Journal of Pragmatics 215 (2023) 131–144

Contents lists available at ScienceDirect

Journal of Pragmatics

journal homepage: www.elsevier.com/locate/pragma

Attention to the source domain of conventional metaphorical expressions: Evidence from an eye tracking study

Ana Werkmann Horvat [a], Marianna Bolognesi [b,*], Nadja Althaus [c]

[a] Faculty of Humanities and Social Sciences, University of Osijek, L. Jagera 9, 31000 Osijek, Croatia
[b] University of Bologna, Italy
[c] University of East Anglia, UK

ARTICLE INFO

Article history:
Received 10 February 2023
Received in revised form 16 July 2023
Accepted 21 July 2023

Keywords:
Attention
Metaphor processing
Eye tracking
Deliberateness

ABSTRACT

This study investigates whether the metaphorical status of conventional expressions can be reactivated when elements of the source domain are present in the context. In indirect metaphors the source domain (or literal meaning) is not expressed (e.g., *The father cut the budget*). The literal meaning of cutting remains latently encoded in the predicate and readers' attention is not required to move from the finance domain to the domain of physical cuts. Such conventional metaphoric expressions are likely to be processed via lexical disambiguation of a polysemous (metaphorical) verb. Using an eye tracking combined with a forced-choice semantic relatedness task we investigated whether by adding linguistic material referring to the source domain (e.g., *father cut the budget like grass*), we can direct readers' attention to the source domain of the metaphorical predicate and stimulate them to interpret conventional metaphorical expressions by means of cross-domain mapping. The results indicate that in the reactivated condition participants dwell on the object (*budget*) significantly longer in their second run and when they regress to it after the final region than where there is no source domain activation. These findings may offer new insight into the limited experimental evidence related to the deliberate metaphor theory.

1. Introduction

The power of metaphorical language in driving readers'/hearers' attention is particularly interesting where one term of the metaphor is presented in the message, and the other one is cued by the context. These are the most frequent cases of metaphors found in language, and they are commonly called indirect metaphors (Steen et al., 2010). For instance, by saying that *a budget has been cut*, we mean that financial support has been limited. We use the verb *cut* metaphorically (nothing has been literally cut). Such expression is quite conventional, up to the point that it may hardly be processed by thinking of literal cuts, involving scissors or other tools. The source domain of this metaphor (in which the literal meaning of cutting would fit) is likely to remain unaddressed by the readers/hearers' attention.

In metaphor research, there exists strong evidence that conventionalized metaphoric expressions are processed without a lot of cognitive effort (see Werkmann Horvat et al., 2021 for an overview), however, there has not been much empirical research investigating whether processing of these types of expressions is different when the source domain is activated.

* Corresponding author.
E-mail addresses: awerkmannhorvat@ffos.hr (A. Werkmann Horvat), m.bolognesi@unibo.it (M. Bolognesi).

https://doi.org/10.1016/j.pragma.2023.07.011

Conversely, novel metaphoric expressions are likely to attract hearers/readers' attention to the source domain of the metaphor, precisely because of their novelty. If instead of cutting the budget we read *prune the budget*, a much less common expression (according to enTenTen20), we are more likely to direct our attention to the literal meaning of pruning, before constructing an interpretation of its metaphorical meaning in the given context.

Yet, language seems to be able to give us the tools for driving hearers/readers attention to the source domain of a metaphoric expression also when this is particularly conventionalized, and therefore it would be naturally understood without the need of directing our attention to the source domain. For instance, one can say *the budget was cut like grass* and this additional linguistic material referring to the literal meaning of cutting may push the readers/hearers to address the literal meaning of cutting, a meaning that they would not otherwise address if the message remained within the target domain of financial budgeting.

The aim of the present study is to understand whether and how the metaphorical status of a conventional expression can be reactivated when elements of the source domain are present in the context. The results of this study are relevant because they may complement the limited empirical evidence related to the deliberate metaphor theory (see next section).

2. Theoretical background

2.1. Deliberate metaphor theory: the basics

The processes involved in metaphor comprehension have been traditionally interpreted in two ways. On one hand, the classic comparison view suggests that metaphors are interpreted by means of a horizontal alignment of the two terms of the metaphor, and the activation of mappings between the source and the target (Lakoff and Johnson 1980; Lakoff et al., 1999; Gentner, 1983; Gentner et al., 2001; Gentner and Markman, 1997; Clausner and Croft, 1997; Grady, 1997). For example, when we read the statement *she devoured the novel*, we would understand the metaphorical use of *devour* thanks to a comparison with the literal meaning of this verb. In its literal sense *devour* is typically used in association with food and voracity, while in the metaphorical sense it can be used in the human domain, in association with books and reading eagerly. Both domains of the metaphor are active in the mind of the speaker during comprehension, according to the comparison view.

In contrast to the comparison view, a second type of process, based on semantic categorization, has been identified and investigated in the literature (Glucksberg and McGlone, 2001; Glucksberg and Keysar, 1990; Glucksberg, 2008). According to the categorization view, the target of the metaphor is treated as a member of a more generic, superordinate category, which is exemplified by the source of the metaphor. For example, the metaphorical meaning of *devour* becomes a hyponym of the generic category that could be summarized as "taking in something quickly and eagerly to meet an urgent need". The two meanings of *devour* play different roles in this comprehension process; the literal meaning provides a superordinate category that can be used to characterize the metaphorical meaning, which in turn constrains the features by which it can be characterized, and blocking irrelevant features (for example, the use of the mouth in the action of intaking). In this perspective, the comprehension of a metaphor proceeds through a process of abstraction.[1]

Because both mechanisms (comparison and categorization) have been supported by empirical data (e.g., Bowdle and Gentner, 2005 for a review), and differences in reaction times have suggested that they both can take place (with categorization processes being faster and more automatic than comparison ones), it became necessary to identify a variable that would determine in which cases a metaphor is processed by comparison, and in which cases it is processed by categorization.

Three main theories have been proposed to solve this puzzle: Bowdle and Gentner (2005) proposed that the degree of a metaphor's conventionality can explain whether a metaphorical statement is processed by means of comparison or by means of categorization, with novel metaphors being processed by comparison and conventional ones by categorization (Bowdle and Gentner, 2005; Gentner and Bowdle, 2008). Glucksberg and Haught (2006) and Jones and Estes (2006) suggested that the variable that can explain whether and when a metaphor is processed by comparison or by categorization is aptness. This is defined as the extent to which the category to which the source of the metaphor belongs, captures an important feature of the target (Blasko and Connine, 1993; Chiappe and Kennedy, 1999; Jones and Estes, 2006). In this view, metaphors that are more apt (novel or conventional ones alike) are processed by categorization, while metaphors that are less apt need to be processed by comparison. Both these views are supported by empirical data, as further discussed in Bolognesi and Werkmann Horvat, (2023).

More recently, a third view has been suggested to account for the two different but co-existing comprehension processes. This third view sees in the metaphor's deliberateness the reason for the activation of each of the two comprehension processes (Steen, 2008a, 2008b, 2011a, 2011b, 2013, 2015, 2016, 2017). A metaphor is deliberate when it is produced with the intent of instructing the addressee to "to adopt an 'alien' perspective on a target referent so as to formulate specific thoughts about that target from the standpoint of the alien perspective" (Steen 2013:180). Deliberate metaphor use is, therefore, "the intentional use of a metaphor as a metaphor" (Steen 2015:67). Steen (2008) emphasizes that, when discussing the use of metaphor, a three-dimensional model needs to be considered, i.e., linguistic, conceptual and communicative dimensions all

[1] Note that the comparison and the categorization views have been traditionally exemplified in relation to direct metaphors, in which both terms are laid out (e.g., *my lawyer is a shark* or *my job is a jail*), while here we use an indirect metaphor, i.e., the polysemous verb *devour* (for the distinction between direct and indirect metaphors, see Steen et al., 2010).

play an important role in understanding the diverse ways in which people use and understand metaphor. The variable deliberateness, in this sense, belongs to the communicative dimension of metaphor use: if a metaphor is used intentionally *as a metaphor*, then it is a deliberate metaphor. Concerning the processing of metaphors, Steen (2008) identifies a paradox: most metaphors are, in fact, not processed metaphorically. Deliberate metaphors are more likely to be processed by means of comparison (thus by paying attention to the source domain), while non-deliberate metaphors are more likely to be processed by means of categorization (thus by not paying attention to the source domain). For instance, when describing aspects related to the Covid pandemic military language has been used very often (Wicke and Bolognesi, 2020; Wicke and Bolognesi, 2021). Expressions like *fighting the invisible enemy*, which are technically metaphorical, are likely to be processed by means of categorization because of their conventionality, and they are probably used as non-deliberate metaphors: speakers do not intentionally use these expressions *as metaphors*. Yet, by signaling linguistically their deliberate use as metaphors, readers may be stimulated to pay attention to the source domain of the war and its related components. For instance, one may add a creative extension to the conventional metaphor, as in *fighting the invisible enemy with biomolecular weapons*, or by adding discursive signals that trigger readers' attention to the expression's metaphoricity, as in *literally* fighting or *really* fighting or *physically* fighting.

