Journal of Articles in Support of the Null Hypothesis
Vol. 17, No. 1

www.jasnh.com

Swallowing the pill and being laid to rest: No advantage for metonymic over metaphoric idioms in primed lexical decisions?

Diana Michl
Sprachwissenschaft
FU Berlin
Kastanienallee 4
14471 Potsdam
diana.michl@fu-berlin.de

Idioms are often metonymic or metaphoric, thus different in non-literalness. Is this reflected in their automatic processing? As metonymic idioms are perceived as more literal and partly read faster, it is possible that reactions to metonymic idioms are generally faster and that this advantage is greatest for literally related words. We conducted two primed lexical decisions with metonymic and metaphoric idioms as primes and literally and non-literally related adjectives. The expected effects were not found. A Bayesian regression analysis suggests there are no differences in metonymic vs. metaphoric idioms. Regardless of idiom type, we found largest processing advantages for non-literally related words and smaller advantages for literally related words. We conclude that processing of non-literal structure is not part of automatic processing.

Keywords: idioms; non-literalness; metaphor; metonymy; semantic processing; priming

Acknowledgements
This study was funded by the author's personal finances which are provided through a scholarship by the Friedrich-Naumann-Stiftung für die Freiheit. I thank two anonymous reviewers for their insightful and helpful comments. First and foremost, I am very grateful to Rebecca Reglin (Universität Potsdam) for her brave and relentless work in helping to create and control the target words, and her excellent recordings. I am also very thankful to Andrey Anikin (University of Lund) for his great advice on Bayesian modelling, to Kate Stone (Universität Potsdam) for her enduring help on conducting power analyses, and to Holger Hill (KIT Karlsruhe) for his helpful advice on SOA vs. ISI.

Introduction

Idioms such as *to have a big mouth* or *to wrap someone around one's finger* are non-literal, i.e. they have a non-literal meaning that is usually intended. Among the most widespread and well-known kinds of non-literal language are metaphors and metonymies, and idioms – "syntactically complex, fixed expressions" (Sailer, 2013) – are indeed often metaphoric or metonymic. In metaphors and metonymies, words are used to express a more or less different meaning than the literal. It is widely acknowledged that metaphors and metonymies differ in their non-literal structure, with metaphors being based on an analogy relation and metonymies being based on a contiguity relation. It is generally agreed upon that metaphors operate between at least two distinct semantic concepts or domains whereas metonymies operate within one semantic concept or domain (Fauconnier & Turner, 1996; Lakoff & Johnson, 1980; Mendoza Ibánez, 2003; Spieß & Köpcke, 2015).

In the metaphoric idiom 'she *breaks the ice*', the 'broken ice' stands for a tense atmosphere that has just been overcome. While a tense atmosphere among people can be metaphorically referred to as 'cold', there is no actual semantic connection between ice as frozen water, and an atmosphere between humans. Conversely, in the metonymic idiom 'he *has to have the last word*', the 'word' stands for the desire to speak and thereby finish any conversation. There is a close semantic connection between 'word' and '(the desire to) speak' as speaking usually happens in the form of words. Thus, a metonymic relationship differs from a metaphoric relationship mainly in the close semantic connection between what is said and what is meant.

Both metaphors and metonymies among themselves differ in their degree of non-literalness: there are rather non-literal and rather literal metaphors as well as rather non-literal and rather literal metonymies. However, the difference in metaphoric compared to metonymic structures seems to be mirrored in how strongly metaphoric and metonymic idioms as separate groups are perceived as non-literal: our earlier rating studies show that native speakers rate metaphoric idioms as distinctly more non-literal than metonymic idioms (Michl, 2019a). Indeed, in a metaphor such as *to swim against the current*, the concept expressed (*current* in a body of water) is literally unrelated to the intended concept (a 'general trend' or 'fashion'). In a metonymy such as *to have an eye for detail*, the concept expressed (*eye*) is literally related to the target concept (i.e. 'talent to acknowledge detail visually') (Michl, 2019a).

But is this conscious differentiation in the ratings also reflected in automatic semantic processing and the lexical access of idioms? Conflicting findings from cognitive linguistics and semantic processing studies exist: it has been well-established that idiomatic expressions as a group have a processing advantage over nonidiomatic language and that their comprehension is automatized to some degree (Canal, Pesciarelli, Vespignani, Molinaro, & Cacciari, 2017; Conklin & Schmitt, 2008, 2012; Cronk & Schweigert, 1992; Sprenger, Levelt, & Kempen, 2006; Swinney & Cutler, 1979; Vespignani, Canal, Molinaro, Fonda, & Cacciari, 2010). Given that idioms are often defined as lexical units or even large words (Sprenger et al., 2006), it is well possible that their meanings are automatically retrieved upon recognition. On the other hand, it has been suggested that the different non-literal structures of metaphors as opposed to metonymies should affect the processing of idioms (Omazić, 2008). Furthermore, empirical research from psychology and psycholinguistics has shown that metaphors are more difficult to process than metonymies (Annaz et al., 2009; Rundblad & Annaz, 2010; see also Klepousniotou, 2002; Weiland, Bambini, & Schumacher, 2014), while empirical research comparing metaphors and metonymies directly is scarce. Indeed, there is some evidence suggesting that metonymies are also processed as more literal (Bambini, Ghio, Moro, & Schumacher, 2013; Michl, 2019b), and in self-paced readings of sentences with embedded idioms, metonymic idioms were also processed faster than metaphoric idioms (Michl, 2019b). While it has been researched whether and how long literal meanings of idioms are accessed in sensicality judgments (Bambini et al., 2013), and primed lexical decisions (Colombo, 1993), it has not been tested whether lexical access of idiom-related words is easier or more difficult depending on how non-literal an idiom is.

It is assumed that lexical access consists of at least two stages. The first is semantic access, thus the choice of the lemma (Caramazza, 1997). Models of lexical access mostly assume that upon hearing or reading a word, several kinds of information are activated via spreading activation (Collins & Loftus, 1975; Gernsbacher & Robertson, 1999). Choosing the correct or most relevant piece of information is also determined by suppression of irrelevant information (Gernsbacher & Robertson, 1999; Giora, 1997). For idioms, information related to the idiomatic meaning is activated. Some evidence from lexical decision suggests a processing advantage of literally related compared to unrelated words to idioms, independent of sentential context (Canal, Pesciarelli, Molinaro, Vespignani, & Cacciari, 2015). Thus, especially in idioms with a more literal meaning or a large overlap between idiomatic and literal meanings, literal information should be activated as well, and it should be activated more strongly than in the non-literal idioms because it is more relevant. This can be tested in primed lexical decision tasks.

