Universiteit Utrecht

SEEING IS UNDERSTANDING: differences between deliberate and non-deliberate metaphors revealed in eye movements during literary reading

Bachelor's thesis

Clarissa de Vries (4273559)

Opleiding: BA Taalwetenschap

Eerste begeleider: Joost Zwarts

Tweede begeleider: Roel Willems (RU)

Contents

Abstract

When reading literature, we are confronted with creative metaphors that make us see the same words in a different way. Deliberate Metaphor Theory (DMT, Steen, 2008; 2011; 2017) proposes that these kind of metaphors, that serve a communicative function *as metaphor*, are radically different from metaphors that are not used with this communicative intention. The current study investigates differences in processing between these two types of metaphors, compared to literal expressions. Using the DMIP procedure, we identified metaphors in two literary stories. Participants (N=72) read these stories. Subsequently, a Linear Mixed model was created with gaze duration as the dependent variable, and metaphor type as the independent variable. Our results show that deliberate metaphors were read slower than non-deliberate metaphors, and that both metaphor types were read slower than literal expressions. Furthermore, these effects were moderated by reading experience. Higher reading experience led to faster reading for deliberate metaphors, but not for non-deliberate metaphors and literal expressions. These results support DMT, and provide a starting point for further research on deliberate metaphor processing.

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

Introduction

For a long time, metaphors have been regarded as a stylistic variation, employed in literary fiction and poetry in a deviant manner (Aristotle, 335 BC; Mukaiovslj, 1932). With the introduction of Conceptual Metaphor Theory (CMT, Lakoff & Johnson, 1980), the view on metaphors changed: CMT takes a different perspective on metaphor, claiming that metaphor is highly prevalent in everyday language use. According to CMT, linguistic metaphors are manifestations of conceptual metaphors which form the basis of our thought. For instance, saying 'our relationship is going nowhere' reveals that we think of relationships in terms of journeys (on a conceptual level). We encounter these types of metaphors every day. So, as opposed to something special, CMT highlights the idea that metaphors are actually highly prevalent. Conceptual Metaphor Theory has inspired numerous experimental studies investigating the cognitive underpinnings of metaphor (e.g. Bowdle & Gentner, 2005; Lai, Curran & Menn, 2009; Columbus et al., 2015; Cardillo et al., 2012). A potentially important theoretical distinction which has been overlooked in the empirical work to date is that between deliberate and non-deliberate metaphors. In Deliberate Metaphor Theory (DMT), metaphors are crucially distinguished based on their communicative function (Steen, 2008, 2011, 2017; Reijnierse et al., 2017). More specifically it is proposed that metaphors will be processed cognitively in radically different ways depending on whether the metaphor is used deliberately or not by the language producer. Proponents of DMT indeed question whether highly conventionalized metaphors are processed as metaphor at all (Steen, 2017). Experimental research on deliberate metaphor is scarce (Semino & Steen, 2008; Gibbs, 2015a). A distinction which is related to deliberateness, the one between novel and familiar metaphors, *has* been investigated empirically (e.g. Bowdle & Gentner, 2005; Lai, Curran & Menn, 2009; Cardillo et al., 2012) revealing differences in processing between the two. For instance, Lai et al. (2009) used Event Related Potentials (ERP) to show that both novel and conventional metaphors lead to an increased negativity in the N400 component, compared to literal utterances. However, while conventional metaphors later converged with literal utterances, novel metaphors led to a sustained negativity, revealing differences between the two. In addition, Bambini et al. (2018) investigated the processing of literary metaphors, which are one instance of deliberate metaphors, in Italian fiction and poetry. Using ERP, they found differences in the time course of processing of literary metaphors versus literal utterances, comparable to that of Lai et al. (2009).

Here we extend previous literature on metaphor processing by explicitly focusing on the deliberate versus non-deliberate distinction, which is deemed as the defining feature whether a metaphor will be processed as a metaphor or not. We compared the looking times (gaze durations measured with eye-tracking) of deliberate versus non-deliberate metaphors while participants read short literary stories. Before we go into more detail on the current study, we will elaborate some more on the theoretical background.

Theoretical background

Metaphor in literature and poetry especially, has been studied since ancient times. A skilled writer was recognized by Aristotle, by his mastery of the metaphor: *"It is the one thing that cannot be learnt from others; and it is also a sign of genius, since a good metaphor implies an intuitive perception of the similarity in dissimilars."* (*Poetics 22*, c. 335 BCE)

Until centuries after Aristotle, a skilled use of metaphor has been seen as a literary technique that finds similarity between seemingly different entities or concepts and brings these to our attention. In literary theory, the use of metaphors has been regarded as a form of foregrounding. Foregrounding (Mukaiovslj, 1932) is a literary technique that is used to call some feature of a word or concept to the attention and can occur on different levels, for instance the phonetic level (alliterations, assimilations), or on the semantic level, of which metaphors are an example.

