cook, A. (2010). *Shakespearean neuroplay: Reinvigorating the study of dramatic texts and performance through cognitive science*. NY: Palgrave Macmillan doi: 10.1057/9780230113053

Crane, M. T. (2001). *Shakespeare's Brain: Reading with Cognitive Theory*. Princeton University Press.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*, 48(4), 1227–1237. doi: 10.3758/s13428-015-0651-7

Crossley, S. A., Kyle, K., & McNamara, D. S. (2017). Sentiment Analysis and Social Cognition Engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behavior Research Methods*, 49(3), 803–821. doi: 10.3758/s13428-016-0743-z

Crystal, D. (2012). *Think on my Words: Exploring Shakespeare's Language (Canto Classics)*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139196994

Cupchik, G. C. (1986). A decade after Berlyne: New directions in experimental aesthetics. *Poetics*, 15(4–6), 345–369. doi: 10.1016/0304-422X(86)90003-3

De Beaugrande, R. A. (1979). Toward a general theory of creativity. *Poetics*, 8(3), 269–306. doi: 10.1016/0304-422X(79)90036-6

Delmonte, R. (2016). Exploring Shakespeare's Sonnets with SPARSAR. *Linguistics and Literature Studies*, 4(1), 61–95. doi: 10.13189/lls.2016.040110

Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, 40(2), 139–157. doi: 10.1023/A:1007607513941

Dixon, P., & Bortolussi, M. (2015). Measuring literary experience. *Scientific Study of Literature*, 5(2), 178–182. doi: 10.1075/ssol.5.2.03dix

Fellbaum, C. (Ed.). (1998). WordNet: An electronic lexical database. Cambridge: MIT Press.

Forgács, B., Bohrn, I., Baudewig, J., Hofmann, M. J., Pléh, C., & Jacobs, A. M. (2012). Neural correlates of combinatorial semantic processing of literal and figurative noun noun compound words. *Neuroimage*, 63(3), 1432–1442. doi: 10.1016/j.neuroimage.2012.07.029

Frank, S. L. (2013). Uncertainty reduction as a measure of cognitive load in sentence comprehension. *Topics in Cognitive Science*, 5(3), 475–494. doi: 10.1111/tops.12025

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, 140, 1–11. doi: 10.1016/j.bandl.2014.10.006

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford University Press. doi: 10.1093/0195138929.001.0001

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods*, 36(2), 193–202. doi: 10.3758/BF03195564

Graesser, A. C., & McNamara, D. S. (2010). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 3(2), 371–398. doi: 10.1111/j.1756-8765.2010.01081.x

Graesser, A. C., Dowell, N., & Moldovan, C. (2011). A computer's understanding of literature. *Scientific Study of Literature*, 1(1), 24–33. doi: 10.1075/ssol.1.1.03gra

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

Grünbein, D. (1996). Katze und Mond. In D. Grünbein (Eds) *Galilei Vermisst Dantes Hölle und Bleibt an den Maßen Hängen* (pp. 116–128). Frankfurt AM: Suhrkamp.

Habermas, T., & de Silveira, C. (2008). The development of global coherence in life narratives across adolescence: Temporal, causal, and thematic aspects. *Developmental Psychology*, 44(3), 707–721. doi: 10.1037/0012-1649.44.3.707

Harker, W. J. (1997). Literary texts as models: Implications for the empirical study of literature. In S. Tötösy de Zepetnek & I. Sywenky (Eds.), *The systemic and empirica lapproach to literature and culture as theory and application* (pp. 51–65). Edmonton, Canada: University of Alberta.

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. doi: 10.1016/j.neubiorev.2014.06.011

Hope, J., & Witmore, M. (2004). The very large textual object: a prosthetic reading of Shakespeare. *Early Modern Literary Studies*, 9(3), 1–36.

Hsu, C. T., Conrad, M., & Jacobs, A. M. (2014). Fiction feelings in Harry Potter: Haemodynamic response in the mid-cingulate cortex correlates with immersive reading experience. *Neuroreport*, 25(17), 1356–1361. doi: 10.1097/WNR.0000000000000

Hsu, C. T., Jacobs, A. M., Citron, F. M., & Conrad, M. (2015). The emotion potential of words and passages in reading Harry Potter – An fMRI study. *Brain and Language*, 142, 96–114. doi: 10.1016/j.bandl.2015.01.011

Hutcheon, L. (2012). *A Theory of Adaptation*. 2nd edition with Siobhan O'Flynn. London and New York: Routledge.