Semantic priming effects are due to facilitation without inhibition (Posner & Snyder, 1975; Spencer & Wiley, 2008) and are in favor of the information more relevant or closer to the prime. They occur because closely related, relevant information is automatically and strongly activated. Semantic priming effects are robust between idioms and words semantically or conceptually related to the idiomatic meaning (Cacciari & Tabossi, 1988; Caillies & Butcher, 2007; Sprenger et al., 2006; Titone, Holzman, & Levy, 2002). The priming effects from idioms lead to facilitated semantic processing of ensuing target words that are semantically closely related (Titone et al., 2002). It has been found that participants respond faster to figuratively (i.e. idiomatically) related targets than to literally related targets if all idioms are equally salient, meaning equally recognizable, well-known and comprehensible (Laurent, Denhières, Passerieux, Iakimova, & Hardy-Baylé, 2006). On the other hand, priming effects have been found to be almost equal on literally and non-literally related words (Colombo, 1993). It has also been suggested that literal meanings of idioms are indeed accessed during immediate processing (van de Voort & Vonk, 1995) which indicates some processing advantage of literally related compared to unrelated words.

However, how strongly literally related words are activated compared to non-literally (i.e. idiomatically) related words also depends on contextual effects. If an idiomatic interpretation is biased, then non-literally related words are activated more strongly whereas in a neutral context, this effect is weaker (Cacciari & Tabossi, 1988; Colombo, 1993). Similarly, our self-paced reading experiments suggest that the difference between the reading speeds of metonymic compared to metaphoric idioms is larger in neutral sentences than in sentences that bias an idiomatic interpretation. While the advantage of metonymic compared to metaphoric idioms was still present in the biasing condition, it was clearly smaller (Michl, 2019b).

Lexical decision experiments with idioms as primes usually compare reaction times of idiomatically vs. unrelated target words. If literal meanings are of interest, literal target words are compared to non-literal or unrelated words. The stronger the activation of a word, the stronger the priming effect and the faster the reaction to the word should be. Our research question is: How strongly are non-literal and literal words activated, depending on how non-literal the idiom prime is? Concretely, are literal words processed faster when preceded by a metonymic idiom compared to a metaphoric idiom? Moreover, is the processing difference between unrelated and literal words larger for metonymic idioms because they are more literal?

If the different degree of non-literalness in metonymic compared to metaphoric idioms is indeed mirrored in a primed lexical decision task, then we expect the following differences in reaction times: Literally related words should be processed faster when preceded by a metonymic idiom than when preceded by a metaphoric idiom, because metonymic idioms are more literal. At least for the metonymic idioms, but possibly for metaphoric idioms as well, literal words should be faster than unrelated words because their relation is still closer to the idiomatic meaning. This would mean that metonymic idioms are indeed processed as more literal.

If, one the other hand, differing non-literalness is not reflected in the priming effects, then merely non-literally related words should evoke the fastest reaction times, which is a common finding in idiom priming experiments when an idiomatic interpretation is biased. Reaction to literal words generally should be slower due to weaker priming effects because literally related words as these are less relevant to the idiomatic meanings. These findings would be independent of idiom type in the prime sentence, and they would suggest that idiomaticity – being an idiom and being recognized as a familiar idiom – is the only driving force influencing semantic processing at this early processing stage.

A number of other properties also affect processing in idioms and in single words. In idioms, *familiarity* determines processing ease in fixed expressions in general (Cronk & Schweigert, 1992; Gibbs, 1980; Libben & Titone, 2008; Nippold & Taylor, 2002; Schweigert, 1986; Tabossi, Fanari, & Wolf, 2009; Titone & Connine, 1994). It refers to how well-known an idiom is. A related influential property is *comprehensibility* (equivalent to 'meaningfulness', Titone & Connine, 1994). It refers to the ease with which a recipient can understand an idiom or make sense of it (Katz, Paivio, Marschark, & Clark, 1988). Other factors influencing processing ease in lexical decision depend on the mode of presentation. For example, processing can be impeded by internal syntactic modifications of the idiom. This effect was found with idioms whose content was modified, that were presented by rapid serial visual presentation mode, and where the lexical decision was to be made on a word within a sentence (van de Voort & Vonk, 1995). High compositionality of idioms – the degree to which as idiomatic meaning can be derived from its individual words – can facilitate processing ease in lexical decision. This was found by Caillies and Butcher (2007) who also used rapid serial visual presentation and neutral prime sentences that did not bias an idiomatic interpretation of the idiom in the sentence. On the other hand, the opposite effect was found by Titone and Libben (2014) in a cross-modal priming lexical decision, and only at 1000ms after offset of the non-biasing prime sentence. The authors themselves argue that compositionality is not likely to affect idiom priming at the automatic stage (Libben & Titone, 2008; Titone & Libben, 2014). This shows that effects of syntactic flexibility and compositionality that emerge in one experiment do not necessarily come into effect within even similar experimental designs. Yet their possible emergence should be considered during the experimental design stage.

In words, *frequency* (which is generally seen as tightly linked to familiarity) is known to affect processing ease, thus the more frequent, the more easily a word is processed (Andrews, 1989; Carreiras, Mechelli, & Price, 2006; Chumbley & Balota, 1984). Increasing *length* of words also leads to slower responses (New, Ferrand, Pallier, & Brysbaert, 2006). Lastly, *orthographic neighbors* and their frequency can affect processing ease (Vergara-Martínez & Swaab, 2012). Neighborhood effects can arise through words derived through letter transposition in another (Acha & Perea, 2008; Andrews, 1989, 1997; Johnson, 2009) which is likely to produce words with a different semantic content (Levenshtein, 1966). This affects semantic processing and priming.

We introduce method, design, and results of two primed lexical decision experiments after one another, with a short discussion for each, then continue to re-analyze both sets of findings with Bayesian regression analyses. After a comparison of all findings, we provide a general discussion and conclusion.