A different perspective on metaphors is introduced with the publication of *Metaphors We Live By* (Lakoff & Johnson, 1980). Instead of viewing metaphor as a purely stylistic device, used in the literary and the rhetoric domain, Conceptual Metaphor Theory (CMT) sees metaphor as underlying thought and language. CMT argues that our thoughts are structured by so-called conceptual metaphors, for instance IDEAS ARE CONTAINERS, LIFE IS A JOURNEY, THINKING IS EATING. These conceptual metaphors map source domains (that are in general more concrete, and related to bodily actions, such as EATING) onto target domains (these are often more abstract, in this case THINKING) and are manifested in linguistic utterances (e.g. *food for thought; raw facts; hard to swallow)* (cf. Lakoff, 2014, for a recent, introductory overview of CMT). This means that when we hear the utterances mentioned above, we unconsciously use our embodied understanding of EATING to infer the conceptual meaning of THINKING. According to CMT, metaphor is not something special or deviant, but very common in daily language use, and *fundamental* for our functioning.

The view of metaphor as foregrounding and as very common in everyday language, seem to contradict each other. CMT emphasizes the continuity between literary metaphors and conventional metaphors, arguing that the poetic metaphor is indeed different from everyday thought in that it extends and elaborates on conventional metaphors, but that it can be traced back to the same conceptual metaphors we use in everyday language (Turner & Lakoff, 1989). Proponents of a 'discontinuity view' argue that CMT cannot account for the cases where poetic metaphors are so novel and creative that they go beyond our regular metaphorical knowledge (Semino & Steen, 2008).

The use of metaphor

The two views that have been described above contrast the nature of metaphors (creative and special versus conventional). However, metaphors can also be distinguished based on their communicative function. With Deliberate Metaphor Theory (DMT), Steen (2008, 2011, 2017) stresses the importance of the communicative function of metaphors, and proposes a three-dimensional model

of metaphors, including language, thought, and communication. In this model, deliberate metaphors are distinguished from non-deliberate metaphors. Metaphors in literature may often be of the former type, serving, as Aristotle puts it elegantly, to reveal 'similarity in the dissimilar' (*Poetics 22*, c. 335 BCE). For instance, see (1).

(1) The spotlights that illuminated the scene from the fire truck, gave the scene something unreal, as if the street below us were a quay, and we were standing at the railing of a large ship on which someone had become unwell. [1]
(Fragment from *The people that had everything delivered*, Van Essen (2014).

In (1), the source domain HARBOR is mapped onto the target domain of STREET. The street is a quay, and the house (where the main character is located) is a ship. When reading this metaphor, it is obvious that the metaphor is not used as just the conventional way of expressing the situation on the street, but it is used deliberately, with the communicative function as a metaphor. As a consequence of the deliberate use of the metaphor, we are provided with an alternate perspective on the utterance and shortly attend to the source domain of the metaphorical utterance. In the example above, we can actively imagine the house as a ship that is moving on a tumultuous sea, being illuminated by headlights. This illustrates the importance of the source domain, which is the defining feature of deliberate metaphors according Reijnierse et al. (2017): 'A metaphor can be defined as potentially deliberate, when the source domain of the metaphor is part of the referential meaning of the utterance in which it is used.' (2017, p. 8). In the example above, the deliberateness of the metaphor is also indicated by the elaboration of the source domain, using the words '*railing'* and '*ship'*, and by signaling that a metaphor is about to present itself, using the words the words 'as if'. These signal words have been called Metaphor Flags (cf. Steen et al., 2010).

Deliberate metaphors can be contrasted with conventional metaphors, such as in (2), taken from a Dutch newspaper.

(2) Property tax burden rises with 2.5 percent this year.[2]
(Reijn, 2018)

In (2) the word 'rises' is used metaphorically, as it describes INCREASE OF AMOUNT (target domain) in terms of MOVEMENT UPWARDS (source domain). This is one example relating to a well-known conceptual metaphor, MORE IS UP. For this type of metaphor, it can be questioned whether the reader actively pays attention to the source domain for assigning meaning to the word

[1] Translated from Dutch. Original sentence: "De schijnwerper die vanaf de brandweerwagen alles bijlichtte, gaf het toneel iets onwezenlijks, alsof de straat beneden ons een kade was en wij aan de reling stonden van een groot schip waarop iemand onwel was geworden."

[2] Translated from Dutch. Original sentence: "Ozb-heffing stijgt dit jaar met 2,5 procent."

'rises'. In contrast with example (1) above, most of us probably do not actively imagine prices moving upwards when reading this sentence. What's more, there is already a conventionalized metaphorical meaning of 'to rise' as *'to increase in size, amount, quality, or strength'* (Macmillan English Dictionary Online). DMT proposes that non-deliberately used metaphors are processed not by a cross-domain mapping of the source domain onto the target domain, as there is no focus on the metaphorical nature of the utterance. Instead, they hypothesize that non-deliberate metaphors are processed through lexical disambiguation, much like other polysemous words (Steen, 2017).