Hutchins, R. (Ed.) (1952). *Great books of the Western world* (54 vols.), Chicago: University of Chicago Press.

Jacobs, A. M. (2011). Neurokognitive Poetik: Elemente eines Modells des literarischen Lesens [Neurocognitive poetics: Elements of a model of literary reading]. In R. Schrott and 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). Towards a neurocognitive poetics model of literary reading, in R. Willems (Ed.) *Towards a Cognitive Neuroscience of Natural Language Use* (pp. 135–159). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107323667.007

Jacobs, A. M. (2015b). Neurocognitive poetics: Methods and models for investigating the neuronal and cognitive-affective bases of literature reception. *Frontiers in Human Neuroscience*, 9:186. doi: 10.3389/fnhum.2015.0018

Jacobs, A. M. (2015c). The scientific study of literary experience. *Scientific Study of Literature*, 5(2), 139–170. doi: 10.1075/ssol.5.2.01jac

Jacobs, A. (2016a). The scientific study of literary experience and neuro-behavioral responses to Literature: Reply to Commentaries. *Scientific Study of Literature*, 6(1), 164–174. doi: 10.1075/ssol.6.1.08jac

Jacobs, A. M., and Kinder, A. (2015). Worte als Worte erfahren: wie erarbeitet das Gehirn Gedichte (Experience words as words: how the brain constructs poems), in A. Pompe (Ed.) *Kind und Gedicht* [Child and Poem] (pp. 57–76). Berlin: Rombach.

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 & Symbol*, in press.

Jacobs, A. M., and Schrott, R. (2015). Gefesselt im Kopfkino: Von Kippschaltern, Madeleine Effekten und Don Quichote Syndromen bei der Immersion in Textwelten (Captivated in the mind's cinema: Of trigger-switches, Don Quichote syndroms and immersion in text worlds). Available online at: FIKTION.CC

Jacobs, A. M., Võ, M. L. H., Briesemeister, B. B., Conrad, M., Hofmann, M. J., Kuchinke, L., Lüdtke, J. & Braun, M. (2015). 10 years of BAWLing into affective and aesthetic processes in reading: what are the echoes?. *Frontiers in Psychology*, 6:714. doi: 10.3389/fpsyg.2015.00714

Jacobs, A. M., Lüdtke, J., Aryani, A., Meyer-Sickendieck, B., & Conrad, M. (2016a). Mood-empathic and aesthetic responses in poetry reception. *Scientific Study of Literature*, 6(1), 87–130. doi: 10.1075/ssol.6.1.06jac

Jacobs, A., Hofmann, M. J., & Kinder, A. (2016b). On elementary affective decisions: To like or not to like, that is the question. *Frontiers in Psychology*, 7:1836. doi: 10.3389/fpsyg.2016.01836

Jacobsen, T. (2006). Bridging the arts and sciences: A framework for the psychology of aesthetics. *Leonardo*, 39, 155–162. doi: 10.1162/leon.2006.39.2.155

Jakobson, R., & Jones, L. G. (1970). *Shakespeare's Verbal Art in Th'Expence of Spirit* (No. 35). Walter de Gruyter.

Jakobson, R., & Lévi-Strauss, C. (1962). "Les chats" de Charles Baudelaire. *L'homme*, 2, 5–21. doi: 10.3406/hom.1962.366446

Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review*, 7(2), 257–266. doi: 10.3758/bf03212981

Kintsch, W. (2012). Musings about beauty. *Cognitive Science*, 36(4), 635–654. doi: 10.1111/j.1551-6709.2011.01229.x

Kintsch, W., & Mangalath, P. (2011). The construction of meaning. *Topics in Cognitive Science*, 3(2), 346–370. doi: 10.1111/j.1756-8765.2010.01107.x

Kiss, G. R., Armstrong, C., Milroy, R. and Piper, J. (1973). An associative thesaurus of English and its computer analysis. In Aitken, A. J., Bailey, R. W. and Hamiltonsmith, N. (Eds), *The Computer and Literary Studies*. Edinburgh: University Press.

Kivy, P. (1991). *Music alone: Philosophical reflections on the purely musical experience*. Cornell University Press.