The value of deliberate metaphor theory (DMT) is hotly debated, with some scholars failing to find empirical evidence in support of this theory and thus suggesting that this is not a viable route to distinguish between processing strategies (notably, Gibbs, 2015a; 2015b). Other scholars, instead, have proposed reinterpretations of existing empirical evidence in a perspective that supports DMT (e.g., Cuccio, 2018), and outlined sets of linguistic signals that would provide evidence of deliberate metaphor use in discourse (Krennmayr, 2011). Due to the limited experimental results concerning DMT, more research is needed that explores the effect of linguistic elements that signal a deliberate use of metaphor in metaphor processing strategies.

2.2. Deliberate metaphor theory and language processing

Deliberateness differentiates those metaphors that are used with the intent of changing the reader's/hearer's standpoint on a given topic, from those metaphors that are used without this specific intent. As Steen argues, deliberate metaphors work as "perspective changers" (Steen, 2016: 116), in that they provide an external perspective on a topic, which stimulates the reader/listener to draw attention to the source domain referent of the metaphor. To achieve this communicative goal, deliberate metaphors are very often signaled in the discourse (De Vries et al., 2018). Similes are a typical example of how a figurative comparison may be used deliberately in discourse because of the explicit comparison between source and target domain referents (Reijnierse, 2017).

Consider for instance the statement *the salesman dropped the price like a bomb.* In the first part of this statement the verb *dropped* is used metaphorically, in association with the object *price*, which cannot be literally dropped. The metaphor *dropped the price*, alone is not signaled in the discourse, and it is lexicalized in language and thus quite conventional. These features suggest that it is a non-deliberate metaphor, which is used in discourse without the specific communicative purpose of changing the listener's standpoint on this financial move. However, the second part of the statement, *like a bomb*, has a precise communicative function: that of revitalizing the literal meaning of dropping, by adding an additional direct object (the bomb) which can be literally dropped. This communicative operation, arguably leads the reader/listener to recast her interpretation of the verb *dropped*, activate the literal meaning of this verb, which is selected by the new object (*bomb*), and eventually re-process the metaphor *dropped the price* as a metaphor, and thus by cross-domain comparison. The signal introduced in the discourse, therefore, suggests a deliberate use of the (conventional) metaphoric expression *dropped the price*.

In terms of processing, while the first part of the statement (containing a conventional metaphor), according to some theories (e.g. Bowdle and Gentner 2005 and similar), is likely to be processed by categorization, and thus quickly and without paying attention to the literal meaning of the verb *dropped*, the second part of the statement is constructed in such a way that the conventional metaphoric expression is revitalized and the literal meaning of *dropping* gets activated, so that the conventional metaphoric expression might be re-processed by means of comparison between the literal and the metaphorical meanings of the verb *dropped,* each selected by an object (the bomb and the price).

The predictions about the activation of the source domain (the literal meaning of *dropped* in this example) during metaphor processing can be usefully addressed in experimental psycholinguistic research, using eye tracking techniques. In a pioneering study on this topic, De Vries et al. (2018) used the Deliberate Metaphor Identification Procedure to identify potential deliberate metaphors in two literary stories and then investigated the reading behavior of 72 participants who were asked to read these literary stories. These texts were previously analyzed and within them the authors had identified deliberate metaphors, non-deliberate metaphors and non-metaphorical words. The authors found that deliberate metaphors were read significantly slower than non-deliberate metaphors and they found significant differences between more and less experienced readers. Moreover, non-deliberate metaphors were read more slowly than non-metaphorical words. The difference in reading times in this case was significant but this needs to be taken with caution ($p = 0.03$, compared to $p < 0.001$ for the difference in reading times between deliberate and non-deliberate metaphors). The slower reading times for deliberate compared to non-deliberate metaphors are interpreted as a sign that participants processed deliberate metaphors by means of cross-domain comparison: extra time is required to focus the attention on the source domain of the metaphor as a separate domain of reference. Conversely, non-deliberate metaphors do not function as metaphors at the level of communication, and can be processed via simple lexical disambiguation, and therefore do not require extra processing time.

These preliminary results, based on two literary texts, suggest that deliberateness is a variable that may produce different types of processing strategies. In line with these preliminary results, our aim is to further investigate the different processing strategies potentially involved in reading deliberate and non-deliberate metaphorical expressions.

2.3. The current experiment

Based on DMT, conventional metaphorical expressions that are not deliberate are likely to be processed via categorization. Conversely, metaphorical expressions that are produced with the intent of drawing readers' attention to the metaphor per sé, as a metaphor, are likely to be processed by means of cross-domain mapping, and therefore with more cognitive effort. We assume that cross-domain mapping involves more cognitive effort since connections need to be formed between the two domains, while this should not be the case for processing via categorization.

This paper investigates whether contextual clues from the source domain can draw attention to the metaphorical status of a very conventionalized metaphorical expression, and whether this affects metaphorical processing. To explore this, we use an eye-tracking reading paradigm combined with a semantic relatedness task. This approach allowed us to investigate the attention directed at different regions during reading but also the activation of various meanings in the semantic relatedness task.

In this experiment, we explored a particular case of deliberate metaphor, one that is signaled by a linguistic cue following the metaphor-related word. This experiment assumes incremental linguistic processing, i.e., the interpretation of a given unit is assumed to change its meaning as the context changes so initial processing (by lexical disambiguation/categorization) might change to more complex processing (by cross-domain mapping/comparison) once additional linguistic material is encountered. We presented participants with sentences in which a conventional metaphor is either reactivated (reactivated condition) or not (neutral condition) by a linguistic expression introduced by the signal *like*, followed by a noun or adverbial expression that was semantically related to the literal meaning of the verb or neutral (i.e., not related to either of the meanings, usually a temporal phrase or similar). See Table 1 below for examples.

Table 1
Critical sentence conditions with regions.

Sentence condition	Regions				
	Noun	Verb	Object	Like	Final noun/adverb
Reactivated	The father	cut	the budget	like	grass
Neutral	The father	cut	the budget	like	last month.

Participants were invited to read the sentences and then press a button that led them to a semantic relatedness task. Here they saw two nouns on the screen. These words were either not related to the preceding sentence (UNR) or related to the literal meaning of the sentence topic (LIT) or related to the metaphorical meaning of the sentence topic (MET). All of the three possible combinations of these occurred in the experimental trials (LIT/UNR, MET/UNR, LIT/MET), and depending on the preceding sentence there was always one possibility that was the target, i.e., the expected answer (LIT in LIT/UNR; MET in MET/UNR; MET in LIT/MET), and one that was the distractor, i.e., the answer that is not expected. The instruction was to press a button corresponding to the item that appeared to them most related to the previously presented sentence. For instance, after reading *The father cut the budget like grass* participants would see two of the following three words: *deficit* (related to the metaphorical meaning of cutting, labelled as MET), *meadow* (LIT) or *beach* (UNR).