The distinction between metaphors on the ground of deliberateness is a controversial one (Gibbs, 2011; 2015a; Müller, 2011). While DMT argues that not all metaphorical utterances are in fact processed metaphorically, Gibbs & Chen (2017) write that there is considerable evidence that even highly conventional metaphors reflect an embodied metaphorical processing. For instance, Gibbs & Nelson (2007) showed that doing or imagining body movements facilitated comprehension of metaphors that were related to this movement. Similar results have been found for other action related metaphors (Glenberg & Kaschak, 2002) and the results have also been reflected on a neurological level (e.g. Samur et al., 2015). However, experimental results have also been found that do not support a true metaphorical processing for conventional metaphors (e.g. Keysar, Glucksberg & Horton, 2000). The debate on whether or not conventional metaphors are processed as metaphors (i.e. by embodied cross-domain mapping) is still open (Chatterjee, 2010).

A second criticism is that until recently, there has not been an objective way to determine what metaphors are deliberate and what metaphors are not (Gibbs, 2011). With the introduction of the Deliberate Metaphor Identification Procedure (DMIP, Reijnierse et al., 2017), this problem is on its way to be solved.

One final critique is that DMT is in fact not at all supported by empirical evidence. This has also been acknowledged by DMT proponents (Semino & Steen, 2008) and a call for psycholinguistic research on deliberate metaphors has been made. Below, I will elaborate on the progress that has been made on the territory of experimental research on deliberate versus conventional metaphors, and other comparable distinctions, such as between novel and conventional metaphors.

The processing of metaphor

Experimental evidence on DMT is not widespread (Semino & Steen, 2008; Gibbs, 2015a), especially not in a literary context. However, substantial experimental research has been conducted on the different processing of novel versus conventional metaphors, which may be comparable to the difference between non-deliberate and deliberate metaphors. While DMT focusses on the communicative function of the metaphor, the distinction between novel and conventional focusses on the familiarity of the metaphor itself. Of course, a novel metaphor can only be used deliberately, and a conventional m metaphor may often be used non-deliberately.

One important theory reflecting differences between novel and conventional metaphor processing is Career of Metaphor Hypothesis, Bowdle & Gentner (2005). They argue that for novel metaphors, a mapping between the source and the target should be established ad hoc. As metaphors become conventionalized, this mapping is stored and transferred so that the metaphorical meaning becomes conventionalized. Behavioral evidence for this hypothesis comes from Gentner & Wolff (1997), who found that priming of the source domain led to faster metaphor recognition for novel metaphors but not for conventional metaphors, revealing an asymmetrical processing of novel and conventional metaphors, and Blank (1988) who found longer reaction times for novel than for conventional metaphors.

Neurological evidence for an asymmetrical processing between novel and conventional metaphors comes among others from a large body of ERP research. For novel metaphor comprehension, many studies have found larger N400 components (Coulson & Van Petten, 2002; Goldstein, Arzouan & Faust, 2007; Rataj, 2014; Lai et al., 2009) than for conventional metaphors and literal utterances. For example, Lai et al (2009) contrasted the processing of conventional and new metaphors with both literal sentences and anomalous metaphors. They found a more negative N400 component of anomalous, novel and conventional metaphors compared to literal sentences in the early time-window (320-440 ms). However, in the late time-window the N400 component 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.

The ERP response to metaphors could also be manipulated. Goldstein, Arzouan & Faust (2012) showed that when the meanings of metaphors were explained to the participants beforehand, N400 components for novel metaphors were reduced and, interestingly, N400 components for conventional metaphors were increased so that they resembled each other. These findings show that context is very important when investigating metaphor comprehension, which is also illustrated by Steen (1994) who showed that individuals recognize more metaphors when they believe that they are reading fiction, than in journalistic discourse.

Similarly, using eye-tracking previous research has shown that metaphorical sentences are read slower than literal sentences (Inhoff, Lima & Caroll, 1984; Columbus et al., 2015; Olkoniemi, Ranta & Kaakinen, 2016) but when presented in a longer context introducing the metaphor, these effects were reduced (Inhoff et al., 1984).

Few psycholinguistic studies have been conducted that pay attention to the communicative function that DMT stresses. Gibbs (2015b) made an attempt to empirically investigate deliberate metaphors on the linguistic level (for a critique on this study see Steen, 2015). A more ecological

approach to deliberate metaphors was taken by Bambini et al. (2018), who studied the time-course of literary metaphor processing in Italian poems and novels within a literary context, compared to their literal counterparts. They found a more negative N400 component for deliberate literary metaphors, compared to their literal counterparts, which was followed by a sustained negativity. These findings agree with previous research and suggest a long-lasting effect of figurative meaning. However, Bambini et al. (2018) only compared deliberate metaphors to literal utterances, and did not include non-deliberate metaphors.