Method

Experiment 1

To explore the effects of non-literalness and the strengths of different semantic relationships between idioms and words, we conducted a lexical decision task with a cross-modal semantic priming. The primes were German sentences containing metonymic and metaphoric idioms. The idioms were embedded in sentences to provide a more natural context as in natural situations, idioms would rarely be encountered on their own, but rather as part of a sentence. They were presented in their canonical word order, unmodified in their content, and the idiomatic interpretation was always biased. To further contribute to a somewhat natural situation and to avoid effects of compositionality, any possible spill-over or sentence wrap-up from reading, the primes were presented auditorily and as complete sentences. The targets strings were adjectives and participles and legal nonwords resembling adjectives and participles.

Semantic priming effects under short interstimulus intervals (ISI) up to 350 to 400ms are attributed to automatic spreading activation (Collins & Loftus, 1975) and unconscious processing, whereas priming effects ISI over 400ms are linked to conscious, controlled, or strategic processing that can also be based on expectancy (Carter, Hough, Stuart, & Rastatter, 2011; Hutchison et al., 2013; Spencer & Wiley, 2008).[1] Thus we used a 200ms interstimulus interval (ISI) which should be too short to allow priming effects caused by expectancy-based strategies, as has also been found by Moss, Tyler, Hodges, and Patterson (1995). Moreover, at this short ISI, both the literal and the idiomatic meanings of the idioms should still be activated (Smolka, Rabanus, & Rösler, 2007), and automatic priming effects should still be reasonably large, given that they tend to decrease between ISIs of 0ms to ISIs of 400ms (Carter et al., 2011). Lastly, no effects of decomposability in idiom primes are expected at this short ISI, as such an effect has only been found with an ISI of over 1000ms in a cross-modal priming lexical decision (Titone & Libben, 2014).

Material

Idiom primes. 74 German idioms (28 metonymic, 46 metaphoric idioms) were embedded in sentences that biased an idiomatic interpretation. All idioms were taken from an idiom database containing 244 metonymic and metaphoric idioms (Michl, 2019a) selected from a modern German idiom dictionary (Schemann, 2011). This database had to be created as extracting metaphors and metonymies from corpora remains a great challenge for many reasons (Stefanowitsch & Gries, 2007), and is likely even more difficult for suitable idioms. For detailed information on creation of this database, please refer to Michl (2019a). The idioms selected for the present experiment were matched according to several properties and have been used before in two self-paced reading tasks (Michl, 2019b). Importantly, especially metonymy has a number of subtypes. In the idioms, however, we almost exclusively found the metonymic forms 'part for whole' and 'whole for part'. Only these two forms were used in our materials for all processing experiments.

Familiarity and comprehensibility. The idioms in the original database were rated by two different groups of participants on familiarity (n = 96) and comprehensibility (n = 86). Participants rated on 5-point Likert scales how often they encountered an idiom (group 1) or how comprehensible it was to them (group 2). On the 5-point Likert scale for the frequency of encountering idioms, participants could choose from 'hardly ever' to 'very frequently', or instead choose the separate answer 'never encountered it before'. The Likert scale for comprehensibility had 5 possible answers ranging from 'extremely difficult to understand' to 'extremely easy to understand'. Only idioms scoring 'highly' to 'extremely familiar/easy to understand' (based on median and mean ratings) were selected for the present experiment.

Non-literalness. 104 participants rated the idioms in the original database on how non-literal they were on a 5-point Likert scale. This Likert scale ranged from 'extremely literal' to 'not at all literal'. Metonymic idioms were rated as 'fairly literal' on average (M = 2.5, SD = 0.17), whereas metaphoric idioms were on average rated as 'hardly literal' (M = 4.0, SD = 0.19) (taken from Michl, 2019a).

Structure. The idiom sentences were short main clauses of the structure 'Prepositional phrase + Idiom', specifically 'Prepositional phrase + [verb + subject noun + (potential adverb) + (preposition) + noun]'. In a few cases, the idiom was longer, so the structure was '[verb + subject noun + (potential adverb) + (preposition) + noun + preposition + noun]'. Please see the Appendix for examples of the material. It was thus ensured that the last word was always the idiomatic noun, after which the target string would follow. For the sake of high comparability, the sentences were ensured to be of nearly equal length (M = 49.5 letters, SD = 1.7). After the sentences were constructed, two individuals reported whether they could detect an idiom in each. Sentences were then adapted until only the idiomatic reading was prominent. Lastly, idioms were easily recognized in all cases but two. These two sentences were excluded from the material, leaving 72 idioms (26 metonymic).

Plausibility. All prime sentences were rated for plausibility on a 5-point scale by three adult German native speakers who did not participate in the final experiment. All sentences were modified until they received a mean rating of at least 2 ('quite plausible').

Target strings. We paired each of the 72 idiom primes with four target strings, namely three different German adjectives and one nonword, thus creating 288 test items. Most importantly, the words were chosen to be either related to the idiomatic meaning of the idiom prime (thus non-literally related), related to the literal meaning of the idiom prime (thus literally related), or unrelated to the idiom prime. Where it was impossible to find an adjective fitting in the intended manner, we chose a past or present participle instead. We controlled the strings for properties that affect processing.

Frequency. Frequency measures of all lemmas were obtained from the DLEXDB corpus for German language. Advantages of

[1] The literature on this time course is abundant and rather unanimous, but it usually refers to the much more frequently discussed stimulus-onset asynchrony (SOA), not on the interstimulus interval (ISI). Except for (Moss et al., 1995), all authors quoted here specifically refer to the ISI, yet themselves quote authors who refer to the SOA. It seems that ISI and SOA are sometimes used interchangeably despite their different definitions. Compare, for example, Carter et al. (2011) and Groot, Thomasson, and Hudson (1986).

Table 1
Means and standard deviations of length and frequency by condition

idiom type -> condition	Length (in letters) metonymic	Length (in letters) metaphoric	Frequency (log) metonymic	Frequency (log) metaphoric
unrelated	9.14 (1.33)	8.84 (1.23)	0.50 (0.45)	0.62 (0.47)
literal	8.09 (1.30)	8.88 (1.59)	0.47 (0.84)	0.02 (0.86)
non-literal	9.59 (1.35)	9.87 (2.01)	0.53 (0.61)	0.46 (0.82)
nonword	9.15 (0.68)	9.24 (0.75)	0 (0.00)	0 (0.00)

DLEXDB are its balanced composition, its modernity, and its large size (see www.dlexdb.de or Michl, 2019a). Mean normalized log-10 frequency of the lemmas was 0.42 (SD = 0.73). 'Normalized' refers to the occurrence of one word per 1000 words of running text. Log frequency measures by condition and idiom type can be found in Table 1.