Our general research questions are the following:

1. Is the metaphorical status of a conventional metaphorical expression reactivated when elements of the source domain are present in the context?
2. Does metaphorical reactivation cause processing delays and reading regressions?

Our predictions consider two possible outcomes:

1. The null hypothesis, which suggests that the literal meaning of the conventional metaphoric expression is not reactivated by the presence of elements from source domain because the metaphoric expression is conventionalized and therefore more likely to be processed by means of semantic categorization. In this case, we expect to observe no differences in the processing of the reactivated and neutral sentences, namely T*he father cut the budget like grass* (reactivated) and *The father cut the budget like last month* (neutral). We will be looking at total dwell times, as well as at first and second run dwell

times[2] in the verb region, object region, *like* region and the final region. As for the semantic relatedness question, we outline our predictions in Table 2.

2. The alternative hypothesis, which suggests that the literal meaning of the polysemous verb presented in a highly conventionalized metaphoric expression (e.g., *cut the budget*) is reactivated by the presence of words that evoke the source domain of the metaphor, placed at the end of the sentence. This would push the readers to go back and dwell longer on the verb after a first reading, showing a reactivation of the literal meaning of this verb, a focused attention on the source domain, and arguably a processing strategy by comparison and cross-domain mapping. We predict that in this case participants will exhibit different processing strategies (i.e., longer regressions and longer dwell times in the reactivated condition compared to the neutral condition) once the elements of the source domain are encountered. As for the forced choice relatedness question, we outline our predictions in Table 2. In this table, we word our predictions in terms of whether more or less effort needs to be invested to select the target choice. Here we assume that the lower number of expected answers and looks directed towards the target signify greater effort and potentially different processing strategies.

Table 2
Predictions for forced choice question.

Sentence	Target & distractor	Prediction
The nurse digested the news like milk. (reactivated)	Media (MET) & dairy (LIT)	**if metaphor not reactivated**: effortless to select MET **if metaphor reactivated:** effortful to decide which is related
The singer fed his ego like a pig. (reactivated)	Pride (MET) & dome (UNR)	**if metaphor is not reactivated**: effortless to select MET due to the lack of competition from UNR **if metaphor is reactivated**: effortless to select MET due to the lack of competition from UNR
The father cut the budget like grass. (reactivated)	Lawn (LIT) & magic (UNR)	**if metaphor is not reactivated:** effortful to say which is related **if metaphor is reactivated:** effortless to select LIT
The politician sold his soul like every term. (neutral)	Spirit (MET) & silver (LIT)	**metaphor is not reactivated:** effortless to select MET
The murderer spread the fear like last month. (neutral)	Danger (MET) & bush (UNR)	**metaphor is not reactivated:** effortless to select MET
The victim buried the emotion like the first time. (neutral)	Funeral (LIT) & golf (UNR)	**metaphor is not reactivated:** effortful to decide which is related

In Table 2 we assume that if there is a reactivation of the literal meaning, and therefore the conventionalized metaphoric expression is revitalized in the reactivated condition, then the readers have to activate both domains in their mind: the target domain (e.g., the finance domain), as well as the source domain (e.g., the garden domain). Because both domains are active in their mind, both words related to the literal and metaphorical meanings are meaningful and get into competition. This may result in a smaller number of expected responses and less looks to the target. In the cases where it is easier to decide due to lack of competition between plausible word candidates, we assume more expected responses and more looks to the target, as described in Table 2.

3. Methods

3.1. General notes

In this experiment, eye movements were tracked while participants read sentences containing a conventional metaphorical expression. After the sentences, participants saw two words on the screen and were instructed to press the button to say which of these words is related to the sentences they read. We have also recorded reaction times and answers to questions. This means that when interpreting the experimental conditions, we will be relying both on the eye tracking data from the reading task and the answers to the relatedness question (accuracy and eye tracking data).

Participants were recruited at two different locations. The eye-trackers used at both locations were Eyelink1000, with Experiment Builder and Data Viewer software packages. The monitor used at the first location was Viewsonic G90FB, while the monitor used at the second location was a Dell Ultrasharp U2412M monitor with settings changed so the size of the screen matches the monitor used at the first location (4:3 aspect ratio). The ethics approval for this study was obtained through the Ethics board of the Authors' institutions.

[2] Total dwell time represents the total time the participant spends fixating on a given region, the first run dwell time is the time participant spends on a given region the first time they encounter it, the second run dwell time is the time the participant spends on a given regions the second time they encounter it. Total dwell time tells us about the overall level of attention and importance that a reader gives to a specific region, while first-run and second-run dwell times offer more specific information about the reader's attentional priorities and processing strategies.

3.2. Participants

We recruited 48 native speakers of English aged 18–50 with normal or corrected to normal vision, with no known language, neurological or hearing disorders. The participants were compensated for their time at the usual University rate. The participants were tested on two different locations (each location n = 24) using the same experimental stimuli and eye-tracker, and same size screens.[3]

3.3. Stimuli creation and experimental design

The critical stimuli consisted of 24 sentences across two conditions: reactivated condition and neutral condition. Strictly speaking, the critical parts were not the whole sentences, but the endings of the sentences since this is where the conditions differed. The reactivated condition is a condition in which a conventional metaphor is reactivated by using elements from the source domain in the context. An example is*: The woman devoured the novel like pizza,* where *devoured the novel* is the metaphorical expression, and *like pizza* is the element from the source domain. The neutral condition included a conventional metaphor with no elements from the source domain. An example of this is: *The woman devoured the novel like the first time.* In the reactivated condition the last phrase of the sentence was always a noun, while in the neutral condition the last phrase of the sentence was always a temporal phrase or an adverb (see Table 3 for examples). The conventional metaphorical expressions consisted of a polysemous verb and its object (from Werkmann Horvat et al., 2021). This type of structure was used in Steen (2008:229) to exemplify sentences in which metaphorical expressions might be processed in different ways depending on the deliberate use of elements from the source domain. After reading the sentence, the participants had to make a choice between two possible options: the choices were either connected to the source domain, the target domain or unrelated to the metaphor in the sentence. For example, in the case of *devour the novel,* the choices that participants would see were *book* (related to the target domain, i.e., to the metaphorical meaning), *food* (related to the source domain, i.e., to the literal meaning) and *door* (unrelated).

Table 3
Examples of stimuli with the choice screen.

Sentence condition	Sentence					Target-related	Source-related	Unr.
	1	2	3	4	5			
reactive.	The father	**cut**	**the budget**	like	grass	debt	lawn	magic
reactive.	The man	**climbed**	**the ranks**	like	a fence.	navy	wire	storm
reactive.	The nurse	**digested**	**the news**	like	milk.	media	dairy	seat
neutral	The father	**cut**	**the budget**	like	last month.	debt	lawn	magic
neutral	The man	**climbed**	**the ranks**	like	in his youth.	navy	wire	storm
neutral	The nurse.	**digested**	**the news**	like	never before.	media	dairy	seat

The words that were given as choices for the relatedness task were all balanced across conditions for the various lexical measures, showing no significant difference on a single-factor ANOVA test for length in CELEX [$F(2, 69) = 0.748$, $p = .477$], Kučera and Francis Corpus [$F(2, 69) = 1.532$, $p = .223$], and BNC [$F(2, 69) = 0.103$, $p = .902$], showing no significant difference between the frequencies of the words across the three conditions in any of the traditional frequency corpora (target-related, source-related, unrelated), as well as exhibiting no significant differences in length across the three conditions [$F(2, 69) = 0.838$, $p = .437$].

We also ran a norming study in Qualtrics to check for relatedness of target-related terms to the metaphorical meaning of the expression, and source-related terms to the literal meaning of the expression. However, note that participants never actually saw any literal expressions in the eye tracking experiment (they saw the *like* phrase), but we believed it was necessary to check whether the source-related terms were judged to be semantically related to the literal meaning of the expression. This means that the expressions included in the norming study were pairs of 1) metaphorical expression + target-related word (*digest news* + *media*), 2) literal expression + source-related word (*digest milk* + *dairy*) 3) metaphorical expression + unrelated word (*digest news* + *seat*) 4) literal expression + unrelated word (*digest milk* + *seat*).

