Eye movements reveal readers' sensitivity to deliberate metaphors during narrative reading

Running head: Deliberate metaphor processing in narrative reading

Eye movements reveal readers' sensitivity to deliberate metaphors during narrative reading

Metaphors occur frequently in literary texts. Deliberate Metaphor Theory (DMT; e.g., Steen, 2017) proposes that metaphors that serve a communicative function *as* metaphor, are radically different from metaphors that do not have this function. We investigated differences in processing between deliberate and non-deliberate metaphors, compared to non-metaphorical words in literary reading. Using the Deliberate Metaphor Identification Procedure (Reijnierse, et al., 2018), we identified metaphors in two literary stories. Then, eye-tracking was used to investigate participants' (N=72) reading behavior. Deliberate metaphors were read slower than non-deliberate metaphors, and both metaphor types were read slower than non-metaphorical words. Differences were controlled for several psycholinguistic variables. Differences in reading behavior were related to individual differences in reading experience, and absorption and appreciation of the story. These results are in line with predictions from DMT, and underline the importance of distinguishing between metaphor types in the experimental study of literary reading.

Keywords: deliberate metaphor, eye-movements, literature, metaphor processing, literary reading

Introduction

Foregrounding is considered one of the central characteristics of literary texts (e.g., Leech, 2008; Mukařovský, 1932). Through stylistic deviation – for instance at the phonetic, grammatical, or semantic level – certain elements in literary texts can be made to stand out from the rest of the discourse (e.g., Miall & Kuiken, 1994). Experimental research has shown that foregrounded elements, such as alliteration, unusual word order, and metaphor, cause (some) readers to read slower and to experience parts of a text as more surprising, more striking, and so on, in comparison with 'ordinary' language use (e.g., Hakemulder, 2004; Miall & Kuiken, 1994; van Peer, 1986; van den Hoven, Hartung, Burke, & Willems, 2016; and see Jacobs, 2015a, 2015b, for a neurocognitive processing model). Much of the research into foregrounding takes into account a broad range of foregrounding devices. As a result, it is difficult to pinpoint which devices – and in which particular form or quantity – cause foregrounding effects. Studies that took the amount of foregrounding elements in a sentence into account found that this affects readers' evaluations of the aesthetic structure of the sentence, as well as their cognitive and attitudinal responses (van Peer, Hakemulder, & Zyngler, 2007), and that it influences readers' absorption into the story world (Kuijpers, Hakemulder, Tan, & Doicaru, 2014). To date, however, research that specifically examines the effects of a single foregrounding device are scarce (but see Bambini, Canal, Resta, & Grimaldi, 2018, for foregrounding effects of literary metaphor).

In this paper, we aim to contribute to a better understanding of the cognitive processing of one rhetorical figure that is typically associated with foregrounding in literary texts, namely metaphor (e.g., Leech, 2008). Metaphor is considered one of the most prominent ways in which words or phrases can be foregrounded (e.g., Semino & Steen, 2008), because they enable us to

describe one thing in terms of something else by drawing similarities between seemingly different concepts. In this study we investigate whether readers are 'sensitive' to metaphors while reading literary narratives, and if so, how this influences their experience of the narrative. Specifically, we use eye-tracking to investigate participants' reading times (gaze durations) for different kinds of metaphors. The distinctions we draw between these different kinds of metaphors is inspired by work on Deliberate Metaphor Theory (e.g., Steen, 2017), as we will introduce below. In our work we bring together insights from foregrounding research and metaphor research and extend the previous literature on both literary reading as well as metaphor processing.

Three dimensions of metaphor (in fiction)

For a long time, metaphors were regarded primarily as a form of stylistic variation or deviation from 'normal' language use, employed particularly in literary fiction and poetry (Aristotle, trans. 1940; Mukařovský, 1932). However, this perspective changed as a result of the 'cognitive turn' in linguistics (e.g. Steen, 2011). Instead of viewing metaphor as a purely stylistic device, or a matter of language alone, cognitive approaches to metaphor highlight the idea that much of our thinking is metaphorical as well, and that metaphors are actually highly prevalent in everyday language use (Lakoff & Johnson, 1980; Ortony, 1979). According to Conceptual Metaphor Theory (hereafter CMT; Lakoff & Johnson, 1980; and see Lakoff, 2014, for a recent overview of CMT), our thoughts are structured by conceptual metaphors, such as IDEAS ARE FOOD, LIFE IS A JOURNEY, and TIME IS MONEY (Lakoff, Espenson, & Schwartz, 1991).[1] These conceptual metaphors map generally more concrete source domains (such as FOOD) onto generally more abstract target domains (such as IDEAS), and are manifested in language through

[1] We use small capitals to indicate conceptual domains (e.g., Lakoff, 1993).

linguistic metaphors (e.g. "food for thought", "raw facts", "swallow a claim"; Lakoff & Johnson, 1980, p. 47). According to CMT, metaphor is not something special, but something very common in daily language use, and fundamental for how we think and act (Lakoff & Johnson, 1980).

Discourse-analytical and corpus-analytical research has confirmed the claim from CMT that metaphors are a frequent phenomenon in many different types of language use, including in educational discourse (e.g., Cameron, 2003), political discourse (e.g., Charteris-Black, 2005), business discourse (e.g., Koller, 2003), advertising (e.g., Leech, 1966), and doctor-patient conversations (e.g., Tay, 2013). Together, these and other studies show that metaphors are not just present in literature, but also in other domains (Semino, 2008). Interestingly, and counterintuitively perhaps, literary fiction does not contain more linguistic metaphors than other registers (Steen, et al., 2010). In fact, both academic prose and news texts contain relatively more linguistic metaphors than literary fiction.[2]

Despite the fact that literature is not unique in the presence of metaphor, many contemporary scholars agree that the use of metaphors in literature is important. It has been suggested that the use of metaphor in literature differs from the use of metaphors in other registers (e.g., Goatly, 1997; Leech, 2008; Semino, 2008; Semino & Steen, 2008). Metaphors in literature are deemed "more creative, original, interesting, complex and rich than those found outside literature" (Dorst, 2015, p. 3; see also Semino & Steen, 2008). That is, metaphors in

[2] These results were obtained by applying the Metaphor Identification Procedure Vrije Universiteit (MIPVU; Steen, et al., 2010) to the VU Amsterdam Metaphor Corpus, a corpus of almost 190,000 words sampled from texts from four different registers that are present in the British National Corpus. Academic prose contained the highest percentage of metaphor-related words (17.85%), followed by news texts (16.34%), literary fiction (11.49%) and everyday conversations (7.33%; Steen, et al., 2010).

literature may be used for foregrounding, and may therefore have different properties than metaphors in other registers.

To account for the different properties of metaphors, various researchers have suggested to consider metaphor not just as a cognitive and linguistic tool that allows people to think and talk about one thing in terms of another (as is the case in CMT), but to also take into account the communicative functions that metaphors may have in discourse (e.g., Cameron 1999, 2003; Goatly, 1997; Semino, 2008; Steen, 2008, 2011). Consider example (1) below.

(1) "The searchlight that illuminated everything from the fire truck, gave the scene something unreal, as if the street below us were a quay (…)."[3]

(van Essen, 2014)

Example (1) contains various words that are used metaphorically, in that they describe one thing (the target domain) in terms of something else (the source domain). A street is described in terms of a quay, and the effect that a searchlight causes (i.e. making a situation look unreal) is described in terms of a person passing something to someone ('gave'). These linguistic metaphors can be seen as expressions of underlying conceptual metaphors. For 'quay', the associated source domain concept of HARBOR comes from a different domain than the target domain concept of STREET. For 'gave', the source domain of CONTROL OVER AN OBJECT RELATIVE TO A POSSESSOR comes from a different conceptual domain than the target domain of CAUSATION (Lakoff, Espenson, & Schwartz, 1991).

[3] Translated from Dutch. The original text runs as follows: "De schijnwerper die vanaf de brandweerwagen alles bijlichtte, gaf het toneel iets onwezenlijks, alsof de straat beneden ons een kade was (…)."

'Quay' and 'gave' are similar in that they constitute linguistic expressions of underlying mappings between two distinct conceptual domains. However, the two examples are also different from each other. 'Quay' is part of an explicit metaphorical comparison between a street and a quay, as indicated by the metaphor marker 'as if'. This metaphor presents a new perspective onto the target domain situation that is being described, and as such stands out from the rest of the discourse. By contrast, 'gave' is conventionally used to describe causing an effect (e.g., Lakoff & Johnson, 1980). There is no indication in (1) that suggests that the metaphorical use of the verb 'to give' deviates from 'normal' language use. In fact, this constitutes one of the ways in which people talk about causing an effect (see also Cameron, 2003, p. 100).