Length. Since reading speed is affected by the length of units by about 15ms per character (Just, Carpenter, & Woolley, 1982; Schmitt & Underwood, 2004), we attempted to match strings by their number of letters. Mean length of all strings was 9.1 letters (SD = 1.45). We ensured that no word had <6 letters and <2 syllables. Mean number of syllables was 2.7 (SD = 0.6). Length measures by condition and idiom type can be found in Table 1.

Orthographic neighbors and neighborhood frequency. We checked whether the words had any orthographic neighbors according to Coltheart's N. A word is a Coltheart neighbor to another if it can be formed through substitution of one letter (Levenshtein, 1966). Overall, the words had between 0 and 4 Coltheart neighbors (M = 1.55, SD = 0.84, while 147 out of 216 had no orthographic neighbors, and 9 had 3-4). For each target word, the cumulative normalized log frequencies of its orthographic neighbors combined were calculated. The mean cumulative log-10 frequency was -0.23 (SD = 0.62), thus in all cases, the orthographic neighbors together were much lower in frequency than the target words themselves. All nonwords were ensured to have no orthographic neighbors.

Fillers. We constructed 30 filler pairs, each consisting of a non-idiomatic prime sentence and a nonword. This was done to roughly balance the amount of nonwords and words.

Procedure

The 102 sentences were recorded in a neutral tone by a young female speaker. Each audio file was about 2.7 seconds long and ended immediately after the final phoneme. In sum, we had 318 stimuli: 30 fillers and 288 test items. These 288 test pairs were rotated over four lists by Latin square design to ensure that each participant was exposed to every target type equally often, but heard every idiom prime only once. For example, the idiom prime *In der Lehrzeit hat Marie ihre eigenen vier Wände* occurred in all four lists but was followed by a different target word string in each list. *Seine eigenen vier Wände haben* ('to have one's own four walls') means 'to live on one's own' and indicates a level of relative personal independence. In this example, list 1 had the non-literally related target word *eigenständig* ('independent'). This means that the non-literally related word was related to the non-literal meaning of the idiom, 'to live on one's own'. List 2 had the literally related word *räumlich* ('room-like'), as a room is often referred to by *four walls*. This means that the literally related word was related to the literal meaning of the idiom, 'to have one's own four walls/own room'. List 3 had the unrelated word *gefühllos* ('insensitive'), and list 4 had the nonword **wopelhaft*. The words were selected and checked by three independent German native linguists.

Finally, each list consisted of 72 test items (idiom prime with target string) plus the same 30 fillers (prime with nonword). Each list thus contained 54 words and 48 nonwords. In sum, the experiment comprised 102 individually randomized trials for each participant. The CTRL keys were assigned the answer options 'WORD' and 'NOT A WORD'.

Each trial began with the audio playback of the prime sentence. During playback, a fixation cross was displayed in the center of the computer screen for about 2700ms. 200ms after the audio ended (ISI = 200ms), the target string was displayed in the center of the screen, along with a reminder where on the keyboard the answer options were. Then participants had to press one of the CRTL keys to make a lexical decision. This keypress started the next trial. Reaction times were measured from the onset of the display of the string. If no key was pressed after 7000ms, the trial was timed-out and the next trial began. No feedback was given on the correctness of the answer.

Participants were assigned one of the four lists randomly. After filling in the consent form, they were given headphones and seated comfortably about 60cm before the computer screen. They were instructed to listen to the sentences, look at the fixation cross, read the string that would replace the cross, and indicate as quickly and accurately whether it was a German word or not. They were told several times to leave their index fingers on the keys indicating 'WORD' and 'NOT A WORD' at all times during the testing phases. After six practice trials, the actual experiment began. After 60 trials, a break was preset whose length participants themselves could determine. The lexical decision session lasted about ten minutes.

The experiment was implemented in

Table 2
Means and standard deviations of reaction times in ms

idiom type -> condition	Experiment 1 all	Experiment 1 metonymic	Experiment 1 metaphoric	Experiment 2 all	Experiment 2 metonymic	Experiment 2 metaphoric
unrelated	691 (248)	698 (239)	687 (252)	861 (304)	875 (309)	853 (301)
literal	732 (333)	719 (300)	738 (350)	853 (344)	837 (354)	862 (338)
non-literal	668 (256)	661 (249)	672 (260)	822 (320)	822 (319)	822 (321)
nonword	749 (283)	728 (256)	761 (296)	949 (347)	936 (346)	956 (348)

the computer program DMDX 5.1.3.3 (J. Forster & Forster, 2003). The material was presented in marine blue 38-point serif font against a light blue background.

Participants

36 early monolingual native speakers of German recruited from the universities of Potsdam and Tübingen participated in the experiment for financial reimbursement or course credit. Most were students of various subjects of study. 75% were female, 25% were male; 81% were right-handed. All had normal or corrected-to-normal vision and were on average 28.6 years old (*SD* = 9.5 years).

Analysis

We conducted all analyses with R 3.6.1 (R Core team, 2019) and R Studio 1.2.1335 (R Studio team, 2019) and prepared data using the R package 'car' 3.0-3 (Fox, Weisberg, & Price, 2019). Responses in the lexical decision task were excluded if they were false, timed-out or slower than 2500ms[2]. 4.8% of data were thus excluded. Means and SDs in milliseconds were obtained for all conditions, both globally and depending on idiom type. To predict the speed of the RT depending on global as well as individual factors, we conducted linear mixed effects regressions, using the R packages 'lme4' 1.1-21 (Bates, Maechler, Bolker, & Walker, 2019) and 'lmerTest' 3.1-0 (Kuznetsova, Brockhoff, & Haubo Bojesen Christensen, 2019). Mixed effects regression allows accounting for individual variances among participants and among items within a single model (Baayen, Davidson, & Bates, 2008; Field, Miles, & Field, 2014). RT were inversed (von der Malsburg, 2018) to meet the regression model assumptions and to achieve better model fit, making results more robust and reliable. For comparison, however, final analyses were also performed on raw RT (Lo & Andrews, 2015), which yielded very similar results. Condition was a 4-level factor with 'unrelated', 'literally' and 'non-literally related words', and 'nonwords', with unrelated words serving as the reference level. Idiom type was a factor with the 2 levels 'metaphoric' and 'metonymic'. Numeric variables such as length, frequency, and age were scaled. Some factors might be correlated, for example, many frequent words tend to be short, or a high number of orthographic neighbors might have a higher cumulative frequency than a low number of orthographic neighbors. Thus, we ran Pearson's product-moment correlations between length and frequency, and between number of orthographic neighbors and their cumulative frequency. No correlations over $r = -0.31$ were detected. Multicollinearity was not detected in the other variables (all variance inflation factors were between 1.0 and 1.4).