Participants were given the following instructions: *For each phrase mark how related is its meaning to the word on its right on a scale from 1 to 5, where 1 is not at all related, and 5 being extremely related.* On the screen, the participants saw a randomized list of pairs of phrases and words and had to mark on a scale from 1 to 5 how related these are. We recruited 61 native speakers of English for this study. We were interested in seeing whether there is a significant difference between how related metaphorical expressions, e.g., *cut budget*, are to target-related words, e.g., *debt*, ($M = 4.32$, $SD = 0.45$) vs. literal expressions, e.g., *cut grass*, to source-related words, e.g., *lawn*, ($M = 4.45$, $SD = 0.48$). There was no significant difference between relatedness judgements on a paired t-test ($t = 2.069$, $p = 0.285$). We were also interested in seeing whether there are differences between how related unrelated terms, e.g., *magic*, are to metaphorical ($M = 1.24$, $SD = 0.20$) and literal expressions ($M = 1.20$, $SD = 0.19$). A paired t-test showed there were no significant differences between relatedness judgements ($t = 2.069$, $p = 0.314$).

[3] For a brief discussion on the effect of location on the experimental manipulation, see the files on OSF.

Overall, there were two sentence conditions and three choice conditions, i.e., the experimental design was 2 × 3. Six experimental lists were created, and each participant only saw a single sentence in only one condition once. The order of the sentences was randomized, and the lists were balanced with respect to the position of the choice words they saw (left or right on the screen). Before starting with experimental trials, participants saw 3 practice sentences. Each participant also saw 24 filler sentences of the following type: *Katy cleaned the house like her mother.*

3.4. Procedure

After providing written consent, the participants were given the following instructions: *You will see sentences on the screen. After the sentence disappears two words will appear on the screen, one on the left, one on the right. You will need to choose which one of these two words is connected in meaning to the sentence you've just read. You will do this by pressing the buttons LEFT or RIGHT on the keyboard.* The participants were seated in front of the eye tracker at approximately 65 cm distance from the screen. A 5-point calibration procedure was conducted. After this they were presented with 51 trials (see Fig. 1). Each trial began with a fixation cross displayed for 1000 ms. The location of the fixation cross was the left edge of the sentence on the following screen. This was done to ensure that participants did not fixate on words in the middle of the sentence prior to reading the sentence from the beginning. The target sentence was displayed immediately after the offset of the fixation cross and remained on the screen for 3000 ms. The sentence display was followed by a fixation cross that was shown centrally for 1500 ms, after which two target words were shown side by side, horizontally centered at 0.25 and 0.75 of the screen width, respectively, and vertically centred in the middle of the screen. The words remained on screen for 3000 ms and participants' manual responses during this time were recorded. After the offset of the words the screen remained blank for 100 ms before the next trial began.

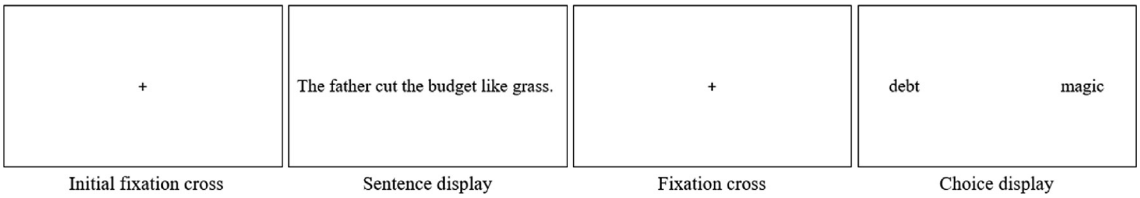

Fig. 1. Example of an experimental trial.

4. Results

4.1. Methods of analysis

The data were analyzed using linear mixed-effects models for the eye tracking data (dwell times and fixation sequence analysis) and a general linear mixed-effects model with a binomial distribution for relatedness answers with the lme4 package, version 1.1.-27 (Bates et al., 2015) in R, version 4.1.0. (R Development Core Team, 2011).[4] All fillers were excluded from the analysis, but no additional trials were removed.

In general, for all linear mixed-effects models, we included random effects of subjects and items on the intercepts, as well as random effects of subjects on the slopes for the sentence portion of the trial. In the cases in which the model failed to converge or where there were singularity issues, we removed the random effects of subjects on the slope.[5] For each analysis, we created a base model, which included an intercept and the two random factors (subject, item). In a second step, we then added a fixed effect of sentence condition. These two models were compared to see whether the addition of sentence condition effect improved the model fit. The “best fit” is the most complex model for which a significant improvement over the previous one was found. The Chi Square figure in the following LME tables refers to the result of the model comparison, i.e., base model vs. the model with added effects. For instance, in Table 5 this means that for the dwell times in the verb region, the sentence condition has no effect, while it does have an effect on the dwell times in the object region.

4.2. Sentence screen: eye tracking data

4.2.1. Dwell time analyses

For dwell times analyses in the sentence part of the trial, we fitted separate models for the dependent variables Total Dwell Time, First Run Dwell Time and Second Run Dwell Time, for each of the regions of interest *Verb*, *Object*, *Like* and *Final*. For each of these we conducted stepwise model comparisons as described above, with the last model including sentence condition as

[4] R code available on OSF: https://osf.io/7yfau/?view_only=9e599bdc665a4a0d8b3d283fe2cc5229.

[5] For complex models, where the number of parameters is too large to be reliably estimated with the given data set, the fitting procedure often fails to converge to an optimal fit. The resulting model is difficult to interpret and model simplification is therefore preferable (Barr, 2013).

the fixed effect, and subjects and items as random effect, as well as by-subject slopes for sentence condition. For the *Final* region we also included the fixed effect of length. Preliminary analyses showed that this was the dominating factor, which was added to the model before the fixed effect of sentence condition to avoid an intermediate model without the fixed effect of length which could not be a good fit and would therefore yield a misleading outcome. In Table 5 we report outcomes for those regions of interest (ROIs)/runs where we predict there might be differences caused by the experimental manipulation, i.e., we do not report the results for the first run and total dwell time in ROIs *Verb*, *Object* and *Like* since this is before the participants encountered the critical part of the sentence. The full analysis for these regions/runs is available at: https://osf.io/7yfau/?view_only=9e599bdc665a4a0d8b3d283fe2cc5229, while in Table 5 we only report the best fit models.

In Table 4 we report the raw mean dwell times for all relevant regions and runs. It shows how long on average the participants spent looking at certain regions (i.e., Verb, Object, Like, Final) in different runs (total, 1st run, 2nd run) in the two conditions (reactivated/neutral). In total we report 6 best fit models in Table 5 (each compared to a simpler model through a stepwise comparison, details available on OSF). Random effects of subjects were removed from the slopes due to a singularity issue in the following runs: verb region (second run dwell time), object region (second run dwell time), and due to convergence failure in the *like* region (second run dwell time[6]).

Table 4
Raw mean dwell times for regions and conditions.

Region	Total Dwell Time		1st run Dwell Time		2nd Run Dwell Time	
	Reactivated	Neutral	Reactivated	Neutral	Reactivated	Neutral
Verb	443.86	427.61	215.13	215.89	155.27	148.31
Object	562.67	455.30	346.62	312.44	168.97	122.20
Like	295.15	262.99	275.82	259.99	244.30	248.20
Final	317.54	460.69	258.34	343.73	56.07	99.30

Table 5
Mixed model analysis for all regions for dwell times in the order of reading (sentence condition reactivated is compared to the base level of sentence condition neutral).