The current study adds to the findings above, using eye-tracking to investigate differences in processing between metaphorical and literal expressions. Specifically, we aimed at uncovering differences between deliberate and non-deliberate metaphors, compared to literal utterances. To increase the ecological validity of our study, we used naturalistic stimuli in their context: Dutch short literary stories.

Based on the evidence discussed above and the arguments provided by DMT, we hypothesized that deliberate metaphors would elicit longer gaze duration than both non-deliberate metaphors and non-metaphor related words. Regarding the difference between non-deliberate metaphors and non-metaphor related words, it was expected that there would also be a difference in looking times, as DMT hypothesizes that non-deliberate metaphors are processed through lexical disambiguation, and as previous research has shown a distinction in the processing of literal versus conventional metaphorical sentences.

Secondly, we took into account individual differences in reading experience. Although deliberate metaphors are not more frequent in fiction than in other genres (Dorst, 2015), they show a different distribution in literary texts: fiction contains more direct metaphors than other genres (Dorst, 2015). As experienced readers have hypothetically had more exposure to direct metaphors than less experienced readers, we hypothesized that this would influence looking behavior, and lead to different looking times for deliberate metaphors between experienced and non-experienced. We expected that this difference would not hold for conventional metaphors, as these are not typical for literary fiction.

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. 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 quality of the eye-tracking data was too poor, leading to the rejection of 7

participants. Also, for technical purposes we only selected participants that had data points for all words in both stories for the current study, rejecting 30 participants. This left 72 participants (59 females) for the analysis. The mean age of the remaining participants was 23.31 years old (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. The study was approved by the local ethics committee.

Materials

Initially, three Dutch literary stories were visually presented to the participants. For practical purposes we chose to use only two of the stories for the current study. The selection of the stories was made for the prior experiment mentioned above (Mak & Willems, submitted), based on the length of the stories and the probability that the stories would be unknown to the target group (to ensure that all participants would read the stories for the first time). The first story, *The people that had everything delivered* (2013; in Dutch *De mensen die alles lieten bezorgen*), was written by the Dutch author Rob van Essen, and contained 2988 words. The second story, *Signs and Symbols* (1949; 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 2143 words. All participants read both stories (in counterbalanced order), which took them about 10 – 15 minutes per story. None of the participants reported having read any of the two stories before.

Metaphor identification

Metaphors in the stories were identified using the DMIP procedure (Reijnierse et al., 2017). This method uses a 5-step procedure to identify potentially deliberate metaphors, and is an extension of the MIPVU procedure (Steen et al., 2010) that identifies metaphor related words (MRW's) in texts and in discourse. As the MIPVU procedure takes the word as the lexical unit of analysis, this unit was chosen for the current study as well.

In total there were 5071 words. Of these words, determiners and connectives were always marked as literal utterances. For each remaining word, the contextual meaning in the text was first established. Next, the dictionary was used to determine if the word had a more basic contemporary meaning than the contextual meaning. We defined basic meanings as more concrete, often relating to bodily action and describing a more precise meaning (as opposed to vague), following MIPVU. When there was a more basic meaning listed in the dictionary than the contextual meaning, the word was marked as a metaphor related word (MRW). Next, it was determined whether the metaphor was potentially deliberate or not. In order to do this, the dictionary was consulted again. When the contextual meaning of the word was listed in the dictionary as a conventional meaning of the word, the word was marked as a conventional metaphor. When this was not the case, it was marked as a

potentially deliberate metaphor. I will illustrate the procedure using the example sentence from one of the stories.

(3) The spotlights that illuminated the scene from the fire truck, *gave* the *scene* something unreal, as if the street below us were a *quay*, and we were standing at the *railing* of a *large ship* on which *someone* had *become unwell*.

In (3), all words that were marked as an MRW are in italics. Of these words, both *gave* and *scene* are marked as non-deliberate, as their contextual meaning is in the dictionary as a conventionalized metaphorical meaning. For *give*, this metaphorical meaning is 'to cause a general result or effect', as opposed to the more basic meaning of 'to put something in someone's hand, or to pass something to someone' (Macmillan). For *scene*, the contextual meaning is 'a view that you can see in a picture or from the place where you are', while the more basic meaning is 'a part of a play, book, film etc. in which events happen in the same place or period of time' (Macmillan). For *quay*, no meaning is listed in the dictionary that is comparable to its contextual meaning in the sentence. What's more, the word is preceded by a metaphor flag. Consequently, this word is marked as a deliberate metaphor. In the last phrase of the sentence, the source domain is elaborated. For *standing*, it could be argued that it belongs to the source domain, but the fact that the main character was also standing in the house leads us to conclude that *standing* should be analyzed as referring to the main character, which is not the source domain. For the following words, this cannot be said. *Railing*, *large*, *ship*, *someone*, and *become unwell* all refer to entities or actions in the source domain. Therefore, they are marked as deliberate metaphors. Note here that this leads to 7 deliberate MRW's, in the case of one 'big' metaphor, which is the consequence of marking metaphors at the word level.