Koelsch, S., Jacobs, A. M., Menninghaus, W., Liebal, K., Klann-Delius, G., von Scheve, C., & Gebauer, G. (2015). The quartet theory of human emotions: An integrative and neurofunctional model. *Physics of Life Reviews*, 13, 1–27. doi: 10.1016/j.plrev.2015.03.001

Krumhansl, C. L. (1997). An Exploratory Study of Musical Emotions and Psychophysiology. *Canadian Journal of Experimental Psychology*, 51(4), 336. doi: 10.1037/1196-1961.51.4.336

Kuchinke, L., Fritzemeier, S., Hofmann, M. J., & Jacobs, A. M. (2013). Neural correlates of episodic memory: Associative memory and confidence drive hippocampus activations. *Behavioural Brain Research*, 254, 92–101. doi: 10.1016/j.bbr.2013.04.035

Kuhlmann, M., Hofmann, M. J., Briesemeister, B. B., & Jacobs, A. M. (2016). Mixing positive and negative valence: Affective-semantic integration of bivalent words. *Scientific Reports*, 6:30718. doi: 10.1038/srep30718

Kuzmičová, A. (2014). Literary narrative and mental imagery: A view from embodied cognition. *Style*, 48(3), 275–293.

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. doi: 10.1037/0033-295X.104.2.211

Lasswell, H. D., & Namenwirth, J. Z. (1969). *The Lasswell value dictionary*. New Haven: Yale University Press.

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, 95(4), 489–508. doi: 10.1348/0007126042369811

Leder, H., & Nadal, M. (2014). Ten years of a model of aesthetic appreciation and aesthetic judgments: The aesthetic episode – developments and challenges in empirical aesthetics. *British Journal of Psychology*, 105(4), 443–464. doi: 10.1111/bjop.12084

Leder, H., Markey, P. S., & Pelowski, M. (2015). Aesthetic emotions to art-What they are and what makes them special: Comment on "The quartet theory of human emotions: An integrative and neurofunctional model" by S. Koelsch et al. *Physics of Life Reviews*, 13, 67–70. doi: 10.1016/j.plrev.2015.04.037

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), e0124550. doi: 10.1371/journal.pone.0124550

Levinson, J. (1997). Music and negative emotion. In J. Robinson (Ed.), *Music and Meaning* (pp. 215–41). Ithaca, NY: Cornell University Press.

Lindquist, K. A., MacCormack, J. K., & Shablack, H. (2015). The role of language in emotion: Predictions from psychological constructionism. *Frontiers in Psychology*, 6:444. doi: 10.3389/fpsyg.2015.00444

Liu, S., Erkkinen, M. G., Healey, M. L., Xu, Y., Swett, K. E., Chow, H. M., & Braun, A. R. (2015). Brain activity and connectivity during poetry composition: Toward a multidimensional model of the creative process. *Human Brain Mapping*, 36(9), 3351–3372. doi: 10.1002/hbm.22849

Loh, W. Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1), 14–23. doi: 10.1002/widm.8

Lüdtke, J., & Jacobs, A. M. (2015). The emotion potential of simple sentences: Additive or interactive effects of nouns and adjectives?. *Frontiers in Psychology*, 6:1137. doi: 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. doi: 10.1037/a0036826

Marin, M. M. (2015). Crossing boundaries: Toward a general model of neuroaesthetics. *Frontiers in Human Neuroscience*, 9:443. doi: 10.3389/fnhum.2015.00443

Marin, M. M., Lampatz, A., Wandl, M., & Leder, H. (2016). Berlyne revisited: Evidence for the multifaceted nature of hedonic tone in the appreciation of paintings and music. *Frontiers in Human Neuroscience*, 10:536. doi: 10.3389/fnhum.2016.00536

Martindale, C. (1975). *The romantic progression: The psychology of literary history*. Halsted Press.

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*, 47(4), 292–330. doi: 10.1080/01638530902959943

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press. doi: 10.1017/CBO9780511894664

McQuarrie, E. F., & Mick, D. G. (1996). Figures of rhetoric in advertising language. *Journal of Consumer Research*, 22(4), 424–438. doi: 10.1086/209459

Mcquire, M., Mccollum, L., & Chatterjee, A. (2017). Aptness and beauty in metaphor. *Language and Cognition*, 9(2), 316–331. doi: 10.1017/langcog.2016.13

Meireles, R. C. (2005). The hermeneutics of symbolical imagery in Shakespeare's sonnets. Unpublished Dissertation, University. Porto Alegre, Bresil.