The distinction between metonymic and metaphoric is dichotomous, so to allow for a more fine-grained examination on how perceived non-literalness in idioms might influence their processing, we included mean non-literalness ratings of the idioms as a numeric fixed effect in the final models. This was a post-hoc decision. Given that non-literalness rating and idiom type somewhat correlate, the effect of idiom type was always tested with and without rating.

Besides theoretical considerations, models were chosen to be minimal, based on the relevance and significance of factors as well as the goodness of fit measures Akaike information criterion AIC (the lower, the better the model fit) and R^2 (the higher, the better the model fit). R^2 was calculated using the R package 'MuMin' 1.43.6 (Barton, 2019). Using the R package 'HLMdiag' 0.3.1 (Loy, 2015), final models were also checked for large Cook's distances (Cook's d ≥ 1) and run without high leverage points (leverage > 0.06). This neither changed results nor increased model fit, thus all high leverage points were kept.

Results and Discussion

Raw mean reaction times were fastest for non-literally related words and faster for unrelated words than for literally related words. Means and sd's of raw reaction times can be found in Table 1. Five item-specific factors had no significant effect: number of orthographic neighbors, cumulative neighbor frequency, grammatical category of the target strings (adjective or participle), and mean ratings of non-literalness. Thus, these predictors were excluded from further analysis. The final model then contained random slopes for log frequency by participant and random intercepts for participants and for target strings. Fixed effects were participants' age, length and frequency of the target strings, and the 4×2 interaction of condition (unrelated, literally and non-literally related word, and nonword) and idiom type (metonymic and metaphoric). Age, length, and frequency proved to be significant. For the regression output, see Table 3. While there were main effects for non-literally related words and nonwords, the interaction term had no effect. This means that compared to unrelated words primed by a metaphoric idiom, literally or non-literally related words and nonwords were not processed significantly faster or slower when primed by a metonymic idiom. Figure 1 ("Experiment 1") shows the plotted interaction. To test for simple effects of the single predictors, this model was also tested without the interaction term and with the fixed effects of condition and idiom type as additive predictors instead. The same effects were found: age, length, frequency had significant effects, while neither idiom type nor non-literalness ratings were significant; non-literally related words were significantly faster that unrelated words, nonwords were significantly slower, and literally related words were not significantly different from unrelated words. The conditional R^2 indicated 50.3% of variance to be covered by both the interaction model and the additive model indicating the interaction term did not improve model fit.

There was no evidence that idiom type, hence more literal or

[2] We refrained from more rigorous trimming of the slower responses for the following reasons: It is not a suitable solution if the data are intended to remain realistic, it can create other bias problems (Ulrich & Miller, 1994), and in addition, the effects of interest can be in the tails and thus could be accidentally discarded, which can even reduce power (Ratcliff, 1993). Lastly, more rigorous trimming also cannot mitigate the size of large standard deviations completely.

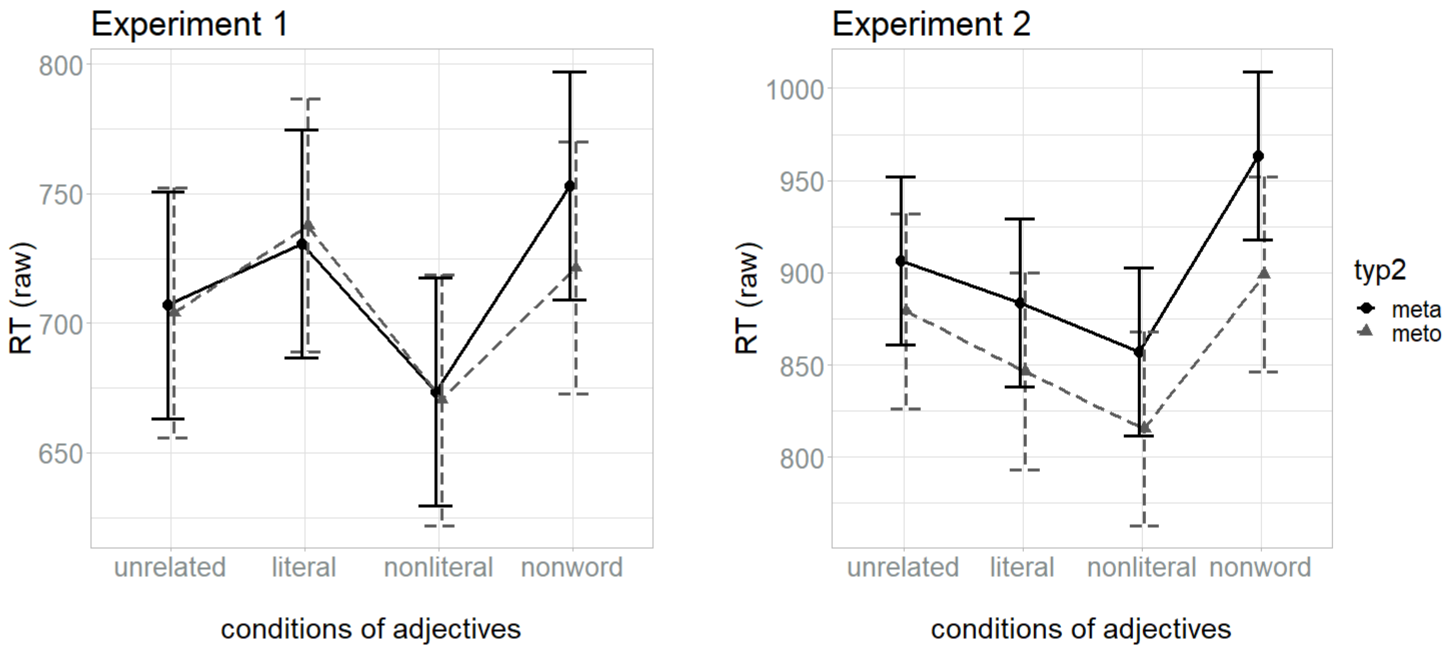

Figure 1. Interaction of semantic relation and idiom type, frequentist models

non-literal structure, affects how fast ensuing words are processed, whether they are idiomatically or literally related. There were also no main effects of literal words or idiom type, while there was a significant main effect of non-literal words. Unsurprisingly, results show the longer and less frequent a word is, the more slowly it is processed. In addition, increasing participant age leads to slower reaction times.