To account for the difference between the two metaphors in (1), it has been suggested to add a third dimension to CMT's two-dimensional model of metaphor in language and thought, namely that of communication (Steen, 2008, 2011, 2017). The resulting three-dimensional model has led to the development of Deliberate Metaphor Theory (henceforth: DMT). At the level of communication, DMT distinguishes between deliberate and non-deliberate metaphors. Deliberate metaphors serve a communicative function *as metaphor* between language users. As such, deliberate metaphors require "distinct attention to the source domain as a separate domain of reference" (Steen 2017, p. 2), and stand out *as metaphors* in communication. Non-deliberate metaphors, by contrast, do not have such a function. Consequently, they do not require separate attention to the source domain of the metaphor, but instead stay 'on topic', i.e. within the target domain. From a communicative perspective, non-deliberate metaphors thus do not stand out *as metaphors* (Reijnierse, Burgers, Krennmayr, & Steen, 2018; Steen, 2011). The notion of deliberateness as presented by DMT can be investigated from both a semiotic and a processing perspective. These two perspectives are explained in more detail below.

The identification of potentially deliberate metaphors in language use

When taking a semiotic approach to deliberate metaphor, analysts may carry out textual analyses to determine for each linguistic metaphor in a text whether it is used *as* a metaphor in communication. To this end, a systematic and reliable identification procedure has recently been developed, called the Deliberate Metaphor Identification Procedure (DMIP; Reijnierse, et al., 2018). In brief, DMIP requires analysts to determine whether "the source domain of the metaphor is part of the referential meaning of the utterance in which it is used" (Reijnierse, et al., 2018, p. 136). If it is, the metaphor is identified as potentially deliberate.[4] If it is not, the metaphor is identified as non-deliberate. Presence of the source domain in the referential meaning of an utterance can be traced by investigating whether there are any cues present in the co-text/context of a metaphor.

The rationale behind DMIP is now briefly explained based on the linguistic metaphors in example (1) that were introduced above. More detailed step-by-step analyses illustrating the application of DMIP are provided in Appendix A. The explicit metaphorical comparison in (1) is signaled by the phrase 'as if', that introduces an autonomous source domain referent ('quay') into the discourse. Both the metaphor marker 'as if' and the source-domain perspective that is introduced into the discourse by 'quay' count as cues to suggest that both the target domain of streets and the source domain of quays are part of the referential meaning of the utterance. This is different for 'gave', for which there is no indication in the co-text (or context) of the utterance suggesting that the source domain meaning of controlling an object plays a role in the referential

[4] DMIP uses the term 'potentially deliberate' to emphasize the semiotic nature of the analysis. Potentially deliberate metaphors are hypothesized to be processed *as metaphors* by language users. However, to examine whether that is indeed the case, further psycholinguistic work needs to be carried out – as we explain in the next section.

meaning of the utterance. Although 'gave' in (1) thus counts as a metaphor at the linguistic and conceptual levels of metaphor analysis, it cannot be identified as a potentially deliberate metaphor at the communicative level. Thus, in (1) 'quay' is used *as* metaphor at the level of communication, making it a case of potentially deliberate metaphor, 'gave' is *not* used *as* metaphor at the level of communication, making it a case of non-deliberate metaphor.

In a recent corpus-linguistic study, DMIP was applied to all metaphors in the VU Amsterdam Metaphor Corpus (see footnote 2; Reijnierse, Burgers, Krennmayr, & Steen, in press). Enriching a corpus that was already coded for all linguistic metaphor with annotations for potentially deliberate metaphor enabled a comparison of the distribution of potentially deliberate versus non-deliberate metaphors in four different registers: academic prose, news texts, literary fiction, and everyday conversations. Results of this analysis showed that literary fiction contained the highest percentage of potentially deliberate metaphors of all four registers, and that significantly more potentially deliberate metaphors were present in this register than might be expected by chance (Reijnierse, et al., in press; see also Dorst, 2015; Steen, et al., 2010).[5] These results provide support for the idea that metaphors in fiction are indeed 'different' from metaphors in other registers, as was suggested earlier (see Dorst, 2015; Semino & Steen, 2008). Specifically, these results suggest that metaphors in literary fiction are more often deliberately used *as* metaphors compared to metaphors in other registers.

In the current paper, we deployed DMIP to identify all potentially deliberate and non-deliberate metaphors in two Dutch literary stories. The resulting annotations were subsequently used to investigate how readers process these types of metaphors in literary reading.

[5] Fiction contained the highest percentage of potentially deliberate metaphors (6.83% of all metaphors in the register), followed by news texts (6.63%), academic prose (2.23%) and everyday conversations (1.37%).

(Deliberate) Metaphor processing

In addition to the semiotic approach to deliberate metaphor that was discussed in the previous section, the notion of deliberateness as presented by DMT can also be investigated from a processing perspective. When approaching deliberate metaphor from a cognitive processing perspective, DMT predicts that deliberate metaphors are processed differently than non-deliberate metaphors because they are used *as* metaphors in communication between language users (Steen, 2017). Consequently, DMT hypothesizes that deliberate metaphors are processed as online cross-domain mappings between source and target domains. Non-deliberate metaphors, by contrast, are hypothesized to be processed by categorization or lexical disambiguation, much like other polysemous words (e.g., Steen, 2017).

As a developing theoretical framework, DMT is in need of research that empirically investigates the distinction between deliberate and non-deliberate metaphor (Gibbs, 2015a; Steen, 2017). To the best of our knowledge, little empirical work has been done on the cognitive processing of deliberate versus non-deliberate metaphors (see Gibbs, 2015b; Krennmayr, Bowdle, Mulder, & Steen, 2014). There is, however, a body of work using ERP or behavioral methods which suggests that metaphors are processed differently as compared to non-metaphorical language. Moreover, it has been found that novel metaphors are processed differently compared to conventional metaphors (e.g., Blank, 1988; Bowdle & Gentner, 2005; Cardillo, Watson, Schmidt, Kranjec, & Chatterjee, 2012; Lai, Curran, & Menn, 2009).

Using Event Related Potentials (ERP), various studies yielded neurophysiological evidence for an asymmetrical processing between novel and conventional metaphors. For novel metaphor comprehension, many studies have found larger N400 components (e.g., Arzouan, Goldstein & Faust, 2007; Coulson & Van Petten, 2002; Lai, et al., 2009; Rataj, 2014) than for

conventional metaphors and literal utterances. For example, Lai et al. (2009) contrasted the processing of conventional (e.g., 'Every point in my argument was attacked') and novel metaphors (e.g., 'Every second of our time was attacked') with both literal sentences (e.g., 'Every soldier in the frontline was attacked'), and anomalous metaphors (e.g., 'Every drop of rain was attacked'). They found a more negative N400 component to anomalous, novel, and conventional metaphors compared to literal sentences in the N400 time-window (320-440 ms). However, in a later time-window the ERP of conventional metaphor processing converged with the literal sentences, while the novel metaphors remained at the same level as anomalous sentences. This showed a difference in processing of novel versus conventional metaphors, but also demonstrated that conventional metaphors were not processed the same as literal sentences.

Similar results for novel metaphors versus literal utterances were reported by Bambini, et al. (2018), who investigated the time-course of literary metaphor processing in Italian poems and novels within a literary context, compared to non-metaphorical utterances of the same words. Comparing literary metaphors in the form 'A of B' (e.g., 'Grass of velvet') to their literal counterparts (e.g., 'Throne of velvet'), they found a more negative N400 component for literary metaphors than for literal expressions, which was followed by a sustained negativity (see Forgács, et al., 2012 for a related study using fMRI).

Using eye-tracking, previous research has shown that metaphorical sentences are read slower than literal sentences (Columbus, et al., 2015; Olkoniemi, Ranta, & Kaakinen, 2016). These differences were moderated by familiarity: for highly familiar metaphors, the differences in gaze durations were smaller than for unfamiliar metaphors (Columbus, et al., 2015).

The current study adds to the findings above and simultaneously aims to empirically test predictions from DMT about the differences in processing between deliberate and non-deliberate

metaphors by using eye-tracking to investigate literary reading. Based on the evidence discussed above and the predictions provided by DMT, we hypothesized that deliberate metaphors would elicit longer gaze durations than both non-deliberate metaphors and non-metaphorical words (H1). Since metaphors have been overall found to increase reading times (Columbus, et al., 2015; Olkoniemi, et al., 2016), we also expected a gaze duration difference between non-deliberate metaphors and non-metaphorical words. We hypothesized that non-deliberate metaphors would elicit longer gaze durations than non-metaphorical words (H2). We also expected that possible differences in gaze durations between types of metaphors would interact with reading habits (H3), because more experienced readers have had more exposure to metaphors in literature. Finally, we explored connections between metaphor and reading experience. To this end, we investigated whether possible differences in gaze durations between metaphor types would be influenced by individual differences in appreciation of the story, and absorption into the story. As previous research showed interactions between these two measures and foregrounding elements, we were interested in their relation to metaphor processing.