We used the online version of Van Dale Hedendaags Nederlands for the metaphor identification. This dictionary was chosen over the 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 we are interested in contemporary meaning. However, in some cases there was some doubt concerning the most basic meaning of the word. In those cases, the Dikke van Dale was consulted as a back-up for determining the basic meaning.

The DMIP procedure resulted in the following distribution of words over metaphor categories shown in table 1.

Table 1. Distribution of all literal utterances, non-deliberate metaphors and deliberate metaphors, per story. MRW = Metaphor Related Word.

	The people that had everything delivered		*Signals & Symbols*		*Total*	
	Number	%	Number	%	Number	%
Literal utterance	2717	91,85	1863	88,17	4580	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	2958	100	2113	100	5071	100

As can be seen in table 1, about 10% of the words is metaphorically used. Of this 10%, 7% is a conventional metaphor, and slightly less than 3% (127 words) is a potentially deliberate metaphor.

Reliability of coding. All stories were coded by the author (the first coder). To secure a reliable implementation of the DMIP method, there was first a training phase in which the analysis of two sample sentences by the first coder was compared to the analysis of a second, experienced, coder. After the training phase, all 5071 words were coded by the first coder. A random sample of 1013 words, which is 20% of the data, was then also coded by the second coder to be subjected to an inter-rater reliability analysis. Of these words, 13 were excluded from inter-rater reliability analysis because they were marked as either a Metaphor-flag, or as non-analyzable words[3] by both coders. This left 1000 words for the analysis. Results showed that the raters agreed on 93.8% of the words. A Cohen's Kappa test showed that the agreement was "substantial" (Landis & Koch, 1977) between coders (κ=0.708 [95% CI, 0.640 to 0.778], $p<.001$). These results confirmed that the DMIP procedure had been accurately implemented in the coding of the materials for the current study, and that the materials could therefore reliably be taken into the analysis.

Experimental procedure

The experimental procedure is explained more elaborately in Mak & Willems (submitted), and will be discussed briefly here. 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 can, as they would do outside the laboratory. The participants first read one story, and subsequently filled out some questionnaires relating to that story (which are not relevant for the current study). Next, the participants read a second story and again filled out questionnaires relating to that story. At the end of the experiment, participants filled out general questionnaires, among which a comprehension check consisting of 3 multiple choice questions per story, that should be easy to answer if the story was read

[3] These are for instance words in another language, in our case English, or abridged words

with normal attention. 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).

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. Interest areas were defined as boxes per word with a small margin, and were automatically determined by the Experiment Builder software. There was no space between interest areas, the boundaries of the interest areas were centered between horizontally and vertically adjacent words.

Design

This study has a within-subject design, with metaphor type (no metaphor; conventional metaphor; deliberate metaphor) as the independent variable, and gaze duration as the dependent variable. All participants (N=72) read two stories, in a counterbalanced order. Word frequency was controlled for, 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). As sentence position influences gaze duration (Rayner et al, 1989; Kuperman et al., 2010), this was also controlled for. We calculated relative position in the sentence as the absolute position in the sentence, divided by the number of words in the sentence.

Additionally, individual differences in reading experience were controlled for.
It was hypothesized that the speed of reading of deliberate metaphors would be different for the experienced reader, as he or she comes across these metaphors more often than non-experienced readers. We did not expect an interaction between reading experience and conventional metaphors, as these metaphors are also often encountered outside literature. 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 the participants were able

to recognize, they earned a point. For each fake name, they got a penalty point. The ART is a good implicit measure for reading experience.

Results

The descriptive statistics of gaze durations for the three types of expressions are summarized in table 2.

Table 2

Descriptives of gaze duration of literal expressions, non-deliberate metaphors, and deliberate metaphors.

	Literal expression	Conventional metaphor	Deliberate metaphor
Mean	251.8	250.8	306.1
Standard Deviation	138.9	133.5	184.7
Median	220	220	260
Range	4254	2694	2296

Note. Gaze duration was measured in milliseconds.

To calculate the effects of metaphor type on gaze durations, a Linear Mixed Model was constructed, using the lme4 package in R (Bates et al., 2014). Mixed Modelling allows for greater by-item and by-participant variability than the ANOVA statistics (cf. Baayen, Davidson & Bates, 2008), which is preferable for our experimental design. In this model, gaze duration was taken as the dependent measure, and Metaphor type, lexical frequency, relative position in the sentence and ART score functioned as predictors. Gaze duration, lexical frequency and sentence position were centered. The results of the model are displayed in table 3.