However, the absence of our expected effects cannot be interpreted as their true nonexistence. In fact, our expected effects could be difficult to detect with a sample size (n = 36) too small to detect a small effect. At any rate, findings could be due to chance and experiments should generally be repeated with a different sample of participants, especially when data are subjected to frequentist analyses. To strengthen or disconfirm findings, increase power and get a larger variety of participants, the experiment was repeated with 74 German native speakers from across the whole of Germany.

Table 3
Experiment 1: Regression output with interaction term

variable	b	SE	t	Pr(>|t|)
(Intercept)	*-1.53*	*0.05*	*-33.62*	*<0.001 ****
literal	-0.01	0.03	-0.24	0.815
non-literal	-0.07	0.03	-2.33	0.020 *
nonword	0.10	0.03	3.14	0.002 **
idiom meto	0.02	0.04	0.51	0.613
length	0.03	0.01	3.33	0.001 ***
age	0.14	0.04	3.18	0.003 **
frequency	-0.06	0.01	-6.00	<0.001 ***
literal:idiom meto	0.03	0.05	0.54	0.59
non-literal:idiom meto	-0.03	0.05	-0.61	0.543
nonword:idiom meto	-0.06	0.05	-1.17	0.242

Conditional R^2: 0.503. RT are inversed.

Experiment 2

Material, design, and analysis were the same as in Experiment 1. There was a difference in the implementation: the experiment was internet-based and accessible over a webpage, implemented on Windows 7 and Ibex 0.3.7 (Drummond, 2013). This has the advantage that participants are tested in circumstances more natural and familiar to them as opposed to a laboratory situation. Participants were recruited on the platform Prolific.[3] Presentation of the material and layout were identical to those of Experiment 1.

Participants

74 early monolingual native speakers of German from across Germany participated in the experiment for financial reimbursement. They came from a wide variety of professional and educational backgrounds. 46% were female, 50% were male, 4% were diverse or provided no answer; 92% were right-handed. All had normal or corrected-to-normal vision and were on average 32.3 years old (SD = 10.6 years).

[3] Accessible at www.prolific.ac or app.prolific.co

Analysis, Results and Discussion

Analysis was identical to that of Experiment 1. 6.7% of data were excluded due to incorrect, timed-out or responses over 2500ms. Comparative analyses performed on raw RT yielded highly similar results to those applied to inversed RT. No Cook's d ≥ 1 were found and excluding high leverage points neither changed results nor increased model fit.

Raw mean reaction times were fastest for non-literally related words and similarly fast for unrelated and literally related words. Means and sd's of raw reaction times are in Table 2. Results differed slightly to Experiment 1 with regards to the relevant fixed effects. Age had no effect and was thus excluded. The final model contained random slopes for log frequency by participant and random intercepts for participants and for target strings as random effects, and length, frequency, and the interaction term of idiom type and condition as fixed effects. Table 4 shows the regression output. Length and frequency were significant. The interaction terms were not significant. Main effects were found for non-literally related words, nonwords, and for literally related words ($b = -0.05$, $p = 0.035$), indicating that literally related words were processed slightly faster than unrelated words. In addition, non-literalness ratings had a significant effect, indicating that the more literal an idiom was rated, the faster the corresponding items were processed ($b = -0.03$, $p = 0.028$). The data were also analyzed without the interaction and instead condition and idiom type as additive predictors. Only one difference to the interaction model was found: idiom type was significant, indicating that items containing metonymic idioms were processed faster than items containing metaphoric items ($b = -0.06$, $p = 0.041$). Figure 1 ("Experiment 2") shows the plotted interaction. Conditional R^2 indicated 48.3% of variance to be covered by both the interaction model and the additive model.

In Experiment 2, no evidence is found that literal words are processed faster when preceded by a metonymic idiom compared to a metaphoric idiom. Main effects and simple effects of condition show that non-literal and literal words are processed faster than unrelated words, which indicates that literal meanings of idioms may indeed by activated until after the idiom has been read. A simple effect of idiom type suggests that outside of condition effects, reactions to metonymic idioms are faster than to metaphoric idioms. Similarly, non-literalness ratings of the idioms seem to have an effect on words and nonwords in general, suggesting that higher non-literalness leads to faster processing. The reason might be that high non-literalness might make idioms more easily recognizable.

Table 4
Experiment 2: Regression output with interaction term

variable	b	SE	t	Pr(>|t|)
(Intercept)	*−1.22*	*0.03*	*−38.21*	<0.001 ***
literal	−0.05	0.02	−2.11	0.035 *
non-literal	−0.08	0.02	−3.65	<0.001 ***
nonword	0.07	0.02	3.08	<0.001 **
idiom meto	−0.03	0.04	−0.82	0.412
length	0.02	0.01	2.72	0.007 **
frequency	−0.05	0.01	−7.13	<0.001 ***
swoemitt	−0.03	0.01	−2.20	0.028 *
literal:idiom meto	−0.03	0.04	−0.79	0.429
non-literal:idiom meto	−0.02	0.04	−0.62	0.537
nonword:idiom meto	−0.06	0.04	−1.56	0.120

Conditional R^2: 0.503. RT are inversed.

Correctness

As errors themselves can reveal valuable information about processing difficulties (Cutler, 1981) and excluding all errors in lexical decision analysis has disadvantages (Diependaele, Brysbaert, & Neri, 2012), we analyzed whether idiom type and condition affected correctness of the lexical decisions. A binomial logistic regression predicting the lexical decisions 'word' and 'nonword' was performed. To increase the number of incorrect cases and make results more reliable, the two datasets were analyzed together. Correct vs. incorrect responses were fit as a function of all previous fixed effects and the random intercepts for target strings and participants. Once again, length ($b = 0.17$, $p = 0.07$) and frequency ($b = 0.39$, $p < 0.001$) had effects, but only frequency was significant: the words were more often wrongly classified than the nonwords, and the more frequent words were more often wrongly classified than the less frequent words. Furthermore, longer strings were more error-prone, but this was only a tendency. Literally ($b = 0.59$, $p = 0.06$) and non-literally related ($b = 0.69$, $p = 0.03$) words showed main effects, but only non-literally related words were significant. Idiom type and the interaction term had no effect. Without the interaction term and the separate predictors of condition and idiom type instead, results remained the same. Idiom type remained nonsignificant while non-literally related words were significant ($b = 0.63$, $p = 0.02$) and a non-significant effect was found for literally related words ($b = 0.49$, $p = 0.06$).