Reg	Fixed effects	Estimate	Std. Error	df	t value	Pr(>|t|)
Final	Total Dwell Time					
	Model 1 (best fit): with fixed effect Length					
	Compared to base model: $X^2(1) = 30.93$, $p < .0001$					
	(Intercept)	9.21	59.70	56.27	0.15	.88
	Length	37.93	5.36	33.69	7.08	<.001
	1st Run Dwell Time					
	Model 1(best fit): with fixed effect Length					
	Compared to base model: $X^2(1) = 24.30$, $p = <.001$					
	(Intercept)	26.31	50.52	52.52	0.52	.60
	Length	27.45	4.63	33.32	−5.93	<.001
	2nd Run Dwell Time					
	Model 1 (best fit): with fixed effect Length					
	Compared to the base model: $X^2(1) = 12.30$, $p = .0005$					
	(Intercept)	−2.43	22.46	46.44	−0.11	.91
	Length	8.14	2.08	30.69	3.91	.0004
Verb	2nd Run Dwell Time					
	Model 0 (best fit, no effect of SentCond)					
	(Intercept)	151.96	12.81	50.25	11.86	<.001
Object	2nd Run Dwell Time					
	Model 1(best fit): with fixed effect SentCond					
	Compared to base model: $X^2(1) = 18.95$, $p = <.001$					
	(Intercept)	121.39	13.31	70.68	9.12	<.001
	SentCondReactivated	45.17	10.33	1082.48	4.37	<.001
Like	2nd Run Dwell Time					
	Model 0 (best fit, no effect of SentCond (Intercept)	72.76	6.53	36.23	11.14	<.001

As Table 5 shows, in the verb region the sentence condition has no effect on dwell times, which is demonstrated by the lack of improvement in fit in the model comparison, i.e., the model without any fixed effects is the best model. This means that the experimental manipulation had no effect on dwell times. In the object region, by contrast, the sentence condition affects dwell times, i.e., adding the fixed effect improved the model fit. The effect of sentence condition is significant ($p < 0.001$), with

[6] While the removal of random slopes from just some models introduces a level of inconsistency, it is better to be conservative for those cases where models do not converge, while at the same time keeping the more precise models for those cases where models converge. Since the difference is not in the fixed effects per se and the differences between the simplified models and the maximal models are very small (without differences in significance), we believe that this is an acceptable solution.

the reactivated condition resulting in longer looking in the second run. This is seen from the results in Table 5, but also in line Object, column 2nd Run Dwell Time of Table 4, where it can be seen that the reactivated condition is looked at longer (168.97 ms) than the neutral condition (122.20 ms). In the *like* region the sentence condition also does not have an effect in the second run, i.e., the model without the addition of any fixed effects is the best model. Finally, the best fit models for the ROI *Final* for total dwell time, first and second run were the ones with just a fixed effect of length. This means total and first run dwell time in the ROI are determined mostly by the number of words in the region.

4.2.2. Fixation sequence analyses

We also conducted a fixation sequence analysis (FSA). This type of analysis is mainly used to examine saccades directed from a certain interest area to other interest areas. We were interested in fixations coming from any of the words of the final region into different words of the metaphorical expression, i.e., the *Verb* and *Object* regions, see Fig. 2. Fixations in an early region after saccades from the *Final* region are especially interesting to us because their duration indicates how long participants spend on the relevant target region after having read the final part of the sentence, in which the metaphor is reactivated (or not).

We followed the previously described procedure with respect to the analysis, i.e., the data were analyzed using linear mixed-effects models. We included subjects and items as random effect, as well as by-subject slopes for sentence condition. In the cases in which the model failed to converge or where there were singularity issues we removed the random effects on the slopes. For each LME model, we created a base model, which included an intercept and the two random factors (subject, item). To the base model, we added a fixed effect of sentence condition. These two models were compared to see whether the addition of sentence condition effect improved the model fit. The dependent variable was the dwell time. This is the time that the participants spent looking at the verb or the object after coming back to it from any of the words of the final region. Slopes were simplified for the model for FSA verb because the model did not converge.

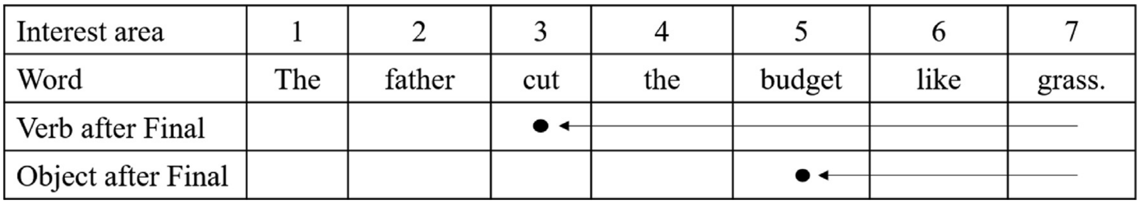

Fig. 2. Regions/words targeted by the FSA analysis.

Table 6
Mixed model analysis for the fixed effect of condition for fixation sequence areas (DV = dwell times).

Area	Predictor	Estimate	Std. Error	df	t value	Pr(>|t|)
FSA	Best fit model (no effect of SentCond):					
VERB after FINAL	(Intercept)	18.43	3.28	52.29	5.617	<0.001
FSA	Best fit model (with fixed effect SentCond):					
OBJECT after FINAL	Compared to base model: $X^2(3) = 15.23$, $p = 0.001$					
	(Intercept)	10.81	2.22	48.00	4.86	<0.001
	SentCondReactivated	9.15	3.16	48.12	2.89	0.006

The FSA analysis results (Table 6) showed that there was no effect of the experimental condition on the verb, meaning that the duration of looking at the verb when regressing from the final region did not differ between reactivated and neutral sentences. This means that in the case of the verb region, the experimental manipulation did not play a role in how long the participants spent on the verb after they would see the final region. On the other hand, the duration of looking at the object did differ between the two conditions, with longer looking at the object region after the final region in the reactivated condition ($p = 0.006$). This is consistent with the dwell time analysis, which showed a significant effect of sentence condition for the object region in the second run, and no significant effects for dwell time in the Verb region. The FSA comprises solely of fixations that are made after a saccade from the FINAL region.

We have also looked at the *number* of regressions to regions verb and object from the final region to see whether the likelihood of regressing is different across conditions (Table 7). For both GLME models we ran a base model and compared it to the model with the main effect of sentence condition.

Table 7
GLME for number of fixations coming from the final regions to region verb and region object (DV = fixation count).

Area	Predictor	Estimate	Std. error	z value	Pr(>|z|)
FSA VERB after FINAL	(Intercept)	−2.96	0.22	−13.27	<0.001
FSA OBJECT after FINAL	(Intercept)	−2.92	0.19	−15.71	<0.001

The results show that in the case of both regions (verb and object), the best models are the base models without any fixed effects. This means that there is no difference for how likely it is that there will be a regression from the final region to the verb or the object region, though in the case of the object region, the effect is approaching significance at 0.06 (see details on OSF), with more regressions going from the final region to the object region in the reactivated condition.

To summarize the analyses for the reading task, the significant effect of experimental condition was found in the *Object* region, with longer dwell times in the reactivated condition than in the neutral condition, as well as longer dwell times in the same region when the participants regress from the final region. This signifies potentially different processing strategies in the two conditions.

4.3. *Forced choice screen: eye tracking data and answer accuracy*

The second part of the analysis focused on the screen that followed the sentence where the participants had to make a choice based on semantic relatedness. For the forced choice task, we examined the proportion of looking at the predicted target out of the sum of the target and distractor dwell times. The data were analyzed using generalized linear mixed-effects model with a binomial distribution. We included subjects and items as random effect. We created a base model, which included an intercept and the two random factors (subject, item). To the base model, we first added a fixed effect of sentence condition ($X^2(1) = 521.17, p < 0.001$), then choice options ($X^2(2) = 100.75, p < 0.001$) and then the interaction. This final model was the best fit model we report in Table 8. These models were compared to see whether the addition of effects improved the model fit. The dependent variable was the proportion of looking directed at the target word (i.e., the expected choice) transformed into log odds ratio.