Menninghaus, W., Bohrn, I. C., Altmann, U., Lubrich, O., & Jacobs, A. M. (2014). Sounds funny? Humor effects of phonological and prosodic figures of speech. *Psychology of Aesthetics, Creativity, and the Arts*, 8(1), 71–76. doi: 10.1037/a0035309

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

Meyer-Sickendiek, B. (2011). *Lyrisches Gespür: Vom Geheimen Sensorium Moderner Poesie* [The lyrical sense of feeling. About the secret sensorium of modern poetry]. Paderborn, Germany: Fink.

Miall, D. S. (1989). Beyond the schema given: Affective comprehension of literary narratives. *Cognition & Emotion*, 3(1), 55–78. doi: 10.1080/02699938908415236

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

Miall, D. S., & Dissanayake, E. (2003). The poetics of babytalk. *Human Nature*, 14(4), 337–364. doi: 10.1007/s12110-003-1010-4

Miller, G. A. (1993). "Images and models, similes and metaphors". In A. Ortony (ed), *Metaphor and thought*, 2nd edition (pp. 357–400). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139173865.019

Miller, G. A. (1993). *Wörter: Streifzüge durch die Psycholinguistik*. Frankfurt: Zweitausendeins.

Millis, K., & Larson, M. (2008). Applying the construction-integration framework to aesthetic responses to representational artworks. *Discourse Processes*, 45(3), 263–287. doi: 10.1080/01638530802070015

Nicklas, P., & Jacobs, A. M. (2017). Rhetorics, neurocognitive poetics and the aesthetics of adaptation. *Poetics Today*, 38, 393–412. doi: 10.1215/03335372-3869311.

Oatley, K. (1994). A taxonomy of the emotions of literary response and a theory of identification in fictional narrative. *Poetics*, 23(1–2), 53–74. doi: 10.1016/0304-422x(94)p4296-s

Oatley, K. (2016). Fiction: Simulation of social worlds. *Trends in Cognitive Sciences*, 20(8), 618–628. doi: 10.1016/j.tics.2016.06.002

O'Sullivan, N., Davis, P., Billington, J., Gonzalez-Diaz, V., & Corcoran, R. (2015). "Shall I compare thee": The neural basis of literary awareness, and its benefits to cognition. *Cortex*, 73, 144–157. doi: 10.1016/j.cortex.2015.08.014

Panksepp, J. (2008). The power of the word may reside in the power of affect. *Integrative Psychological and Behavioral Science*, 42(1), 47–55. doi: 10.1007/s12124-007-9036-5

Pelowski, M., Markey, P. S., Lauring, J. O., & Leder, H. (2016). Visualizing the impact of art: An update and comparison of current psychological models of art experience. *Frontiers in Human Neuroscience*, 10:160. doi: 10.3389/fnhum.2016.00160

Pléh, C. (2003). Narrativity in text construction and self construction. *Neohelicon*, 30(1), 187–205. doi: 10.1023/A:1024178827479

Pragglejaz Group. (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22(1), 1–39. doi: 10.1080/10926480709336752

Richards, I. A. (1929). *Practical criticism: A study of literary judgment*. New York: Harcourt Brace Jovanovich.

Rosenblatt, L. M. (1978). *The reader, the text, the poem: The transactionaltheoryof the literary-work*. Carbondale, IL: Southern Illinois University Press.

Ryan, M. L. (2001). *Narrative as virtual reality: Immersion and interactivity in literature and electronic media*. Johns Hopkins University Press.

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. doi: 10.3758/s13428-013-0426-y

Schrott, R., & Jacobs, A. M. (2011). *Gehirn und Gedicht: Wie wir unsere Wirklichkeiten konstruieren* (Brain and Poetry: How We Construct Our Realities). München: Hanser.

Shakespeare online. n.d. Retrieved from http://www.shakespeare-online.com/sonnets/sonnet-introduction.html)

Shimron, J. (1980). Psychological processes behind the comprehension of a poetic text. *Instructional Science*, 9(1), 43–66. doi: 10.1007/BF00118968

Simonto, D. K. (1989). Shakespeare's Sonnets: A Case of and for Single – Case Historiometry. *Journal of Personality*, 57(3), 695–721. doi: 10.1111/j.1467-6494.1989.tb00568.x

Simonton, D. K. (1990). Lexical choices and aesthetic success: A computer content analysis of 154 Shakespeare sonnets. *Computers and the Humanities*, 24(4), 251–264. doi: 10.1007/BF00123412

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302–319. doi: 10.1016/j.cognition.2013.02.013

Steen, G. (1999). Analyzing metaphor in literature: With examples from William Wordsworth's "I wandered lonely as a cloud". *Poetics Today*, 20(3), 499–522.