Table 3

Coefficients of the model predicting gaze duration by conventional metaphor, deliberate metaphor, lexical frequency, relative position in the sentence, and ART score.

	B	SE	T	P	
Intercept	0.033	0.048	0.673	.503	
Conventional metaphor	0.036	0.014	2.469	.014	*
Deliberate metaphor	0.302	0.020	14.946	<.001	***
Log Frequency	-0.180	0.002	-97.144	<.001	***
Relative position	-0.017	0.002	-8.608	<.001	***
ART score	-0.017	0.005	-2.826	.006	**
ART score*Conventional Metaphor	-0.003	0.002	-1.414	.157	
ART score*Deliberate Metaphor	-0.006	0.002	-2.602	.009	**

Note. For metaphor type, literal expressions functioned as baseline. Significant predictors are marked (* p <.05, ** p <.01, *** p <.001).

To account for spillover effects (cf. Rayner & Duffy, 1986), we created a second model in which the metaphor coding was taken from the previous word. This model showed a positive effect of both deliberate and conventional metaphors in the spillover region on gaze durations. The results of this model are inserted in the appendix.

As expected, lexical frequency and relative position in the sentence were significant predictors of gaze duration. More frequent words were read faster than less frequent words, and words at the end of the sentence elicited shorter gaze durations than words at the beginning of the sentence.

Deliberate metaphors elicited longer gaze durations than literal expressions. Conventional metaphors also elicited longer gaze durations compared to literal expressions. An additional analysis showed that there was a significant difference in gaze durations between deliberate and conventional metaphor (B= 0.238, SE=0.024, df=228.600, t=9.909, p < .001). Deliberate metaphors lead to longer gaze durations than conventional metaphors. These results are plotted in figure 1.

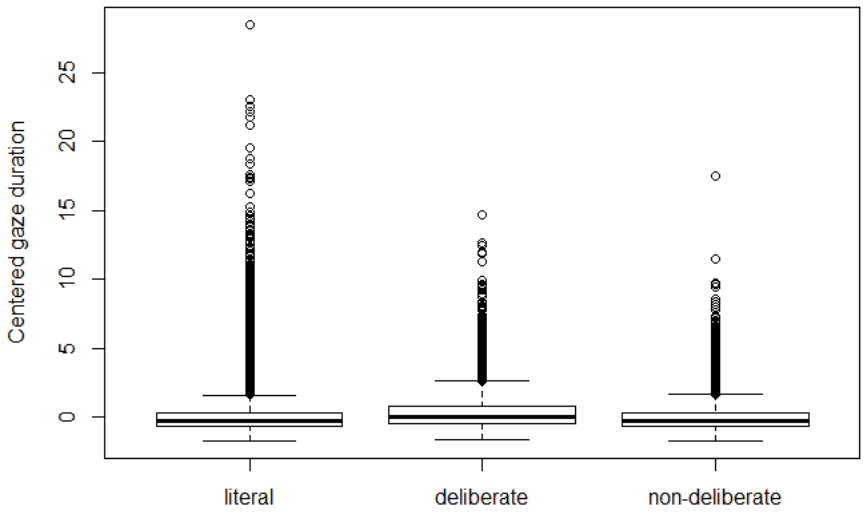

Figure 1. Boxplot of centered gaze durations for literal expressions, deliberate metaphors, and non-deliberate metaphors.

Lastly, individual differences in reading experience 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 literal words. For non-deliberate metaphors, there was no significant effect of reading experience on gaze durations. The means of gaze durations for different types of expressions as predicted by ART score are plotted in figure 2.