Method and materials

Participants

For this study, we used data from an earlier experiment (Mak & Willems, submitted). The topic of investigation of this earlier experiment was unrelated to the topic of the current experiment. For the earlier experiment, 109 participants (86 females) were recruited from the participant database of Radboud University Nijmegen. All participants were native speakers of Dutch and had normal, or corrected to normal vision. Some participants performed insufficiently on the comprehension check and for some of the participants the overall quality of the eye-tracking data was too poor, leading to the rejection of data from 7 participants. For another 30

participants, we did not have a complete data set for the two stories that were part of this study. It was practically challenging and very time consuming to correct this, and we instead decided to perform our analyses only with datasets that were complete. This left data from 72 participants (59 females) for the analysis. The mean age of the participants was 23.31 years (SD = 4.52; range 18 – 40).

Participants received €15 or course credit for their participation in the study. Prior to the experiment, participants were informed about the procedure of the experiment. All participants gave written informed consent in accordance with the declaration of Helsinki. The study was approved by the local ethics committee.

Materials

Initially, three Dutch literary stories were visually presented to the participants. For practical purposes (i.e., amount of time needed to code the stories for metaphor) we chose to use only two of the stories for the current study. The first story, *The people that had everything delivered* (2014; in Dutch *De mensen die alles lieten bezorgen*), was written by Dutch author Rob van Essen, and contained 2,988 words. The second story, *Signs and symbols* (1996; in Dutch *Signalen en symbolen*), was originally written in American English by Vladimir Nabokov, and translated into Dutch by a professional translator (translation has been published). This story contained 2,143 words. All participants read both stories (in counterbalanced order), which took about 10 – 15 minutes per story. None of the participants reported having read any of the two stories before.

Questionnaires

In order to measure individual reading experience, participants filled out several questionnaires. These are described below.

Absorption. Absorption was measured by an extended version of the Story World Absorption Scale (SWAS, Kuijpers, et al., 2014), containing items such as *'When I finished the story I was surprised to see that time had gone by so fast'* and *'I could imagine what the world in which the story took place looked like'*. Six additional questions regarding imagery were added to this scale to measure absorption (Mak & Willems, submitted). The SWAS has been proven to be a reliable measure (Kuijpers, et al., 2014). In total, there were 25 questions on a 7-point Likert scale, ranging from 1 = disagree to 7 = agree, that capture different aspects of absorption: Attention, Emotional engagement, Transportation, and Imagery. The scale is in Dutch.

Appreciation. Appreciation was measured by one general appreciation question (*'How much did you enjoy this story?'*) and six questions concerning enjoyment, adapted from Kuijpers et al. (2014; e.g. *'I was constantly curious about how the story would end'*). Twelve adjectives that could be used to describe the quality of appreciation (*Did you find the story… e.g. 'funny'; 'sad'; 'special'*), adapted from Knoop et al. (2016), were also added. All questions were answered on a 7-point Likert scale, ranging from 1 = disagree to 7 = agree.

Reading experience. Reading experience was measured with the Dutch version of the Author Recognition Test (ART, Stanovich & West, 1989; Koopman, 2010). This test contains a list of 42 authors, 12 of which are fake. For each name participants were able to recognize, they earned a point. For each fake name they incorrectly recognized as referring to an existing author, they got a penalty point. Participants' ART-scores were calculated as the total score of recognized authors, minus the penalty points. The ART is a well-established implicit measure for reading experience (e.g. Acheson, Wells, & MacDonald, 2008).

Metaphor identification procedure

Potentially deliberate and non-deliberate metaphors in the stories were identified using DMIP (Reijnierse, et al., 2018). This procedure consists of a step-by-step method to identify potentially deliberate metaphors in language use. DMIP is an extension of the Metaphor Identification Procedure Vrije Universiteit (MIPVU; Steen, et al., 2010). MIPVU identifies all metaphor-related words (MRW's) in texts and discourse. With DMIP, all MRW's can subsequently be analyzed for potentially deliberate and non-deliberate metaphor. In both MIPVU and DMIP the lexical unit[6] is taken as the unit of analysis. The steps of DMIP are presented below (Reijnierse, et al., 2018, p. 137).

1. Read the entire text to get a general idea of what the text is about.
2. Apply the Metaphor Identification Procedure Vrije Universiteit (MIPVU) to find all metaphorical lexical units (metaphor-related words, or MRWs; see Steen, et al., 2010, for detailed instructions).
3. Look at the first MRW.
4. Determine whether the source domain of the MRW is part of the referential meaning of the utterance in which the MRW is used.
    a. If 'yes', mark the MRW as potentially deliberate and proceed to step 5.
    b. If 'no', mark the MRW as non-deliberate and proceed to step 6.
    c. In case of doubt, mark the MRW as potentially deliberate, and add the code WIDLII (When In Doubt Leave It In; see Steen, et al., 2010). Then, proceed to step 5.
5. If the MRW is coded as potentially deliberate in step 4, describe how the source domain of the MRW is part of the referential meaning of the utterance.
6. Look at the next MRW.

[6] Lexical units typically correspond to words, but some exceptions occur. For Dutch, these include separable complex verbs and pronominal adverbs (Pasma, 2011).

We used the online version of *Van Dale Hedendaags Nederlands* to analyze the data and identify (potentially deliberate) metaphor-related words. This dictionary was chosen over the online version of *Dikke van Dale* because of its practical usability. The *Dikke van Dale* is a more elaborate dictionary that also contains information regarding archaic words and meanings. This information is not relevant for our study, as DMIP is concerned with contemporary word meanings only (Reijnierse, et al., 2018). However, in case of doubt *Dikke van Dale* was consulted as a back-up for determining more basic (i.e. source domain) meanings of lexical units.

The application of DMIP to all lexical units in our dataset yielded a total of 491 metaphor-related words (9.68% of all lexical units), 354 of which were identified as non-deliberate metaphors (6.98% of all lexical units), and 137 as potentially deliberate metaphors (2.70% of all lexical units). Further details are presented in Table

(Table 1 about here)

Note that, while the majority of lexical units were identified as non-metaphorical words, still a sizeable number of items were identified as related to metaphor (~500 lexical units, ~350 as non-deliberate and ~150 as potentially deliberate MRWs). This ensures that the regression coefficients could be estimated for each condition in the statistical analysis.

Reliability of coding. All stories were coded by the first author (the first coder). In order to ensure a reliable implementation of DMIP, there first was a training phase in which the analysis of two sample sentences was compared between the first coder and a second, experienced, coder (the second author of this paper). After the training phase, all 5,071 lexical units in the corpus were coded by the first coder. A random sample of 1,013 lexical units, which

equals 20% of the data, was also coded by the second coder to be subjected to an inter-rater reliability analysis. Of these lexical units, 13 were excluded from inter-rater reliability analysis because they were marked as either a Metaphor Flag (Steen, et al., 2010), or as non-analyzable words[7] by both coders. This left 1,000 words for the analysis. Results showed that the raters agreed on the identification of 93.8% of the lexical units as either non-metaphorical, non-deliberate metaphor or potentially deliberate metaphor. Intercoder reliability testing showed "substantial" agreement between coders (Cohen's $\kappa = .708$ [95% CI, 0.640 to 0.778]; Landis & Koch, 1977).

Experimental procedure

The experimental procedure is explained more elaborately in Mak & Willems (submitted), and will be discussed briefly here. Participants were seated in a sound-proof booth at a desk with artificial light. Participants placed their chin at a chin rest and read the stories from a computer screen. At the beginning of the experiment, participants were instructed to move their head as little as possible, but at the same time to read as naturally as they could, as they would do outside the laboratory. At the beginning of each story, participants performed a 9-point eye position calibration, and after every five sections a drift check was performed. During 1,000 ms before the next section appeared, participants fixated on a fixation cross at the point of the screen the first character of the text would appear.

Participants first read one story, and subsequently filled out the SWAS and the Appreciation questionnaire. Next, participants read a second story and again filled out the SWAS and Appreciation questionnaire. At the end of the experiment, participants filled out general questionnaires, among which a comprehension check consisting of 3 multiple choice questions

[7] These are, for instance, words in another language (in our case English), or abbreviated words.

per story, that should be easy to answer if the story was read with normal attention. If participants failed to answer two or more questions from the comprehension check per story correctly, they were rejected from further analysis. Lastly, participants filled out the Author Recognition Test (ART, Stanovich & West, 1989; Koopman, 2010), measuring reading experience.

Apparatus

Data were collected using a monocular desktop-mounted EyeLink1000plus eye-tracking system. Data were recorded with a sampling rate of 500Hz. Head movements were minimised using a head stabiliser. Participants were seated at 108 cm from the screen (i.e., distance from the eye to the bottom of the screen). The dominant eye was tracked for each participant.