Steen, G. (2002). Metaphor in Bob Dylan's 'Hurricane': Genre, Language, and Style. In E. Semino, & J. Culpepper (eds): *Cognitive Stylistics: Language and Cognition in Text Analysis* (pp. 183–210). Amsterdam: John Benjamins. doi: 10.1075/lal.1.10ste

Steen, G. (2004). Can discourse properties of metaphor affect metaphor recognition?. *Journal of Pragmatics*, 36(7), 1295–1313. doi: 10.1016/j.pragma.2003.10.014

Stockwell, P. (2009). The cognitive poetics of literary resonance. *Language and Cognition*, 1(1), 25–44. doi: 10.1515/LANGCOG.2009.002

Stone, P. J., Bales, R. F., Namenwirth, J. Z., & Ogilvie, D. M. (1962). The general inquirer: A computer system for content analysis and retrieval based on the sentence as a unit of information. *Behavioral Science*, 7(4), 484–498. doi: 10.1002/bs.3830070412

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. doi: 10.3389/fpsyg.2016.00969

Taruffi, L. (2016). Why We Listen to Sad Music: Effects on Emotion and Cognition. (Unpublished doctoral dissertation), Freie Universität Berlin.

The Complete Works of William Shakespeare. n.d. Retrieved from: http://shakespeare.mit.edu/.

Tsur, R. (1998). *Poetic Rhythm: Structure and Performance – An Empirical Study in Cognitive Poetics*. Bern: Peter Lang.

Tsur, R. (2006). Delivery style and listener response in the rhythmical performance of Shakespeare's sonnets. *College Literature*, 33(1), 170–196. doi: 10.1353/lit.2006.0014

Tsur, R. (2008). Deixis in literature: What isn't cognitive poetics?. *Pragmatics & Cognition*, 16(1), 119–150. doi: 10.1075/p&c.16.1.08tsu

Turner, F., & Pöppel, E. (1983). The neural lyre: Poetic meter, the brain, and time. *Poetry*, 277–309.

Ullrich, S., Aryani, A., Kraxenberger, M., Jacobs, A. M., & Conrad, M. (2016). Textual features as basis of the general affective meaning: Exploring the interplay of lexical affective features, dynamic inter-lexical shifts, and the basic affective tone in poetry. In revision.

Van Peer, W. (1986). *Stylistics and psychology: Investigations of foregrounding*. London, UK: Croom Helm.

van den Hoven, E., Hartung, F., Burke, M., & Willems, R. (2016). Individual differences in sensitivity to style during literary reading: Insights from eye-tracking. *Collabra: Psychology*, 2(1), 1–16. doi: 10.1525/collabra.39

Vaughan-Evans, A., Trefor, R., Jones, L., Lynch, P., Jones, M. W., & Thierry, G. (2016). Implicit detection of poetic harmony by the naïve brain. *Frontiers in Psychology*, 7:1859. doi: 10.3389/fpsyg.2016.01859

Vendler, H. (1997). *The art of Shakespeare's sonnets*. Cambridge, MA: Harvard University Press.

Võ, M. L., 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. doi: 10.3758/B.R.M.41.2.534

Võ, M. L., Jacobs, A. M., & Conrad, M. (2006). Cross-validating the Berlin affective word list. *Behavior Research Methods*, 38(4), 606–609. doi: 10.3758/BF03193892

Wainwright, J. (2015). *Poetry: The basics*. Routledge.