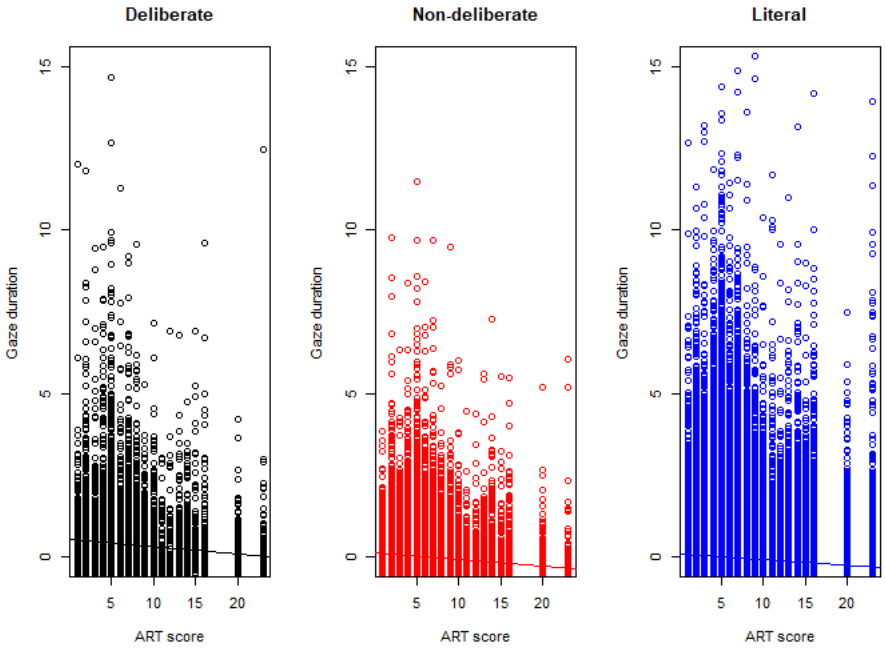

Figure 2. *Scatterplots of centered gaze durations for deliberate metaphors, non-deliberate metaphors and literal expressions predicted by ART scores, and regression line.*

Discussion

The aim of this study was to expose differences in processing between deliberate and non-deliberate metaphors. Specifically, we looked at metaphors in a literary context, comparing gaze durations for the different types of metaphors with literal expressions. The DMIP procedure proved to be a reliable method to identify deliberate and non-deliberate metaphors in literary texts. When we applied the eye-tracking data to our metaphor scoring, 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 literal expressions. These findings are in accord with earlier results from eye-tracking and ERP research (e.g. Columbus et al., 2015; Bambini et al., 2018), and provide empirical support for DMT. As we interpret our findings, they can be explained by DMT, indicating that it takes the reader time to attend to the source domain for deliberate metaphors. For non-deliberate metaphors, no such attention to the source domain is necessary, and consequently these kinds of expressions are read faster than deliberate metaphors. Regarding the difference between non-deliberate metaphors and literal expressions, it must be noted that the difference between the two was rather small.

There are a few possible confounds of our results. First of all, the metaphorically used words in our study were always content words. Non-deliberate metaphors also included function words such as prepositions, and literal expressions included even more function words. This may lead to a possible confound for word class for our results. Secondly, metaphor related words may often be less

expected to occur in a sentence than a literal word. It has been found in multiple studies 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 should be taken into account when studying metaphors. Thirdly, the distribution of words over the three types was not equal. There were more literal expressions than metaphorical expressions in our data, as we used the stories in their full length. There were enough deliberate metaphors to ensure statistical power for the analysis, but to enable a true fair comparison between the groups, the numbers should be even. One possibility to enable this, is to randomly select an even number of literal expressions, conventional metaphors, and deliberate metaphors, and take these into the analysis.

The confounds above can easily be taken into account in future investigations. However, there is one other limitation in the current research. DMT specifically places big interest in the communicative use of metaphors, but as mentioned in the introduction, it is often the case that deliberate metaphors are also novel, and that non-deliberate metaphors are also conventional. To disentangle the effects of familiarity and communicative function of the metaphor, a more controlled paradigm would be preferable, in which conventional metaphors that are used deliberately, and novel metaphors that are used deliberately can be compared. This would give us more information about the specific contribution of these elements to the processing of metaphors. A difficult task for future research is to come up with proper design that match these features, while still using naturalistic stimuli such as literary stories.

A second expectation for the current study was that the effects of expression type on gaze durations would be moderated by individual differences in reading experience. Our results show that this was indeed the case. 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. A larger vocabulary may also have been of influence, as metaphorical words are often less frequent. Also, these effects might not be specific to metaphors, but might apply to other literary techniques as well.

In the current study, metaphors were identified on a single word level using the DMIP procedure (Reijnierse et al., 2017). Hence, it is possible that we did not grasp the effect of deliberate metaphor processing to its full extent, as the context of the metaphor related word is very much of interest in determining the metaphorical meaning. However, for the current study we decided to remain conservative in our metaphor coding, and only take into analysis the metaphorical word itself.

The theoretical debate on deliberateness of metaphors is still open. 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 and other 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, disentangling the effects of familiarity and communicative function of metaphors in their processing will shed more light on the underpinnings of the deliberate metaphor.

Conclusion

This study took a first step in investigating differences in processing between deliberate and non-deliberate metaphors, using the DMIP procedure. By doing so, it added to the current body of research on metaphor processing and takes a more ecological approach, using naturalistic fiction. The results of the study provided empirical support for DMT in the literary context, and paved the way for more research on the roles of different aspects of deliberateness.

References

Aristotle (c. 335 BCE). *The Poetics (Greek: Περι`ποιητικη˜ς).* Translated by I. Bywater (1909), with a preface by G. Murray. http://www.authorama.com/book/the-poetics.html.

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*. Online first 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), 1-23.

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.

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.

Gentner, D., & Wolff, P. (1997). Alignment in the processing of metaphor. *Journal of memory and language*, *37*(3), 331-355. doi: 10.1006/jmla.1997.2527

Chatterjee, A. (2010). Disembodying cognition. *Language and Cognition*, 2, 79–116. doi:10.1515/LANGCOG.2010.004

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

Gibbs, R. W. (2011). Are 'deliberate' metaphors really deliberate?: A question of human consciousness and action. *Metaphor and the Social World*, *1*(1), 26-52. doi: 10.1075/msw.1.1.03gib.