Stimulus presentation

The experiment was presented using SR Research's Experiment Builder software, on a BenQ XL 2420T 24" LED screen. The stories were presented at a resolution of 1024 x 768 (32 bits per pixel), at a refresh rate of 60 Hz. The stories were divided into 30 sections each, that were presented to the participants one at a time. These sections resembled the author's original division of the story into paragraphs as much as possible. Words were presented as black letters on a white background, in a 15-point Calisto MT font corresponding to an on-screen size of 4 mm high for letters such as "m", 6 mm high for capital letters and letters such as "h", and 8 mm high for letters such as "j". For presentation of the sections, minimum margins of 120 pixels were used on all sides. Between different lines on a page, there was 24 mm white space. Participants moved to the next screen by pressing the space bar. Reading speed was hence self-paced. Interest areas were automatically determined by the Experiment Builder software as boxes per word. There was no space between interest areas, the boundaries of the interest areas

were centered between horizontally and vertically adjacent words. Figure 1 shows an example of an eye-tracking page with interest areas.

(Figure 1 about here)

Fixations shorter than 50 ms and longer than 3600 ms were excluded from all analyses (see Luke & Henderson, 2016). Also, data for all first words on eye-tracking "pages" were excluded from analysis, as previous research has shown that these fixations are disproportionately long, due to the after effect of the fixation cross (van den Hoven, et al., 2016).

Experimental design and statistical analysis

This study has a within-subject design, with metaphor type (no metaphor; non-deliberate metaphor; deliberate metaphor) as the independent variable, and gaze duration as the dependent variable. Gaze duration was defined as the sum of all fixation durations in the first pass. We included several covariates known to influence eye movements. These were included in our statistical model in order to control for variance in gaze duration to these covariates which are not of interest to the current study, but which may very well be related to the categorical distinction between metaphor types. Word frequency per word was included as a covariate in the statistical model, as previous research has shown that highly frequent words show shorter gaze durations (see Rayner, 1998, for a review). Word frequency was taken from the SUBTLEX-NL database, as the logarithm of the frequency with which a word appeared in the database (Keuleers, Brysbaert, & New, 2010). Word length (number of characters) was also added as a regressor. Because word position in a sentence influences gaze duration (Rayner, Sereno, Morris, Schmauder, & Clifton, 1989; Kuperman, Dambacher, Nuthmann, & Kliegl, 2010), this was also controlled for. We calculated relative position in the sentence as the absolute position in the

sentence divided by the total number of words in the sentence. Another control variable was Part of Speech (POS). As function words are fixated on less than content words (Carpenter & Just, 1983), we hypothesized that content words would elicit longer gaze durations than function words. The distinction between function and content words was even more relevant for our analysis, given that 83.3% of all deliberate metaphors we identified were content words, compared to 50.6% of all non-metaphorical words and 58% of all non-deliberate metaphors. All words in the two stories were automatically tagged for POS using the FROG toolkit (van den Bosch, Busser, Canisius, & Daelemans, 2007). Subsequently, the words were divided into content words (nouns, adjectives, adverbs and verbs) and function words (all other categories).

Multiple studies showed that words that are not expected to occur given their context are read slower than words that are expected (e.g. Hale, 2001; Goodkind & Bicknell, 2018). This contextual probability has been operationalized as cloze probability (Taylor, 1953; Block & Baldwin, 2010) or surprisal value (Hale, 2001), and some other studies on metaphor processing took into account surprisal value or cloze probability as a covariate. Indeed, Jacobs and Kinder (2018) observed that literary metaphors could be distinguished from non-literary metaphors in terms of surprisal value. Although we are aware that differences exist between non-metaphorical and metaphor-related lexical units in cloze probability and/or surprisal value, we decided not to include this in our model, as we consider surprisal value to be an intrinsic aspect of literary metaphors. Part of deliberate metaphors in the light of foregrounding, is to bring an unexpected element to the front.

Additionally, we investigated the interaction between the effect of deliberate metaphors and individual reading habits. Finally, we aimed to explore the possible connections between the

deliberate metaphor effect and two measures of the individual reading experience: story appreciation and story world absorption.

Our statistical analysis consisted of two separate steps. First, we did an overall group analysis in order to calculate the effects of metaphor type on gaze durations. A Linear Mixed Model was constructed, using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2014). In this model, gaze duration was taken as the dependent measure, and metaphor type, lexical frequency, POS content, relative position in the sentence, word length, and the interaction between ART score and metaphor type functioned as predictors. Subject was entered as a random factor. The predictors lexical frequency, position in the sentence, word length, and ART were centered. Metaphor type was a categorical predictors (3 levels), all other predictors were continuous.

In the second part of our analysis we investigated if and how an individual's 'sensitivity' to metaphor was related to his or her reading experience. To this end, we combined the eye-tracking data with the questionnaires per story. A Linear Mixed Model was created with the same predictors as the one above, but with the notable difference that random effects for metaphor were entered per participant, per story. Put differently, we estimated the effect of Metaphor for each story a given participant read. In this way we could ask whether an increase or decrease in reading speed for metaphors is related to higher or lower appreciation and absorption of a given story by a particular participant. The coefficients from this model were extracted from the model and then applied to new models with absorption and appreciation as predictors for the coefficients, and with subject as a random factor. In other words, in this analysis we first estimated how much a particular participant decreased or increased their gaze duration for, e.g., deliberate metaphor versus non-metaphorical words, and subsequently investigated if this

estimate was related to the reported absorption and appreciation of that particular story. The aim of this analysis is to investigate whether 'sensitivity' to metaphor (deliberate or non-deliberate) is related to appreciation or absorption. To investigate specifically the difference between deliberate and non-deliberate metaphor processing, we subtracted the coefficients of the non-deliberate vs. non-metaphor effect from the deliberate metaphor vs. non-metaphor effect, which leaves us the coefficients of the deliberate vs non-deliberate metaphor effect. The same analysis as above was then applied to these differences between coefficients.

Results

The results of the first analysis step (testing for group effects of Metaphor Type) are displayed in Table 2 and will be briefly described below.

(Table 2 about here)

As expected, lexical frequency, POS content, word length, and relative position in the sentence were significant predictors of gaze duration. More frequent words were read faster than less frequent words, shorter words were read faster compared to longer words, content words led to longer gaze durations than function words, and words at the end of the sentence elicited shorter gaze durations than words at the beginning of the sentence.

Deliberate metaphors elicited significantly longer gaze durations than non-metaphorical words. Similarly, there was a statistically significant difference between gaze durations for non-deliberate metaphors as compared to non-metaphorical words. Finally, Deliberate metaphors

were read significantly slower as compared to non-deliberate metaphors.[8] See Table 3 for the estimated marginal means (and related standard errors).

(Table 3 about here)

Lastly, individual differences in reading habits (as measured by the ART) influenced gaze duration in a negative direction. This effect interacted with the effect of metaphor type on gaze duration: more experienced readers read deliberate metaphors significantly faster than less experienced readers, compared to non-metaphorical words. For non-deliberate metaphors, there was no significant effect of reading experience on gaze durations. This effect was also not significant for the difference between deliberate and non-deliberate metaphors. The gaze durations for different types of expressions as predicted by ART score are plotted in Figure 2.

(Figure 2 about here)

For the second part of our analysis, the aim was to connect our eye-tracking results for metaphor gaze durations to the individual reading experience. As measures for reading experience, we used the Story World Absorption Scale (SWAS, Kuijpers, et al., 2014) and the Appreciation questionnaire. There was considerable variation in the scores on the SWAS and Appreciation (SWAS: *The people that had everything delivered*: $M = 4.93$, $SD = 0.77$;

[8] To account for spillover effects (cf. Rayner & Duffy, 1986), we created a model in which the metaphor coding was taken from the previous word. This model showed a positive effect of metaphor type, deliberate metaphors leading to longer gaze durations in the spillover region than non-metaphorical words as well as longer gaze-durations for non-deliberate metaphors than for non-metaphorical words, and longer gaze durations for the deliberate metaphor versus non-deliberate metaphors. The results of this model can be found in Appendix B.

Appreciation: $M = 4.60$, $SD = 0.78$; *Signs and symbols*: SWAS: $M = 3.55$, $SD = 1.02$; Apprecation: $M = 3.63$, $SD = 1.03$).

As appreciation and absorption correlate strongly (Pearson's $r = .81$, $p<.001$), separate models were created that predicted the coefficients of the metaphor effect, based on SWAS and Appreciation scores. The results of these models are displayed in Table 4.

(Table 4 about here)

As can be seen in Table 4, both SWAS scores and Appreciation scores have a negative effect on the coefficients of metaphor types. Whereas, in general, deliberate metaphors elicit longer gaze durations than non-metaphorical words, a higher degree of absorption leads to a reduction of this effect. This also holds true for the deliberate vs non-deliberate metaphor effect, and for non-deliberate metaphors compared to non-metaphorical words. These effects are plotted in Figure 3.

(Figure 3 about here)

Discussion

The aim of this study was to investigate differences in processing between deliberate and non-deliberate metaphors. Specifically, we looked at metaphors in a literary context, comparing gaze durations for the two types of metaphors with non-metaphorical expressions. We used a well-established method (DMIP; Reijnierse et al., 2018) to objectively identify potentially deliberate metaphors in two literary stories. When these results were applied to the eye-tracking data, we

found that deliberate metaphors elicited longer gaze durations than non-deliberate metaphors, and that both non-deliberate and deliberate metaphors led to longer gaze durations than non-metaphorical expressions. Note that these findings were controlled for differences in reading times to basic word characteristics such as lexical frequency, position of a word in the sentence, and whether the word was a content or a function word.