Wallace, W. T., & Rubin, D. C. (1991). Characteristics and constraints in ballads and their effect on memory. *Discourse Processes*, 14, 181–202. doi: 10.1080/01638539109544781

Wallentin, M., Nielsen, A. H., Vuust, P., Dohn, A., Roepstorff, A., & Lund, T. E. (2011). Amygdala and heart rate variability responses from listening to emotionally intense parts of a story. *Neuroimage*, 58(3), 963–973. doi: 10.1016/j.neuroimage.2011.06.077

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45(4), 1191–1207. doi: 10.3758/s13428-012-0314-x

Westbury, C., Keith, J., Briesemeister, B. B., Hofmann, M. J., & Jacobs, A. M. (2014). Avoid violence, rioting, and outrage; approach celebration, delight, and strength: Using large text corpora to compute valence, arousal, and the basic emotions. *The Quarterly Journal of Experimental Psychology*, 68(8), 1599–1622. doi: 10.1080/17470218.2014.970204

Willems, R. M. (Ed.). (2015). *Cognitive Neuroscience of Natural Language Use*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107323667

Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & Van den Bosch, A. (2015). Prediction during natural language comprehension. *Cerebral Cortex*, 26(6), 2506–2516. doi: 10.1093/cercor/bhv075

Willems, R. M., & Jacobs, A. M. (2016). Caring about Dostoyevsky: The untapped potential of studying literature. *Trends in Cognitive Sciences*, 20(4), 243–245. doi: 10.1016/j.tics.2015.12.009

Yaron, I. (2002). Processing of obscure poetic texts: Mechanisms of selection. *Journal of Literary Semantics*, 31(2), 133–170. doi: 10.1515/jlse.2002.013

Yaron, I. (2008). What is a "difficult poem"? Towards a definition. *Journal of Literary Semantics*, 37(2), 129–150. doi: 10.1515/jlse.2008.008

Zeman, A., Milton, F., Smith, A., & Rylance, R. (2013). By heart an fMRI study of brain activation by poetry and prose. *Journal of Consciousness Studies*, 20(9–10), 132–158.

Appendix

For illustrative purposes, we comment on the decision tree model A3a shown in Figure A1 in some detail here (only for the rightmost branch of the tree). The key question concerns the Affect friends and family variable of SEANCE which is a component score of nine variables, among which General Inquirer affiliation nouns (Stone et al., 1962; for details see Crossley et al., 2017 and www.wjh.harvard.edu/~inquirer/homecat.htm): Is a sonnet's value on that dimension <.49 and features a *certainty component* value of < .19, then the next question is whether the Affect friends and family value is also <.46. Is this the case (as for $N = 62$ sonnets) then the sonnet will be classified as "young man". If it's not the case (i.e., Affect friends and family value is >.46), the Social order component score (11 variables among which General Inquirer need verbs) must be <.59 for a sonnet to fall into that category ($N = 6$). Is the latter score = >.59, then the sonnet is classified as "dark lady ($N = 2$)".

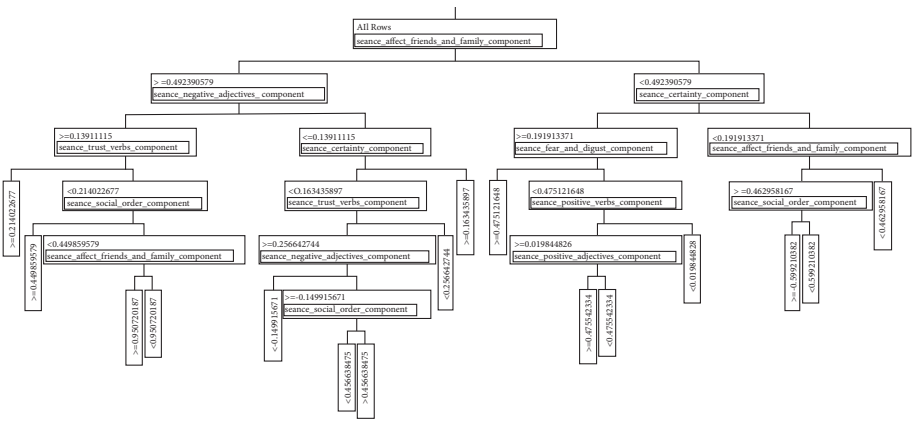


Figure A1. Decision tree for model A3a.