Results of response correctness show that non-literally related words were more likely to be wrongly classified as nonwords than both unrelated and literally related words. Given that the reactions to the non-literal words were faster than to the other words, they seem to be easier to process thanks to their close relation to the idiom. Consequently, our finding might occur because the faster reactions may be more prone to slips, meaning that being fast makes it more likely to accidentally press the wrong key. The same, although weaker, pattern was found for literally related words which would be linked to the same interpretation. Whether the idiom prime was metonymic or metaphoric, however, did not affect the correctness of the lexical decision, which also was not particularly hypothesized.

Null Results and the Bayesian Framework

The findings of Experiments 1 and 2 present only weak and mostly non-significant effects. Consequentially, it is important to ensure whether the findings indicate a truly nonexistent effect, or whether they are in fact false negatives. This might be the case because our standard errors are rather large which makes detecting effects difficult. On the other hand, the frequentist linear mixed modelling performed here is based on null-hypothesis significance testing, which tests whether we should reject or fail to reject the null hypothesis that there is no effect. This is a dichotomous decision determined by the p-value for an effect. There are several weaknesses to this approach. One is that the results and ensuing dichotomous decision rely heavily on parameters like sample size and standard errors. Another is that the p-value can neither express how likely nor how large the effect is in reality, and it does not allow any inference on reality. It is also impossible to obtain evidence FOR the null hypothesis in the NHST framework, i.e. it cannot reveal evidence that effects truly DO NOT exist. In consequence, we can never accept the null, but only fail to reject it. This is unsatisfactory if we want to confirm that there is indeed no effect.

In a case as ours, where effects are very small and mostly nonsignificant and standard errors are large, the risk of a type II error – the probability of falsely concluding no effect exists – is also high. To decrease type II error probability, we would have to massively increase power (von der Malsburg & Angele, 2017), preferably to the desired 80%. Calculations show that with 550 participants and an effect size of −20ms for the interaction term 'literal word with metonymic idiom', power would still be 40%. We observed only 10 and −10ms in the raw RT regressions. For 80% power, we would need an effect size between −30 and −40ms, even with 550 participants. Such requirements are not feasible for us to achieve in real life. Even if they were, increasing the sample size would still not mitigate the disadvantages besides the type-II-error probability.

A solution to deal with more of the NHST weaknesses, Bayesian linear mixed analysis is an adequate method. Contrary to the frequentist framework, evidence FOR the null can be obtained in the Bayesian framework. We can model from the experimental data a likely true effect size and direction. Instead of forcing a dichotomous reject/fail to reject answer, we can evaluate how convincing our specific hypothesis is, based on our data (Marusch, Jäger, Neiß, Burchert, & Nickels, 2019).

In the Bayesian analysis, we first provide the prior information we have about the relevant parameters of our experiment and beyond. In our case, especially effect sizes are relevant and we can provide them by their estimates and standard deviations. This information is translated into prior distributions. Our priors can be highly, weakly, or non-informative, depending on what we know about our effect sizes prior to analyzing the data (Sorensen, Hohenstein, & Vasishth, 2016). By running thousands or tens of thousands of simulations on our experimental data using these prior distributions, the Bayesian regression calculates so-called posterior probability distributions for each of our effects sizes. This means that instead of a point estimate in the experimental frequentist model, we get a large set of possible values for an effect size. This resulting value range projects realistic findings better than a point estimate because in reality, repeating the same experiment under identical conditions would also yield many different effect sizes. 95% credible intervals "represent the range within which we are (1−α) % certain that the true value of the parameter lies"[4] (Sorensen et al., 2016) and thus are a means to make inferences onto reality from our findings. If the credible intervals include 0, an effect is at least doubtful, possibly nonexistent.

Lastly, another means of inference is a Bayes factor which tests the strength of the evidence for a hypothesis. The Bayes factor "is the probability of the data under one model relative to that under another" (Rouder & Morey, 2012). This means that it calculates from the given data the probability of one hypothesis (or specific effect sizes) compared to an alternative.

Bayesian Linear Mixed Regression Analysis

We analyzed the data from both Experiment 1 and 2 using the R packages 'brms' 2.10.0 (Bürkner, 2019) and 'bayestestR' 0.4.0 (Makowski et al., 2019). We applied the same models as in the frequentist analyses, except for the random slopes for frequency. This was decided because it did not change results and decreased work load for the system significantly. Thus, the two Bayesian models had random intercepts for targets and participants while the fixed effects were length, frequency, age, non-literalness rating, and the 4×2 interaction of condition by idiom type. As we found no literature on primed lexical decision for metaphoric versus metonymic idioms, let alone for the interaction of idiom type and condition, we ran the models once with non-informative priors for all effects with M = 0ms and SD = 30ms, which equals a null model.

Some information is available on effects of age, frequency, and length on lexical decisions. Thus we calculated average means and sd's from the following experiments: for length, we used New et al. (2006), for frequency, we used Adelman, Brown, and Quesada (2006), Andrews (1989), Chumbley and Balota (1984), Carreiras et al. (2006), K. Forster and Chambers (1973), Schilling, Rayner, and Chumbley (1998), and for age, we used Bowles and Poon (1988) and Macdonald (2013). Moreover, lexical decisions comparing literal, non-literal, and unrelated words have also shown different effects. Thus, we calculated average means and sd's from following experiments resembling our own (Cacciari & Tabossi, 1988; Caillies & Butcher, 2007; Laurent et al., 2006; Titone et al., 2002; Titone & Libben, 2014). For idiom type, the interaction effects, and non-literalness rating, we kept the uninformative priors. Table 5 presents an overview of the ascertained priors. All calculated models converged on all priors, as can be seen from values of around 1.00 for all 'R hat' values.