One interpretation of our findings is that deliberate metaphors are processed via online cross-domain mapping because they require attention to the source domain of the metaphor as a separate domain of reference based on their communicative function *as metaphor*. Non-deliberate metaphors, which do not function *as metaphors* at the level of communication, do not require such separate attention to the source domain, and can consequently be processed via lexical disambiguation. This difference in processing is then reflected in gaze durations, which are longer for deliberate metaphors than for non-deliberate metaphors (H1). These results are in line with predications from DMT (e.g., Steen, 2017), and can therefore be taken as empirical support for the theoretical distinction between deliberate and non-deliberate metaphors. We also found a small difference in gaze durations between non-deliberate metaphors and non-metaphorical words, indicating that it takes reader more time to process non-deliberate metaphors than non-metaphorical words (H2).

From another angle, our findings are in line with earlier research on foregrounding effects (Miall & Kuiken, 1994; van den Hoven, et al., 2016). The metaphors that we studied in the present experiment can be taken as an instance of foregrounding. Deliberate metaphors are more foregrounded elements as compared to non-deliberate metaphors (or non-metaphorical words). In line with previous work on foregrounding (Jacobs, 2015a, 2015b) we found that foregrounded elements (i.e., deliberate metaphors) led to slower reading (increased gaze durations). We also

showed that not all metaphors are automatically foregrounded, and thus are not processed as such. Metaphors that were foregrounded (i.e., deliberate metaphors) were read significantly slower than non-foregrounded metaphors (i.e., non-deliberate metaphors). These findings provide more insight into the effects of foregrounding, as the distinction on the ground of deliberateness might not be specific to metaphors, but might also hold for other foregrounding elements.

Reading habits also influenced the metaphor effect (H3). Experienced readers read deliberate metaphors faster, compared to non-experienced readers. For non-deliberate metaphors, this difference was not significant. These findings suggest that metaphor interpretation grows easier with experience. It is unclear whether general experience with words (resulting in, e.g., a larger vocabulary size; Chateau & Jared, 2000; Stanovic, West & Harrison, 1995) lies at the basis of this effect, or whether it is experience with figurative language specifically which drives the effect. Also, these effects might not be specific to metaphors, but might apply to other literary techniques as well.

A second aim of the current study was to explore the connections between the metaphor effect and the individual reading experience. Our results showed that the sensitivity to different metaphors was related to individual differences in appreciation of the stories, and absorption into the stories. Individuals that were highly absorbed, and appreciated the stories more, slowed down less when reading metaphors. If metaphors are foregrounding tools (e.g., Leech, 2008), and foregrounding is regarded as a kind of deviation from 'normal' or 'everyday' language use (e.g., Mukařovský, 1932), it makes sense that focusing on metaphors distracts the reader from being fully absorbed in the stories, as absorption requires some sort of flow or attention (cf. Kuijpers, et al., 2014). Similar results have been found earlier. Kuijpers et al. (2014) showed that stories

with more foregrounding elements elicited lower absorption. As absorption and appreciation were highly correlated (as in earlier studies, e.g. Green, Brock & Kauffman, 2004), it was no surprise that they influenced the metaphor effect in a similar way.

Aside from revealing interesting results, this study showed that combining a corpus-analytical approach and a psycholinguistic experimental approach to metaphors can be very productive. DMIP (Reijnierse, et al., 2018) enabled us to identify potentially deliberate and non-deliberate metaphors in a bottom-up way so that we could use naturalistic data, and the eye-tracking data gave us information about the online processing of deliberate versus non-deliberate metaphors.

One limitation of the current study is that we did not take into account the specific properties of deliberateness. Previous corpus-analytical and discourse-analytical research has suggested that potentially deliberate metaphors can take various forms (e.g., Reijnierse, Burgers, Krennmayr, & Steen, submitted; and see Steen, 2017, p. 16). For instance, metaphors can be identified as potentially deliberate because they are part of an extended metaphor, because they are part of a form of wordplay, or because they are novel. To disentangle the possible effects of the different properties of deliberate metaphor, a more controlled paradigm would be preferable, in which different subtypes of deliberate metaphors (such as deliberate conventional metaphors and deliberate novel metaphors) can be compared. This would give us more information about the specific contribution of these elements to the processing of metaphors (see Jacobs & Kinder, 2018, for an interesting approach). A task for future research is to come up with proper design that matches these features, while still using naturalistic stimuli such as literary stories.

In the current study, metaphors were identified on a single word level using DMIP (Reijnierse, et al., 2018). Hence, it is possible that we did not grasp the effect of deliberate

metaphor processing to its full extent, as co-text and context are very much of interest in determining the metaphorical meaning of lexical units. However, for the current study we decided to remain conservative in our metaphor coding, and only took into analysis the metaphor-related word itself.

The theoretical debate on deliberate metaphor is still ongoing (e.g., Gibbs, 2015a, b; Steen, 2017). A next step for experimental research on deliberateness is to take a closer look at different indicators of deliberateness, and how they relate to its processing. Future research should look on the exact role of Metaphor Flags (Steen, et al., 2010) and other co-textual as well as contextual cues for deliberateness, such as elaboration of the source domain, to determine the relative contribution of such aspects on metaphor processing within a literary context. Also, research in other genres (e.g., journalistic discourse, political discourse) is necessary to obtain more information on the processing of deliberate vs. non-deliberate metaphors in general.

Conclusion

This study took a first step in investigating differences in processing between deliberate and non-deliberate metaphors, based on (deliberate) metaphor identification in literary texts by means of DMIP (Reijnierse et al., 2018). In doing so, this study added to the current body of research on metaphor processing and took a more ecological approach, using naturalistic fiction (Willems & Jacobs, 2016; Jacobs & Willems, 2017). We also showed that individual differences in sensitivity to metaphors were related to story absorption and appreciation. As such, results of this study provide empirical support for DMT in the literary context, and pave the way for more research on the roles of different aspects of deliberateness.

Acknowledgements

This study was supported by a grant from the Netherlands Organisation for Scientific Research (NWO), under grant number Vidi 276-89-007. We are thankful to Alessandro Lopopolo for providing the POS tags, Marloes Mak for assistance with data analysis, and Joost Zwarts for providing feedback to an earlier version.

References

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

Aristotle (1940). *The art of poetry* (I. Bywater, Trans.). Oxford: Oxford University Press.

Arzouan, Y., Goldstein, A., & Faust, M. (2007). Brainwaves are stethoscopes: ERP correlates of novel metaphor comprehension. *Brain research*, *1160*, 69-81. doi:10.1016/j.brainres.2007.05.034

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412. doi:10.1016/j.jml.2007.12.005

Bambini, V., Canal, P., Resta, D., & Grimaldi, M. (2018). Time course and neurophysiological underpinnings of metaphor in literary context. *Discourse Processes*. Advance online publication. doi:10.1080/0163853X.2017.1401876

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). Lme4: Linear mixed-effects models using Eigen and S4. *R package version*, *1*(7).

Blank, G. (1988). Metaphors in the lexicon. *Metaphor and Symbolic Activity, 3*(3), 21-36. doi:10.1207/s15327868ms0301_2

Block, C. K., & Baldwin, C. L. (2010). Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. *Behavior Research Methods*, *42*(3), 665-670. doi:10.3758/BRM.42.3.665

Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological Review*, *112*(1), 193-216. doi:10.1037/0033-295X.112.1.193

Cameron, L. (1999). Operationalising 'metaphor' for applied linguistic research. In L. Cameron & G. Low (Eds.), *Researching and applying metaphor* (pp. 3–28). Cambridge: Cambridge University Press.

Cameron, L. (2003). *Metaphor in Educational Discourse*. London/New York: Continuum.

Cardillo, E. R., Watson, C. E., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2012). From novel to familiar: Tuning the brain for metaphors. *Neuroimage*, *59*(4), 3212-3221. doi:10.1016/j.neuroimage.2011.11.079

Carpenter P., & Just, M. A. (1983). What your eyes do while your mind is reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 275-307). New York: Academic Press.

Charteris-Black, J. (2005). *Politicians & Rhetoric*. Basingstoke: Palgrave-Macmillan.

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

Columbus, G., Sheikh, N. A., Côté-Lecaldare, M., Häuser, K., Baum, S. R., & Titone, D. (2015). Individual differences in executive control relate to metaphor processing: An eye movement study of sentence reading. *Frontiers in Human Neuroscience, 8*, 1057. doi:10.3389/fnhum.2014.01057

Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory & Cognition*, *30*(6), 958-968. doi:10.3758/BF03195780

Dorst, A. G. (2015). More or different metaphors in fiction? A quantitative cross-register comparison. *Language and Literature*, *24*(1), 3-22. doi:10.1177/0963947014560486

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

Gibbs, R.W. (2015a). Do pragmatic signals affect conventional metaphor understanding? A failed test of deliberate metaphor theory. *Journal of Pragmatics, 90*, 77-87. doi:10.1016/j.pragma.2015.05.021

Gibbs, R.W. (2015b). Does deliberate metaphor theory have a future? *Journal of Pragmatics, 90*, 73-76. doi:10.1016/j.pragma.2015.03.016

Goatly, A. (1997). *The language of metaphors*. London and New York: Routledge.