Table A1. Overview of all variables of the quantitative narrative analyses reported in this paper. The raw data table can be obtained from the 1st author on demand: ajacobs@zedat.fu-berlin.de

Variable	Explanation
S_ID	sonnet number/name
S_type	sonnet category (*young man* vs. *dark lady*)
Word based Variables	
word_frq_mean	mean value of word frequency according to the Subtlex-database (Brysbaert & New, 2009)
word_frq_sd	standard deviation of word frequency according to the Subtlex-database (Brysbaert & New, 2009)
word_valence_mean	mean value of all single word valence ratings according to the database of Warriner et al. (2013)

word_valence_sd	standard deviation of all single word valence ratings according to the database of Warriner et al. (2013)
word_valence_sum	sum of all single word valence ratings according to the database of Warriner et al. (2013)
word_valence_span	difference between the lowest and highest single word valence rating according to the database of Warriner et al. (2013)
word_arousal_mean	mean value of all single word arousal ratings according to the database of Warriner et al. (2013)
word_arousal_sd	standard deviation of all single word arousal ratings according to the database of Warriner et al. (2013)
word_arousal_sum	sum of all single word arousal ratings according to the database of Warriner et al. (2013)
word_arousal_span	difference between the lowest and highest single word arousal rating according to the database of Warriner et al. (2013)
word_emo_potential_mean	mean value of the product of the valence and arousal rating for all words according to the database of Warriner et al. (2013)
word_emo_potential_sd	atandard deviation of the product of the valence and arousal rating for all words according to the database of Warriner et al. (2013)
word_concreteness_mean	mean value of all single word concreteness ratings according to the Subtlex-database (Brysbaert & New, 2009)
word_concreteness_sd	standard deviation of all single word concreteness ratings according to the Subtlex-database (Brysbaert & New, 2009)
word_processing_fluency_mean	mean value of the product of the concreteness rating and word frequency for all words according to the Subtlex-database (Brysbaert & New, 2009)
word_processing_fluency_sd	standard deviation of the product of the concreteness rating and word frequency for all words according to the Subtlex-database (Brysbaert & New, 2009)
word_surprisal_subtlex_mean	mean value of all single word surprisal values computed with the algorithm of Willems et al. (2015) on the Subtlex-Corpus
word_surprisal_subtlex_sd	standard deviation of all single word surprisal values computed with the algorithm of Willems et al. (2015) on the Subtlex-Corpus
word_surprisal_shakespeare_mean	mean value of all single word surprisal values computed with the algorithm of Willems et al. (2015) on the Shakespeare-Corpus
word_surprisal_shakespeare_sd	standard deviation of all single word surprisal values computed with the algorithm of Willems et al. (2015) on the Shakespeare-Corpus
RID_ primary_p_cognition	number of words related to primary process cognition according to the Regressive Imagery Dictionary / RID (Martindale, 1975)
RID_ secondary_p_cognition	number of words related to secondary process cognition according to the Regressive Imagery Dictionary / RID (Martindale, 1975)
RID_emotions	number of words related to emotions according to the Regressive Imagery Dictionary / RID (Martindale, 1975)

Coh-metrix	
Seven Descriptive Indices	e.g. *number of words* or *word length*
(see http://cohmetrix.com/ for more details)76 Coh-metrix features	e.g. Noun overlap or results from Latent Semantic Analysis (see http://cohmetrix.com/ for more details)
Flesch Reading Ease	according to Coh-metrix
Flesch Kincaid Grade Level	according to Coh-metrix
L2 Readability	according to Coh-metrix
Eight Text Easability Principal Component Scores	e.g. Narrativity, Syntactic complexity (see http://cohmetrix.com/ for more details)
Symbolic Imagery/ Thematic Richnes/Semantic Association	
Symbolic_Imagery_Index	count of symbolic imagery categories according to Meireles (2005)
Simonton_Thematic Richness_ Index	count of 24 specific topics identified by Simonton (1989)
EAT_number_unique_associated_ WRDs_sum	sum of the number of unique associations for each word according to the Edinburgh Associative Thesaurus (EAT; Kiss et al., 1973)
SEANCE	
20 SEANCE components	e.g. component *1* (Negative adjectives) or component *6* (Affect friends and family component, Crossley et al., 2017)
SEANCE_Composite_Score	sum of the values of the 20 SEANCE components (cf. Crossley et al., 2017)
SEANCE_Negative_Mood_Score	sum of the values of the SEANCE components *1* and *7* (cf. Crossley et al., 2017)
SEANCE_Positive_Mood_Score	sum of the values of the SEANCE components *4, 5, 12* and *19* (cf. Crossley et al., 2017)

Note: Please contact the 1st author (ajacobs@zedat.fu-berlin.de) to obtain the raw data described in this table.

Author's address

Arthur M. Jacobs
Department of Experimental and Neurocognitive Psychology
Freie Universität Berlin
Habelschwerdter Allee 45
D-14195 Berlin
Germany

ajacobs@zedat.fu-berlin.de