Table 8
GLME for the proportion of looking at the target word (DV = proportion of looking the expected word with respect to the total sum of looking at target and the distracter).[7]

	Estimate	Prob.	Std. Error	z value	Pr(>|z|)
Best fit model with SentCond * ChoiceOp compared to the previous model: $X^2(2) = 497.57$, p $\leq$ 0.001					
(Intercept)	−0.02	0.49	0.08	−0.29	0.77
sentencecondreactivated	0.02	0.50	0.007	3.31	0.0009
choiceoptmetlit	0.14	0.53	0.07	2.05	0.04
choiceoptmetunr	0.17	0.54	0.07	2.50	0.01
sentencecondreactivated:choiceoptmetlit	−0.20	0.45	0.01	−20.88	<0.001
sentencecondreactivated:choiceoptmetunr	−0.16	0.46	0.01	−16.48	<0.001

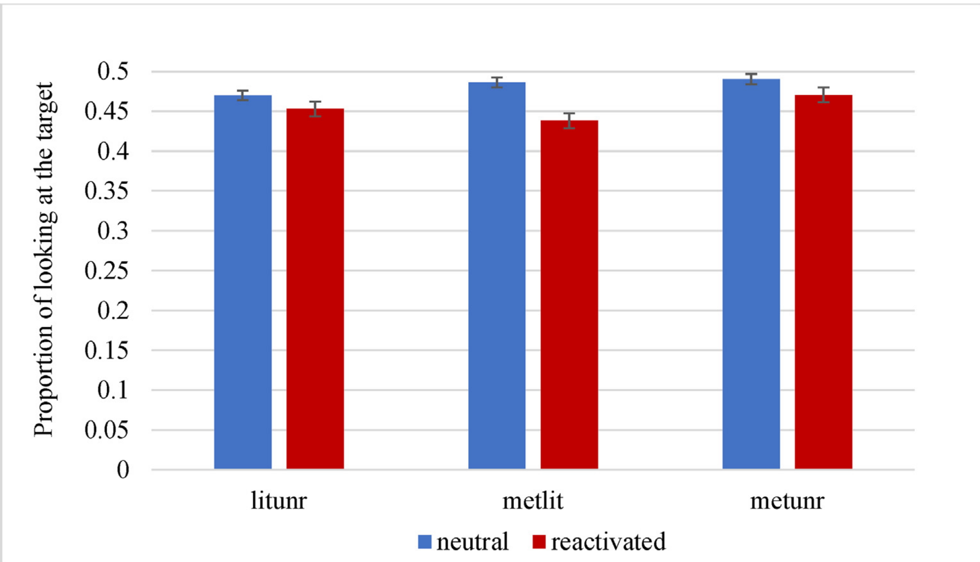

Fig. 3. Proportion of looking at the target word with respect to the total dwell time spent on the screen.

[7] Base levels are sentence condition neutral, choice option literal unrelated. The shortened terms in the table refer to sentencecondreactivted – sentence condition reactivated, choiceoptmetlit – choice option metaphorical literal choiceoptmetunr – choice option metaphorical unrelated, sentencecondreactivated:choiceoptmetlit – sentence condition reactivated:choice option metaphorical literal, sentencecondreactivated:choiceoptmetunr - sentence condition reactivated: choice option metaphorical unrelated.

As can be seen in Table 8 and Fig. 3, the results show that the participants spend more time on the distractor than the target. This might be due to the target being expected, and therefore requiring less effort to process. On the other hand, the distractor (either LIT or UNR) is unrelated so not effortlessly expected. In the reactivated condition, people tend to spend more time on the distractor than the target presumably due to the difficulty of the task (but also due to the direct competition between MET and LIT on that particular screen).

As for the manual responses to the relatedness judgement, on the MET/LIT screen and MET/UNR the target choice was MET, while on the LIT/UNR screen, the target choice was LIT. For instance, for the sentence *The father cut the budget like grass*, on the MET/LIT screen the participant would see the words *deficit*/*meadow*, on the MET/UNR screen *deficit*/*beach*, with *deficit* being the target choice in both cases. On the LIT/UNR screed (*meadow*/*beach*), we considered LIT (*meadow*) the target choice.

We investigated whether sentence condition and choice options affected the participants' likelihood of giving the expected answer or not. We used a generalized linear effects model with a binomial distribution (Table 9) to model the DV Expected Answer (1 vs. 0). We created a base model, which included an intercept and the two random factors (subject, item). To the base model, we first added a fixed effect of sentence condition ($X^2(1) = 3.01$, $p = 0.08$), then choice options ($X^2(2) = 137.92$, $p \leq 0.001$) and then the interaction, which resulted in the best fit model shown in Table 9. These models were compared to see whether additional effects improved the model fit. Below we report the best fit model.

Table 9
GLME for answers to the question (DV = expected answer, base levels: sentence condition neutral, choice option litunr).

	Estimate	Std. error	z value	Pr(>|z|)
Best fit model with SentCond * ChoiceOpt: $X^2(2) = 259.87$, p < .001				
(Intercept)	0.75	0.16	4.55	<0.001
sentenceconditionreactivated	2.36	0.25	9.37	<0.001
choiceoptnmetlit	1.92	0.23	8.27	<0.001
choicecoptmetunr	2.67	0.29	9.27	<0.001
sentencecondreactivated:choiceoptmetlit	−4.74	0.34	−13.87	<0.001
sentenceconditionreactivated: choiceoptmetunr	−2.62	0.42	−6.36	<0.001

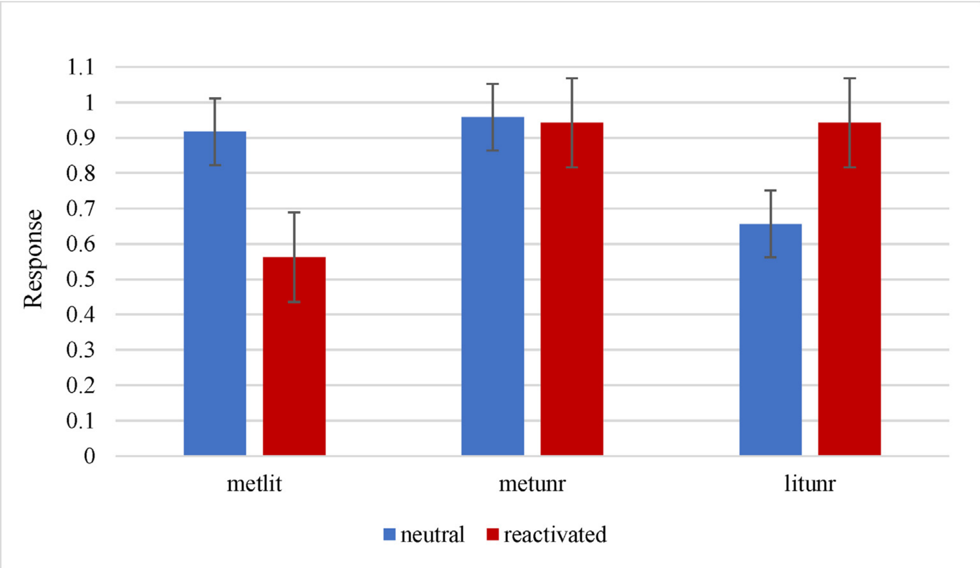

Fig. 4. Responses to the relatedness judgement. Represented as the average value of the answers where the expected answer was coded as 1, unexpected as 0.

As seen in Table 9 and Fig. 4, the results show that participants tend to give more expected answers in the case of the neutral condition when the choices on the screen are MET vs. LIT. This is due to the fact that in the neutral condition there is no competition between the meaning, and it is easier to choose between the two words. There seems to be no difference between the conditions in the case of the MET vs UNR screen which is due to the fact that in this case no competition is expected since the metaphorical meaning is activated in both the reactivated and the neutral condition. Finally, in the case of the LIT vs. UNR screen, the participants give more expected responses in the reactivated condition, possibly because there are no linguistic elements directly connected to the literal and the neutral choices in the neutral condition.

5. Discussion

This study explored whether the conventional metaphoricity of polysemous verbs can be reactivated by the presence of words that are close to the source domain (i.e., elements related to the literal meaning of the verb), resulting in longer processing times of expressions that otherwise would be processed faster.