Goldstein, A., Arzouan, Y., & Faust, M. (2012). Killing a novel metaphor and reviving a dead one: ERP correlates of metaphor conventionalization. *Brain and Language*, *123*(2), 137-142. doi:10.1016/j.bandl.2012.09.008

Goodkind, A., & Bicknell, K. (2018). Predictive power of word surprisal for reading times is a linear function of language model quality. In *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2018)* (pp. 10-18).

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

Hakemulder, J. (2004). Foregrounding and its effect on readers' perception. *Discourse Processes, 38*(2), 193-218. doi:10.1207/s15326950dp3802_3

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational*

Linguistics on Language technologies (pp. 1-8). Association for Computational Linguistics. doi:10.3115/1073336.1073357

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

Jacobs, A. M. (2015b). Towards a neurocognitive poetics model of literary reading. In R. M. Willems (Ed.), *Cognitive Neuroscience of Natural Language Use* (pp. 135-195). Cambridge: Cambridge University Press.

Jacobs, A. M., & Kinder, A. (2018). What makes a metaphor literary? Answers from two computational studies. *Metaphor and Symbol, 33*(2), 85-100. doi:10.1080/10926488.2018.1434943

Jacobs, A. M., & Willems, R. M. (2017). The fictive brain: Neurocognitive correlates of engagement in literature. *Review of General Psychology*, Advance online publication. doi:10.1037/gpr0000106

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

Koller, V. (2003). *Metaphor clusters in business media discourse: A social cognition approach*. Diss. Vienna University. http://www.wuwien.ac.at/inst/english/koller_diss.pdf

Koopman, E.M. (2010). Reading the suffering of others: The ethical possibilities of 'empathic unsettlement'. *Journal of Literary Theory, 4*(2), 235-251. doi:10.1515/JLT.2010.015

Knoop, C. A., Wagner, V., Jacobsen, T., & Menninghaus, W. (2016). Mapping the aesthetic space of literature "from below". *Poetics*, *56*, 35-49. doi:10.1016/j.poetic.2016.02.001

Krennmayr, T., Bowdle, B.F., Mulder, G., & Steen, G.J. (2014). Economic competition is like auto racing. Building metaphorical schemas when reading text. *Metaphor and the Social World, 4*(1), 65-89. doi:10.1075/msw.4.1.04kre

Kuijpers, M. M., Hakemulder, F., Tan, E. S., & Doicaru, M. M. (2014). Exploring absorbing reading experiences. *Scientific Study of Literature, 4*(1), 89-122. doi:10.1075/ssol.4.1.05kui

Kuperman, V., Dambacher, M., Nuthmann, A., & Kliegl, R. (2010). The effect of word position on eye-movements in sentence and paragraph reading, *The Quarterly Journal of Experimental Psychology, 63*(9), 1838-1857. doi:10.1080/17470211003602412

Lai, V. T., Curran, T., & Menn, L. (2009). Comprehending conventional and novel metaphors: An ERP study. *Brain Research*, *1284*, 145-155. doi:10.1016/j.brainres.2009.05.088.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and Thought (2nd ed.)* (pp. 202-251). Cambridge: Cambridge University Press.

Lakoff, G. (2014). Mapping the brain's metaphor circuitry: Metaphorical thought in everyday reason. *Frontiers in Human Neuroscience*, *8*, 958. doi:10.3389/fnhum.2014.00958

Lakoff, G., Espenson, J., & Schwartz, A. (1991). Master Metaphor List. Technical report, University of California, Berkely. Retrieved from: http://araw.mede.uic.edu/~alansz/metaphor/METAPHORLIST.pdf

Lakoff, J. & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174. doi:10.2307/2529786

Leech, G. N. (1966). *English in Advertising: A Linguistic Study of Advertising in Great Britain*. London: Longman.

Leech, G. N. (2008). *Language in literature: style and foregrounding*. Harlow and London: Pearson Education Ltd.

[hier verwijzing naar Luke & Henderson 2016 toevoegen!]

Mak, M. & Willems, R.M. (Submitted). *Different kinds of mental simulation during literary reading: Individual differences revealed with eye-tracking.*

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

Mukařovský, J. (1932/1964). Standard language and poetic language. In: P.L. Garvin (Ed.), *A Prague School reader on esthetics, literary structure, and style* (pp. 17-30). Washington: Georgetown University Press.

Nabokov, V. (1996). *Verzamelde verhalen 1*. Amsterdam: De Bezige Bij.

Olkoniemi, H., Ranta, H., & Kaakinen, J. K. (2016). Individual differences in the processing of written sarcasm and metaphor: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(3), 433-450. doi:10.1037/xlm0000176

Ortony, A. (Ed.). (1979). *Metaphor and Thought*. Cambridge: Cambridge University Press.

Pasma, T. (2011). *Metaphor and register variation. The personalisation of Dutch news discourse*. Oisterwijk: Box Press.

Rataj, K. (2014). Surfing the brainwaves of metaphor comprehension. *Poznan Studies in Contemporary Linguistics*, *50*(1), 55-73. doi:10.1515/psicl-2014-0004

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372-422. doi:10.1037/0033-2909.124.3.372

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

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

Reijnierse, W. G., Burgers, C., Krennmayr, T., & Steen, G. J. (submitted). *The role of co-text in the analysis of potentially deliberate metaphor.*

Reijnierse, W. G., Burgers, C., Krennmayr, T., & Steen, G. J. (in press). Metaphor in communication: The distribution of potentially deliberate metaphor across register and word class. To appear in *Corpora 14*(3).

Reijnierse, W. G., Burgers, C., Krennmayr, T., & Steen, G. J. (2018). DMIP: A method for identifying potentially deliberate metaphor in language use. *Corpus Pragmatics*, *2*, 129-147. doi:10.1007/s41701-017-0026-7

Semino, E. (2008). *Metaphor in discourse*. Cambridge: Cambridge University Press.

Semino, E. & Steen, G. J. (2008). Metaphor in literature. In Gibbs Jr, R. W. (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 232-246). Cambridge: Cambridge University Press.

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

Stanovich, K. E., West, R. F., & Harrison, M. R. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology*, *31*(5), 811. doi:10.1037/0012-1649.31.5.811

Steen, G. J. (2008). The paradox of metaphor: Why we need a three-dimensional model of metaphor. *Metaphor and Symbol*, *23*(4), 213-241. doi:10.1080/10926480802426753

Steen, G. J. (2011). The contemporary theory of metaphor – Now new and improved! *Review of Cognitive Linguistics, 9*(1), 26-64. doi:10.1075/rcl.9.1.03ste

Steen, G. J. (2017). Deliberate Metaphor Theory: Basic assumptions, main tenets, urgent issues. *Intercultural Pragmatics*, *14*(1), 1-24. doi:10.1515/ip-2017-0001

Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., Krennmayr, T., & Pasma, T. (2010). *A method for linguistic metaphor identification: From MIP to MIPVU*. Amsterdam: John Benjamins.

Tay, D. (2013). *Metaphor in psychotherapy. A descriptive and prescriptive analysis.* Amsterdam: John Benjamins.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Bulletin*, *30*(4), 415-433. doi:10.1177/107769905303000401

Van Dale Hedendaags Nederlands Online. http://www.vandale.nl/

van den Bosch, A., Busser, B., Canisius, S., & Daelemans, W. (2007). An efficient memory-based morphosyntactic tagger and parser for Dutch. In P. Dirix, I. Schuurman, V. Vandeghinste, & F. Van Eynde (Eds.), *Proceedings of the 17th Meeting of Computational Linguistics in the Netherlands* (pp. 99-114).

van den Hoven, E., Hartung, F., Burke, M., & Willems, R. M. (2016). Individual differences in sensitivity to style during literary reading: Insights from eye-tracking. *Collabra, 2*(1): 25, 1-16, doi:10.1525/collabra.39

van Essen, R. (2014). *Hier wonen ook mensen.* Amsterdam: Atlas Contact.

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

van Peer, W., Hakemulder, J, & Zygnier, S. (2007). Lines on feeling: Foregrounding, aesthetics and meaning. *Language and Literature, 16*(2), 197-213. doi:10.1177/0963947007075985

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

Appendix

Appendix A

Step-wise analysis with DMIP of the two metaphor-related words in Example (1).

(2) "The searchlight that illuminated everything from the fire truck, gave the scene something unreal, as if the street below us were a quay (…)"[9]

(Van Essen, 2014)

For the sake of convenience, the steps of DMIP are first presented below, followed by the analyses (Reijnierse et al., 2018, p. 137).