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

Gibbs, R. W. (2015b). 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., & Chen, E. (2017). Taking metaphor studies back to the Stone Age: A reply to Xu, Zhang, and Wu (2016). *Intercultural Pragmatics*, *14*(1), 117-124. doi: 10.1515/ip-2017-0005

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic bulletin & review*, *9*(3), 558-565. doi: 10.3758/BF03196313

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).

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

Holyoak, K. J., & Stamenković, D. (2018). Metaphor comprehension: A critical review of theories and evidence. *Psychological bulletin*. Advance online publication. doi:10.1037/bul0000145

Inhoff, A. W., Lima, S. D., & Carroll, P. J. (1984). Contextual effects on metaphor comprehension in reading. *Memory & Cognition*, *12*(6), 558-567. doi: 10.3758/BF03213344

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

Keysar, B., Shen, Y., Glucksberg, S., & Horton, W. S. (2000). Conventional language: How metaphorical is it?. *Journal of Memory and Language*, *43*(4), 576-593. doi: 10.1006/jmla.2000.2711

Koopman, E. (2015). 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

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

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, J. & Johnson, M. (1980). *Metaphors we live by*. Chicago, Illinois: The University Of Chicago Press.

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

Macmillan English Dictionary Online. http://www.macmillandictionary.com/

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

Mukaiovslj, J., 1964 [1932]. Standard language and poetic language. In: P.L. Garvin (ed.), *A Prague School reader on esthetics, literary structure, and style*, 17-30. Washington, DC: Georgetown University Press.

Müller, C. (2011). Are 'deliberate' metaphors really special? Deliberateness in the light of metaphor activation. *Metaphor and the Social World*, 1(1), 61–66. doi:10.1075/msw.1.1.06mul

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

Rataj, K. (2014). Surfing the brainwaves of metaphor comprehension. *Poznan Studies in Contemporary Linguistics*, 50(1), pp. 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.

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

Reijn, G. (2018, February 27). Ozb-heffing stijgt dit jaar met 2,5 procent. CBS: 'Gemeenten stemmen tarieven af op gewenste opbrengst'. *De Volkskrant.* Retrieved from https://www.volkskrant.nl/economie/ozb-heffing-stijgt-dit-jaar-met-2-5-procent-cbs-gemeenten-stemmen-tarieven-af-op-gewenste-opbrengst~a4574050/

Reijnierse, W. G., Burgers, C., Krennmayr, T., & Steen, G. J. (2017). DMIP: A Method for Identifying Potentially Deliberate Metaphor in Language Use. *Corpus Pragmatics*, 1-19. doi: 10.1007/s41701-017-0026-7

Samur, D., Lai, V. T., Hagoort, P., & Willems, R. M. (2015). Emotional context modulates embodied metaphor comprehension. *Neuropsychologia*, *78*, 108-114. doi: 10.1016/j.neuropsychologia.2015.10.003

Semino, E. & Steen, G. Metaphor in Literature (pp. 232-246). In Eds. Gibbs, R., Jr. (2008). *The Cambridge Handbook of Metaphor and Thought.* 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

Steen, G. (1994). *Understanding metaphor in literature: An empirical approach*. New York, NY: Longman Publishing.

Steen, G. (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. (2015). Developing, testing and interpreting deliberate metaphor theory. *Journal of pragmatics*, *90*, 67-74. doi: 10.1016/j.pragma.2015.03.013

Steen, G. (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., Krennmayr, T., & Pasma, T. (2010). *A method for linguistic metaphor identification: From MIP to MIPVU* (Vol. 14). John Benjamins Publishing.

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

Turner, M., & Lakoff, G. (1989). *More than cool reason: A field guide to poetic metaphor*. Chicago, Illinois: University of Chicago Press.

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

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

Appendix

Appendix 1

Coefficients of the model predicting gaze duration by conventional metaphor, deliberate metaphor, lexical frequency, relative position in the sentence, and ART score.

	B	SE	t	p	
(Intercept)	0.040	0.048	0.828	.410	
Log frequency	-0.186	0.002	-101.331	< .001	***
Relative position	-0.015	0.002	-7.877	<.001	***
Deliberate metaphor	0.047	0.022	2.130	0.033	*
Non-deliberate metaphor	0.040	0.015	2.720	0.007	**
ART score	-0.017	0.006	-2.891	0.005	**
Deliberate metaphor*ART score	0.003	0.003	1.233	0.218	
Non-deliberate metaphor*ART score	-0.002	0.002	-1.078	0.281	

Note. For metaphor type, literal expressions functioned as baseline. Significant predictors are marked (* p <.05, ** p <.01, *** p <.001).