In the portion of the experiment where the sentence is presented, the results demonstrated that there are no variations in the amount of time spent processing information between reactivated and neutral conditions when encountering a polysemous verb with a metaphorical meaning (referred to as the *Verb* region). This indicates that the presence or absence of contextual elements from the source domain does not impact the speed at which the verb itself is read. This finding holds true for both instances when the verb is encountered for the first time and when it is encountered again during the second phase after the source material has already been encountered. *Cut the budget like grass* and *cut the budget like always* does not push participants to dwell longer on the verb *cut*.

However, the clearest difference between the neutral and the reactivated condition occurs instead in the *Object* region (*the budget*). This is relevant since the literal or metaphorical meaning of the verb is disambiguated by the object. It is upon processing the object region that participants understand the intended meaning of the verb they just saw. Participants dwell longer on and return to the *Object* region, e.g., *budget* after *cut*, suggesting that elements of the source domain (*grass*) are able to reactivate parts of the metaphorical meaning of the expression *cut the budget*.

The inclusion of elements from the source domain, specifically *grass*, appears to reactivate certain aspects of the metaphorical meaning in the expression *cut the budget*. This reactivation prompts readers to revisit not the metaphorical verb itself, but rather the object that selects the metaphorical interpretation of the verb. Participants likely focus on the object because it provides clarification for the intended meaning of the verb. The frequency of revisiting the object is noteworthy: it reveals that there is no difference in the number of times participants refer back to the object in the neutral and reactivated conditions, suggesting that these conditions are structurally comparable. However, it does show that once participants return to the object, they tend to spend more time on it in the reactivated condition compared to the neutral condition. During this attention to the object, participants may establish an analogy between the two transitive actions (*cutting a budget* vs. *cutting the grass*), thereby activating both the literal and metaphorical interpretations of *cutting*.

Finally, it is important to discuss the findings related to the final region (*grass*/*last month*). Our experimental design, which involved varying numbers of words per sentence, does not allow us to determine whether the sentence condition also influences the processing time for this specific region. This conclusion is based on analyzing all three measures (total dwell time, first run, second run), in all of which the model considering word length as the sole fixed effect provides the best fit and the inclusion of the sentence condition does not enhance the model's fit. If there is an effect of the condition, it is likely not strong enough to be detected in the presence of the prominent influence of word length.

To supplement the dwell time results obtained from the reading portion of the task, we examined the dwell patterns in the choice task and analyzed participants' responses. Overall, the findings indicate that participants directed more visual attention towards the distractor rather than the target (see Fig. 3), possibly suggesting that the target choice was more expected, and consequently, participants spent less time considering it. More specifically, in the reactivated sentence condition, we hypothesized that choosing between the two options would be more challenging and require greater effort due to both domains being active which creates competition between the two choices. This hypothesis was confirmed by the significant main effect of sentence condition, with more attention directed towards the distractor compared to the neutral condition. This implies that in the reactivated condition, the task of making a choice becomes more difficult and demanding.

The disparity in attention between the reactivated and neutral conditions appears to be particularly evident when the choice is between MET & LIT (*debt* & *lawn* for *cut the budget*). In the reactivated condition, we find the lowest proportion of attention directed towards the target due to direct competition with the distractor. Here participants may spend more time on the distractor in an attempt to identify any missed information or to resolve the competition. In the neutral condition, where processing is expected to be less demanding, individuals may find it easier and less mentally taxing to choose the correct target. Consequently, the amount of attention directed towards the target does not significantly differ from the other two screens.

Our alternative hypothesis was that highly conventionalized metaphoric expressions would be read effortlessly until the participants get to the contextual elements at the end of the sentence, which are related to the literal meaning of the metaphoric expression. Upon encountering the elements associated with the source domain, participants would be compelled to shift their attention to the source domain, prompting them to go back to the metaphoric expression and reprocess the metaphor as a deliberate metaphor, by cross-domain mapping between source and target domains.

As expected, we found that the experimental condition in which the end of the sentence is built to reactivate the literal meaning of a polysemous (metaphorical) verb, as in *The father cut the budget like grass* takes longer to process than the condition in which the end of the sentence is neutral, and therefore does not include elements from the source domain, as in *The father cut the budget like last month*. Interestingly, we observed longer dwell times in the first condition (reactivated condition) than in the latter condition (neutral), but such latencies were not observed on the verb region, as we may initially have expected. Latencies were observed in the object region. This is complemented by the results of our FSA analysis, which also shows object region is the region on which participants dwell longer when they regress from the final region.

The obtained results indicate that conventionalized metaphoric expressions presented as deliberate metaphors, where the context directs readers' attention towards the source domain or literal meaning of the expression, are processed in a distinct and potentially more cognitively demanding manner compared to conventionalized metaphoric expressions. We take this as an indicator that the processing strategy can be labelled as comparison, rather than categorization, precisely because the processing by comparison involves the activation of both domains involved in the metaphor, while the processing by categorization involves only one domain: the target domain. Metaphor deliberateness, a variable that can be operationalized by linguistic elements such as pragmatic signals used in the text, is a variable that affects speakers' communicative intentions and listeners' processing strategies. By introducing elements of the source domain within sentences that included highly conventionalized metaphoric expressions, we showed that it is possible to induce the activation of the source domain in the hearer's mind, and therefore push the reader to stop in her tracks and go back to reinterpret the metaphorical expression as a metaphor.

In the second part of our study, in the reactivated condition the participants required a lot of effort to select a word that was related to the meaning of the previous sentence, because both domains were active in the speaker's mind. Conversely, after reading a sentence in the neutral condition it was easier to select the word associated to the target domain, which was arguably the only domain active in the speakers' mind. This is also supported by the proportion of looking at the target choice, since in the neutral condition, there is significantly more looks directed at the target, suggesting that this choice is easier. Finally, when comparing the reactivated sentence condition (*The father cut the budget like grass*) and the neutral condition (*The father cut the budget like last month*), if the choice was between a word related to the literal meaning (source domain) of the polysemous verb and an unrelated word, we expected to see more participants selecting the word related to the literal meaning in the reactivated condition (vs. the neutral condition) because the source domain in this case would be reactivated in the speaker's mind. Where the choice is between literal-related and unrelated, there are more expected responses in the reactivated condition vs. the neutral condition, suggesting that the source elements reactivate the literal meaning of the verb. This did not happen in the neutral condition, in which participants have significantly less expected responses. This suggests that the source domain needs to be overtly activated by the elements of the source domain, and that without the reactivation the processing is easy.

Overall, our results suggest that highly conventionalized metaphorical expressions, if used as deliberate metaphors and therefore in contexts where the reader is pushed to pay attention to the source domain, are processed in such a way that they require a higher cognitive effort (longer dwell times) compared to the cognitive processing required by highly conventionalized metaphorical expressions that are non-revitalized.

Throughout the paper, we assumed that the more cognitively demanding way to process metaphor is by means of comparison and cross-domain mapping, while the less effortful way is by means of categorization, in line with previous research (e.g., Bowdle and Gentner, 2005). Based on this assumption, and in line with the DMT (Steen, 2008) we found that metaphor conventionality alone cannot always explain the difference in processing. In fact, even highly conventionalized metaphoric expressions like those included as stimuli in our investigation, which according to the career of metaphor theory by Bowdle and Gentner would likely be processed by means of categorization, can actually be put in a context that stimulates the reader to pay attention to the source domain and therefore process the metaphoric expression by means of comparison, by revitalizing the source domain. Future research will need to delve deeper into how the source domain could be reactivated, to push the reader to process the metaphor as a metaphor, and therefore by means of cross-domain mappings. In fact, the results of our second part of the study, on word selection, suggest that the source domain has to be overtly elicited.

6. Conclusion

The objective of this study was to examine whether conventional metaphors, typically processed through categorization, might instead elicit more demanding processing when elements of the source domain are used to stimulate the reader to pay attention to the source domain. Through an eye tracking study, we tested the processing of sentences containing polysemous verbs with conventional metaphorical meanings in two cases: when there is no activation of the source domain and when there is an element of the source domain introduced.