7. Read the entire text to get a general idea of what the text is about.
8. Apply the Metaphor Identification Procedure Vrije Universiteit (MIPVU) to find all metaphorical lexical units (metaphor-related words, or MRWs; see Steen, et al., 2010, for detailed instructions).
9. Look at the first MRW.
10. Determine whether the source domain of the MRW is part of the referential meaning of the utterance in which the MRW is used.
    a. If 'yes', mark the MRW as potentially deliberate and proceed to step 5.
    b. If 'no', mark the MRW as non-deliberate and proceed to step 6.
    c. In case of doubt, mark the MRW as potentially deliberate, and add the code WIDLII (When In Doubt Leave It In; see Steen, et al., 2010). Then, proceed to step 5.
11. If the MRW is coded as potentially deliberate in step 4, describe how the source domain of the MRW is part of the referential meaning of the utterance.
12. Look at the next MRW.

[9] Translated from Dutch. The original text runs as follows: "De schijnwerper die vanaf de brandweerwagen alles bijlichtte, gaf het toneel iets onwezenlijks, alsof de straat beneden ons een kade was (…)."

Step 1 Read the entire text to get a general idea of what the text is about.

This example is part of a story about a woman who gets unwell and needs to be transported to the hospital. A fire truck comes to take her from the apartment building.

Step 2 Apply the Metaphor Identification Procedure Vrije Universiteit (MIPVU) to find all metaphorical lexical units (metaphor-related words, or MRWs; see Steen, et al., 2010, for detailed instructions).

Here, we only describe why 'quay' and 'gave' are identified as metaphor-related words. For the sake of clarity, we do not analyze the other lexical units in Example (1).

Gave

Contextual meaning: The verb 'to give' refers to causing a certain effect (Macmillan dictionary), in this case the lights make the scene look unreal.

Basic meaning: The more basic meaning of 'to give' is "to pass something to someone" (Macmillan dictionary).

Contextual meaning vs. basic meaning: The contextual meaning contrasts with the more basic meaning and can be understood in comparison with it. Causing an effect can be understood in terms of passing an object to someone.

Metaphor-related word? Yes (indirect metaphor)

Quay

Contextual meaning: The noun 'quay' refers to "a hard surface next to a sea or river, where boats can stop" (Macmillan dictionary).

Basic meaning: There is no more basic meaning of 'quay'.

Contextual meaning vs. basic meaning: The contextual meaning of 'quay' does not contrast with a more basic meaning. However, the noun is part of a direct metaphorical comparison between streets and quays, as is also indicated by the marker 'as if'.

Metaphor-related word? Yes (direct metaphor)

Step 3 Look at the first MRW.

Gave

Step 4 Determine whether the source domain of the MRW is part of the referential meaning of the utterance in which the MRW is used.

a. If 'yes', mark the MRW as potentially deliberate and proceed to step 5.

b. If 'no', mark the MRW as non-deliberate and proceed to step 6.

c. In case of doubt, mark the MRW as potentially deliberate, and add the code WIDLII (When In Doubt Leave It In; see Steen, et al., 2010). Then, proceed to step 5.

No, for a coherent representation of the referential meaning of the utterance only the target domain meaning of 'to give' is needed.

Step 6 Look at the next MRW.

Quay

Step 4 Determine whether the source domain of the MRW is part of the referential meaning of the utterance in which the MRW is used.

a. If 'yes', mark the MRW as potentially deliberate and proceed to step 5.

b. If 'no', mark the MRW as non-deliberate and proceed to step 6.

c. In case of doubt, mark the MRW as potentially deliberate, and add the code WIDLII (When In Doubt Leave It In; see Steen, et al., 2010). Then, proceed to step 5.

Yes

Step 5 If the MRW is coded as potentially deliberate in step 4, describe how the source domain of the MRW is part of the referential meaning of the utterance.

Two cues are present that suggest that the source domain of 'quay' is part of the referential meaning of the utterance. First, 'quay' is identified as a direct metaphor, which means that there is no contrast between the contextual and more basic meaning of the lexical unit. The lexical unit itself is thus not used metaphorically. However, because 'quay' is part of a metaphorical comparison (the street is compared to a quay), it expresses a mapping between two distinct domains. As a result, 'quay' is present as a source domain referent in the referential meaning of the utterance. A second cue that suggests that the source domain meaning of 'quay' is part of the referential meaning of the utterance in which it is used, is the metaphor flag (or: metaphor marker) 'as if'. This marker explicitly points out that a comparison is being made between two distinct domains.

Appendix B

Effects in the 'spillover region'. The table shows the results of the same model as presented in the main text (results in Table 2), but now using the gaze durations on the next word as dependent variable. This is an analysis on the so-called 'spillover region' which some researcher have argued is an additional interesting dependent variable to consider (see Rayner & Duffy, 1986).

	B	*SE*	*t*	*p*	
(Intercept)	243.807	3.536	68.946	<.001	***
Deliberate MRW vs non-MRW	8.150	1.612	5.056	<.001	***
Deliberate vs non-deliberate MRW	4.754	1.892	2.513	.012	*
Non-deliberate MRW vs non-MRW	3.396	1.066	3.185	.001	**
Log frequency	-8.495	0.394	-21.563	< .001	***
Relative position in the sentence	-0.950	0.266	-3.575	<.001	***
POS content	-4.929	0.679	-7.262	<.001	***
Word length	25.204	0.403	62.534	<.001	***
ART score	-10.192	3.516	-2.899	.005	**
Deliberate*ART vs non-MRW*ART	2.137	1.621	1.318	n.s.	

Deliberate*ART vs non-deliberate*ART	3.329	1.899	1.753	n.s.
Non-deliberate vs non-MRW*ART	-1.191	1.066	-1.117	n.s.

Note. For POS content, function words served as baseline.

Table 1 *Distribution of non-metaphor-related words (non-MRW), non-deliberate metaphor-related words (Non-deliberate MRW), and potentially deliberate metaphor-related words (Deliberate MRW) per story.*

	Story					
	The people that had everything delivered		*Signals & Symbols*		*Total*	
Type of metaphor	*n*	%	*n*	%	*N*	%
Non-MRW	2,717	91.85	1,863	88.17	4,580	90.32
Non-deliberate MRW	195	6.59	159	7.52	354	6.98
Deliberate MRW	46	1.56	91	4.31	137	2.70
Total	2,958	100.00	2,113	100.00	5,071	100.00

Table 2 *Results of the linear mixed model predicting gaze duration by Metaphor Type, lexical frequency, relative position in the sentence, Part Of Speech (POS content), word length and ART score.*

	B	*SE*	*t*	*p*	
(Intercept)	243.396	3.537	68.814	<.001	***
Deliberate MRW vs non-MRW	30.607	1.479	20.69	<.001	***
Deliberate vs non-deliberate MRW	8.470	1.760	16.179	<.001	***

Non-deliberate MRW vs non-MRW	2.137	1.029	2.077	0.038	*
Log frequency	-8.152	0.394	-20.687	<.001	***
Relative position in the sentence	-1.146	0.266	-4.314	<.001	***
POS content	-5.157	0.678	-7.606	<.001	***
Word length	24.733	0.403	61.332	<.001	***
ART score	-9.966	3.518	-2.833	0.006	**
Deliberate*ART vs non-MRW *ART	-3.395	1.465	-2.318	0.020	*
Deliberate*ART vs non-deliberate*ART	-1.618	1.755	-0.922	0.3564	n.s.
Non-deliberate*ART vs non-MRW*ART	-1.777	1.046	-1.698	0.089	n.s.

Note. For POS content, function words served as baseline. Significant predictors are marked (* p <.05, ** p <.01, *** p <.001).

Table 3 *Estimated marginal means and standard errors (in ms) of gaze duration for non-metaphor-related words (Non-MRW), non-deliberate metaphor-related words (Non-deliberate MRW), and potentially deliberate metaphor-related words (Deliberate MRW).*

	Non-MRW	**Non- deliberate MRW**	**Deliberate MRW**
Mean [s.e.]	248.9 [3.52]	251.16 [3.65]	279.68 [3.78]

ents predicting differences in gaze duration for non-metaphor-related words (non-MRW), non-deliberate metaphor-related word

ntially deliberate metaphor-related words (Deliberate MRW), by SWAS and Appreciation scores.

Deliberate MRW vs non-MRW					**Deliberate vs non-deliberate MRW**					**Non-deliberate MRW vs no**		
B	*SE*	*t*	*p*		*B*	*SE*	*t*	*p*		*B*	*SE*	*t*
-7.895	1.215	-6.498	<.001	***	-6.745	1.077	-6.262	<.001	***	-1.060	0.149	-7.137
-7.267	1.377	-5.276	.001	**	-6.319	1.217	-5.192	<.001	***	-1.005	0.168	-5.972

redictors are marked (* p <.05, ** p <.01, *** p <.001).