The findings provide compelling evidence for the reactivation of the source domain, as participants significantly increased their fixation duration on the object associated with the polysemous verbs when exposed to linguistic material from the source domain. This was also supported by the forced-choice task in which the participants found it easier to respond to the task where there was no reactivation of the source domain evidence. This study provides empirical evidence to test DMT for which experimental analyses are still very limited. We hope that in the near future more studies will address this challenge and embrace the discussion of how metaphors are processed, by paying attention to the context in which they are used, rather than by analyzing them in isolation.

Authors' contributions

The study was conceived and designed by AWH and MB. The stimuli were designed by AWH and MB, the data was collected by AWH and analyzed by AWH and NA. The results were discussed and interpreted by AWH, NA and MB. Finally, the paper was drafted by AWH and MB and then edited by NA.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The link to OSF has been included in the manuscript. All data was shared.

Acknowledgements

We would like to thank the members of the Language and Brain Laboratory in Oxford for their assistance and advice, and the School of Psychology at the University of East Anglia. We are especially grateful to Prof Katrin Kohl and Prof Aditi Lahiri for their comments and guidance during the development of this study. This research was funded by a grant from the Arts and Humanities Research Council (Open World Research Initiative: Creative Multilingualism: AH/N004701/1).

References

Barr, D.J., 2013. Random effects structure for testing interactions in linear mixed-effects models. Front. Psychol. 4, 328.
Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Software 67 (1), 1–48.
Blasko, D.G., Connine, C.M., 1993. Effects of familiarity and aptness on metaphor processing. J. Exp. Psychol. Learn. Mem. Cognit. 19 (2), 295.
Bolognesi, M., Werkmann Horvat, A., 2023. The Metaphor Compass. Directions for Metaphor Research in Language, Cognition, Communication, and Creativity. Routledge, London.
Bowdle, B.F., Gentner, D., 2005. The career of metaphor. Psychol. Rev. 112 (1), 193.
Chiappe, D.L., Kennedy, J.M., 1999. Aptness predicts preference for metaphors or similes, as well as recall bias. Psychonomic Bull. Rev. 6 (4), 668–676.
Clausner, T.C., Croft, W., 1997. Productivity and schematicity in metaphors. Cognit. Sci. 21 (3), 247–282.
Cuccio, V., 2018. Attention to Metaphor: from Neurons to Representations. John Benjamins Publishing Company, Amsterdam (G. Steen, Contributor).
De Vries, C., Reijnierse, W.G., Willems, R.M., 2018. Eye movements reveal readers' sensitivity to deliberate metaphors during narrative reading. Sci. Study Literature 8 (1), 135–164. https://doi.org/10.1075/ssol.18008.vri.
Gentner, D., 1983. Structure-mapping: a theoretical framework for analogy. Cognit. Sci. 7 (2), 155–170.
Gentner, D., Bowdle, B., 2008. Metaphor as structure-mapping. In: Gibbs Jr., R.W. (Ed.), The Cambridge Handbook of Metaphor and Thought. Cambridge University Press, Cambridge, pp. 109–128.
Gentner, D., Markman, A.B., 1997. Structure mapping in analogy and similarity. Am. Psychol. 52 (1), 45.
Gentner, D., Bowdle, B., Wolff, P., Boronat, C., 2001. Metaphor is like analogy. In: Gentner, D., Holyoak, K.J., Kokinov, B.N. (Eds.), The Analogical Mind: Perspectives from Cognitive Science. MIT Press, Cambridge, MA, pp. 199–253.
Gibbs Jr., R.W., 2015a. Do pragmatic signals affect conventional metaphor understanding? A failed test of deliberate metaphor theory. J. Pragmat. 90, 77–87.
Gibbs, R.W., 2015b. Does deliberate metaphor theory have a future. J. Pragmat. 90, 73–76.
Glucksberg, S., 2008. How metaphors create categories—quickly. In: Gibbs Jr., R.W. (Ed.), The Cambridge Handbook of Metaphor and Thought. Cambridge University Press, Cambridge, pp. 67–83.
Glucksberg, S., Haught, C., 2006. Can Florida become like the next Florida? When metaphoric comparisons fail. Psychol. Sci. 17 (11), 935–938.
Glucksberg, S., Keysar, B., 1990. Understanding metaphorical comparisons: beyond similarity. Psychol. Rev. 97 (1), 3.
Glucksberg, S., McGlone, M.S., 2001. Understanding Figurative Language: from Metaphor to Idioms (No. 36). Oxford University Press, Oxford, UK.
Grady, J.E., 1997. Foundations of Meaning: Primary Metaphors and Primary Scenes. University of California, Berkeley.
Jones, L.L., Estes, Z., 2006. Roosters, robins, and alarm clocks: aptness and conventionality in metaphor comprehension. J. Mem. Lang. 55 (1), 18–32.
Krennmayr, T., 2011. Metaphor in Newspapers. Netherlands Graduate School of Linguistics.
Lakoff, G., Johnson, M., 1980. The metaphorical structure of the human conceptual system. Cognit. Sci. 4 (2), 195–208.
Lakoff, G., Johnson, M., Sowa, J.F., 1999. Review of Philosophy in the Flesh: the embodied mind and its challenge to Western thought. Comput. Ling. 25 (4), 631–634.
Reijnierse, W.G., 2017. The Value of Deliberate Metaphor. LOT, Utrecht.
Steen, G., 2008a. The paradox of metaphor: why we need a three-dimensional model of metaphor. Metaphor Symb. 23 (4), 213–241.
Steen, G., 2008b. When is metaphor deliberate? In: Johannesson, N.-L., Minugh, D.C., Alm-Arvius, C. (Eds.), Selected Papers from the 2009 Stockholm Metaphor Festival. Department of English, Stockholm University, pp. 43–63.
Steen, G., 2011a. From three dimensions to five steps: the value of deliberate metaphor. Metaphorik.De 21, 83–110.
Steen, G., 2011b. What does 'really deliberate' really mean? More thoughts on metaphor and consciousness. Metaphor Social World 1 (1), 53–56.
Steen, G.J., 2013. Deliberate metaphor affords conscious metaphorical cognition. J. Cognit. Semiot. 5 (1–2), 179–197.
Steen, G., 2015. Developing, testing and interpreting deliberate metaphor theory. J. Pragmat. 90, 67–72.
Steen, G.J., 2016. Mixed metaphor is a question of deliberateness. In: Gibbs Jr., R.W. (Ed.), Mixing Metaphor. John Benjamins Publishing Company, Amsterdam, pp. 113–132.
Steen, G., 2017. Deliberate Metaphor Theory: basic assumptions, main tenets, urgent issues. Intercult. Pragmat. 14 (1), 1–24.
Steen, G.J., Dorst, A.G., Herrmann, J.B., Kaal, A.A., Krennmayr, T., 2010. Metaphor in usage. Cognit. Ling. 21 (4), 765–796.
Werkmann Horvat, A., Bolognesi, M., Lahiri, A., 2021. Processing of literal and metaphorical meanings in polysemous verbs: An experiment and its methodological importance. J. Pragmat. 171, 131–146.
Wicke, P., Bolognesi, M., 2020. Framing COVID-19 how we conceptualize and discuss the pandemic on Twitter. PLoS ONE 15 (9), e0240010.
Wicke, P., Bolognesi, M., 2021. Covid-19 discourse on Twitter: How the topics, sentiments, subjectivity and figurative frames changed over time. Front. Commun. 6, 1–45.

Ana Werkmann Horvat is an assistant professor at the University of Osijek (Croatia). Her research interests lie mainly within the subfields of semantics and psycholinguistics. Recently, she has been focusing on processing of figurative expressions (conventional and creative).

Marianna Bolognesi is an associate professor at the University of Bologna (Italy). Her research focuses on lexical semantics and multimodal communication, and in particular on the relation between language and thought, and on the semantic representation of word meaning in mind. Currently, she is the PI of the ERC project ABSTRACTION.

Nadja Althaus is a lecturer at the School of Psychology, University of East Anglia (UK). Her research investigates language processing and development, using eye tracking and computational models. She is interested in phonology and speech processing, as well as the interface between language and object knowledge in word learning.