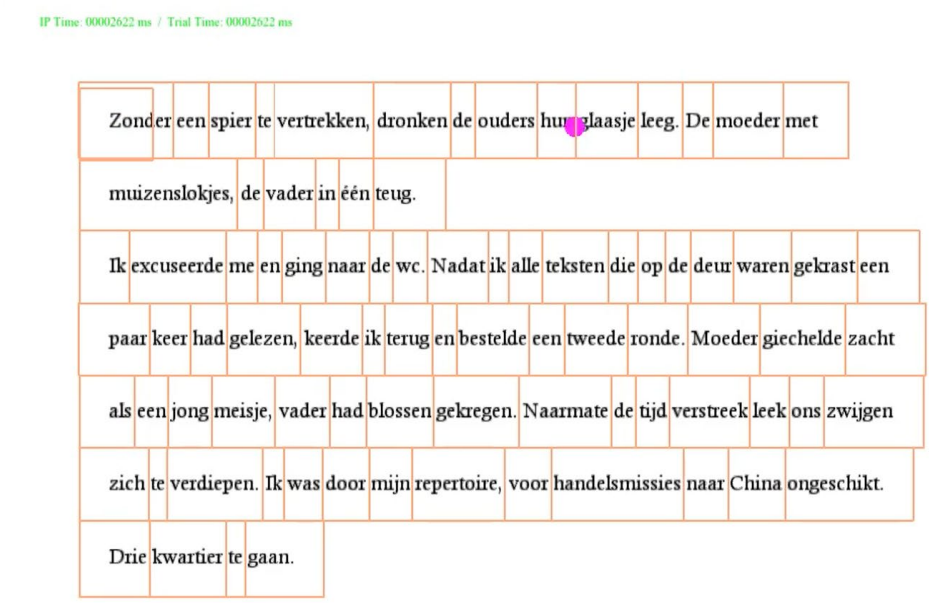

Figure 1. Screenshot of an eye-tracking 'page' with one fixation

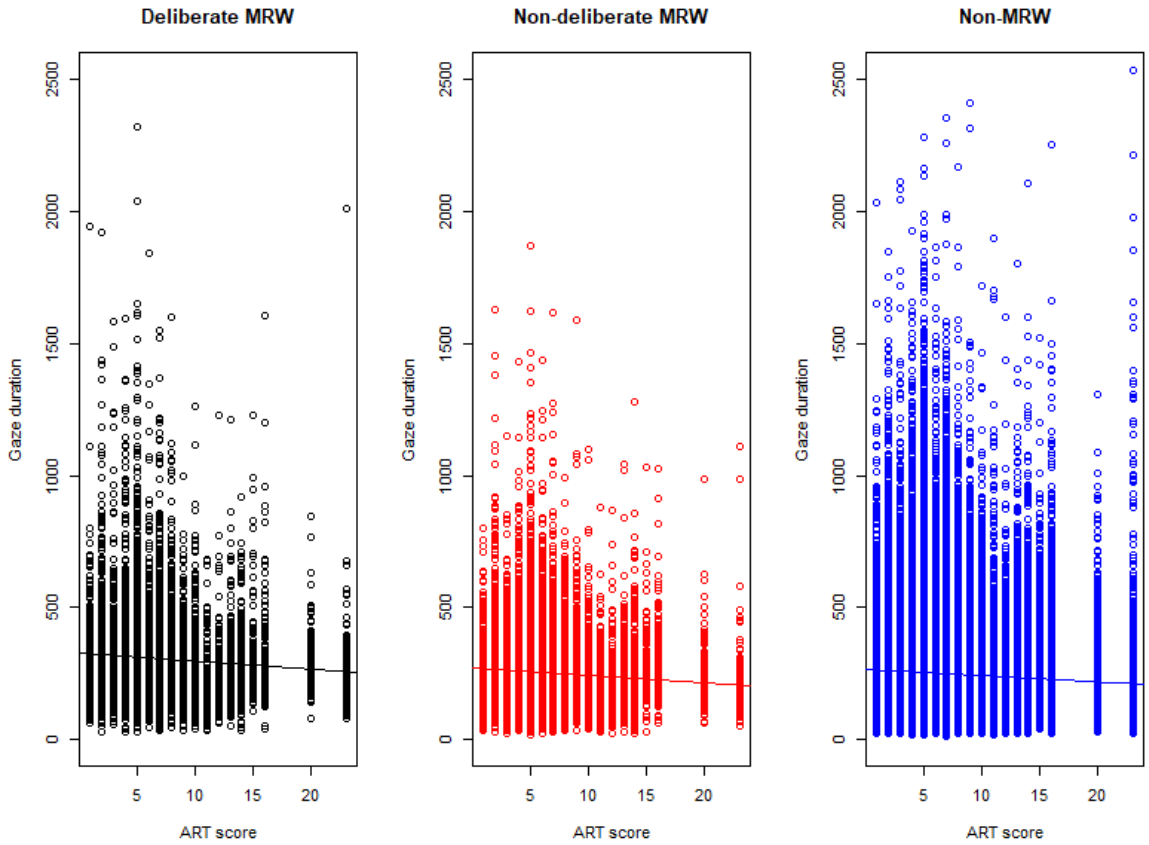

Figure 2. Scatterplots of gaze durations for potentially deliberate metaphor-related words (Deliberate MRW), non-deliberate metaphor-related words (Non-deliberate MRW), and non-metaphor-related words (non-MRW) predicted by ART scores, and regression line.

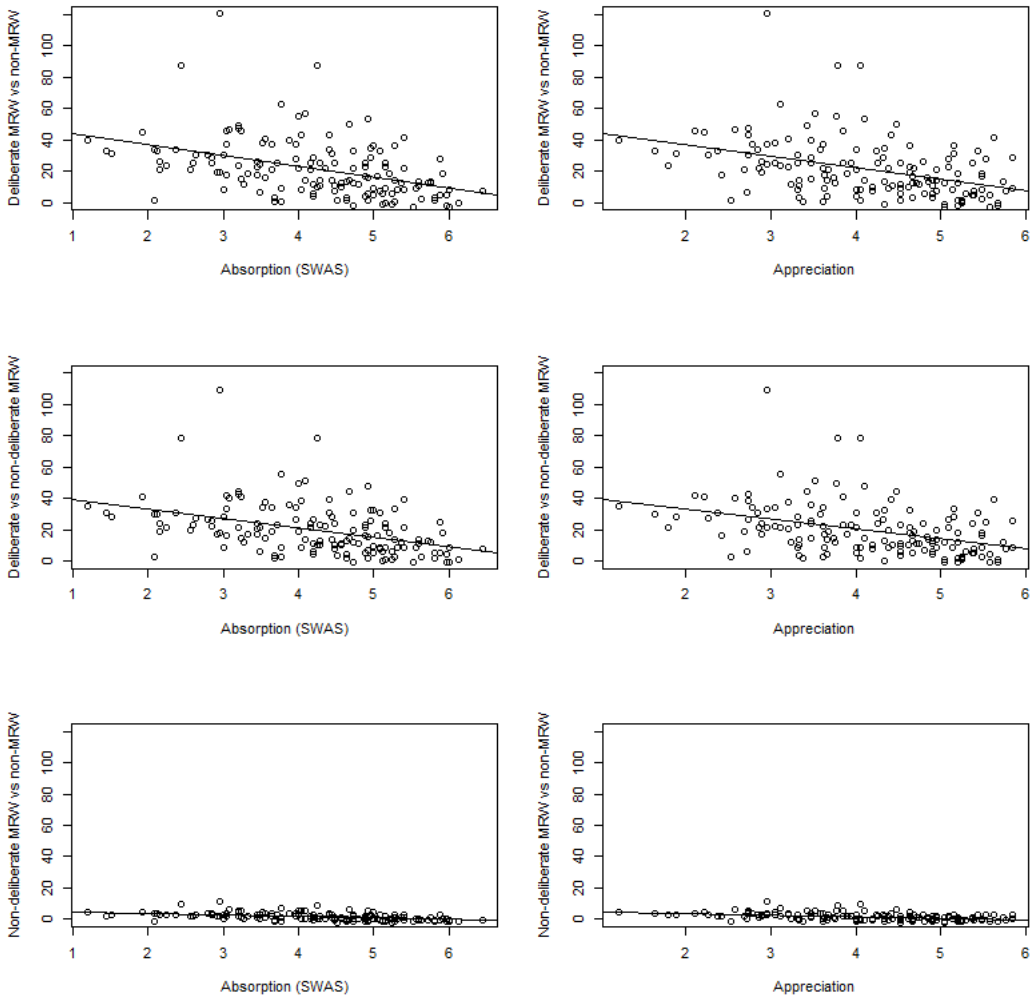

Figure 3. Scatterplots and regression lines for coefficients predicting differences in gaze duration for non-metaphor-related words (non-MRW), non-deliberate metaphor-related words (Non-deliberate MRW), and potentially deliberate metaphor-related words (Deliberate MRW), by absorption (SWAS) and story appreciation.