To investigate model fits, we compared the null with the partly informed model for data 1, and the null with the partly informed model for data 2. Results were extremely similar in both models for data 1 and in both models for data 2, which indicates that the posterior (the resulting effects) are stable. But the Bayes factors showed in both cases that the null models had better fit (BF01

[4] Note that this is not interchangeable with the definition of the confidence interval in the frequentist framework

Table 5
Informed priors, estimated effect sizes by mean and standard deviation

Variable	M	SD
intercept	735	30
literally rel. words (vs. unrelated)	-26	25
non-literally rel. words (vs. unrelated)	-28	25
frequency	-45	15
length	8	3
age	97	25

effects in ms
calculated from other lexical decision studies

and BF02 ≈ 18, indicating that the observed data were around 18 times more likely under the null models). We thus continued to use the null models. These were run using 4 Markov chains and 10000 iterations of the Markov chain Monte Carlo simulations. An additional 1000 warm-up iterations are suggested (Sorensen et al., 2016) to get more precise results. These were separate from the results and excluded from the analysis.

To parallel the frequentist analysis and test for simple effects of single predictors, the Bayesian models were also tested without the interaction term and with the fixed effects of condition and idiom type as additive predictors instead. The same set-up as for the null models was used. Lastly, we tested posterior probabilities (confidence levels) and Bayes factors to evaluate how likely our original hypotheses or null hypotheses were. To exclude extremely small and possibly chance effects close to zero, we defined a null region that included effects sizes ranging from -10 to 10ms. These were classified as 'no effect' and considered congruent with the null hypotheses. In interpreting the Bayes factors, we applied the guidelines by Raftery (1995).

An additional indicator of whether metonymic idioms are processed as more literal is whether the difference between unrelated and literal words is larger for metonymic than for metaphoric idioms. As metaphoric idioms are more non-literal, literal words ensuing them could be treated more similarly to unrelated words, i.e. participants might react to them more slowly, as they react to unrelated words. The probability of this theory can be tested well in a Bayesian paradigm.

Results

Results were similar, yet not identical to the frequentist models. The regression outputs for the interaction models are in Table 6. Posterior probabilities and Bayes factors for individual effect sizes are in Table 7. The interaction terms indicate that responses to the target words were either not or only slightly different for metonymic compared to metaphoric idiom primes, while the large SE's in all cases indicate additional insecurity for any effects found (see Figure 2 for the plotted interactions).

Generally, responses to the literal condition were roughly equally fast with metonymic and metaphoric idioms in data 1 ($\hat{\beta}_{10\ data1} = -4.4$). The Bayes Factor (BF) of 3.45 means that this beta is 3.45 times more likely under the null than under the alternative hypothesis that there is a difference. In data 2, there was a tendency for faster responses with metonymic idioms ($\hat{\beta}_{10\ data2} = -17.2$), yet the effect was uncertain as indicated by the 95%-CI which includes 0. The BF of 1.08 suggests that this finding is about as likely under the null as under the alternative hypothesis. Responses in the non-literal condition, however, seemed to be faster for the metonymic idioms ($\hat{\beta}_{11\ data1} = -17.6$, $\hat{\beta}_{11\ data2} = -14.7$) than for the metaphoric idioms, yet once again, the effect was possible, but uncertain, as

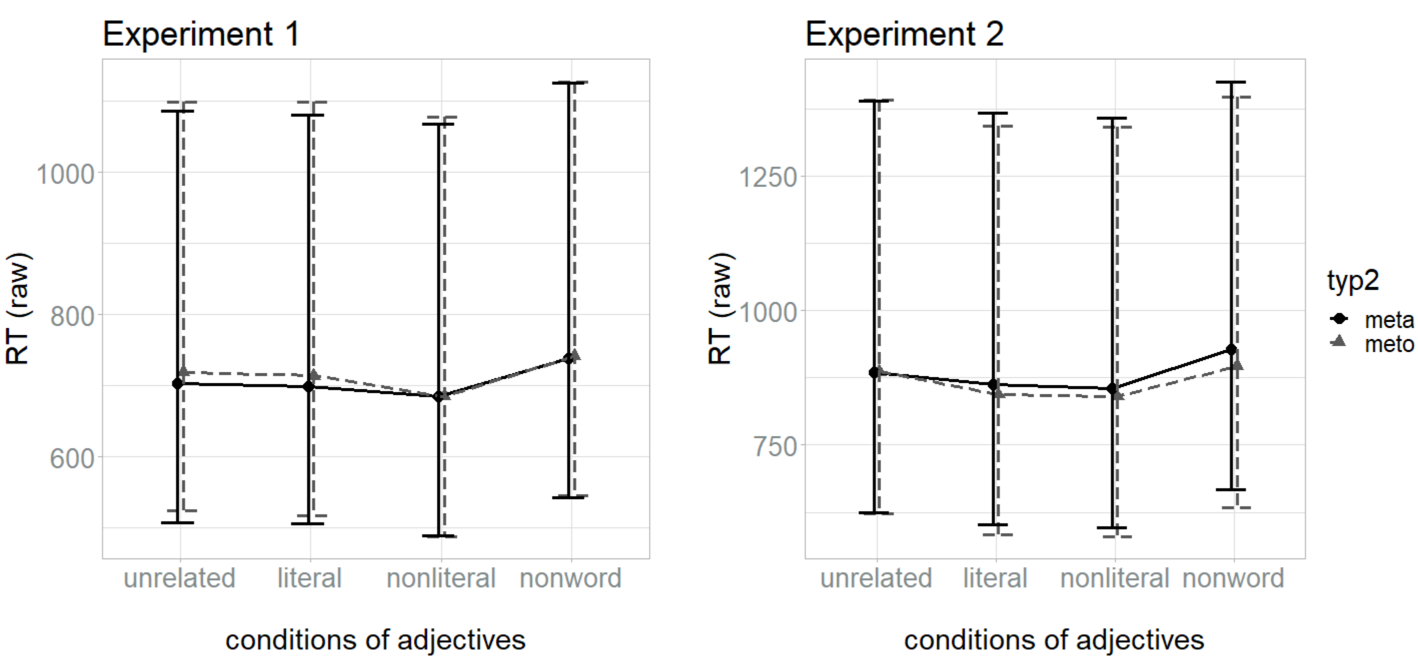

Figure 2. Interaction of semantic relation and idiom type, Bayesian models

Table 6
Bayesian regression output with interaction term

Data 1

variable	$\hat{\beta}$	SE	95%CI lower	95%CI upper	Rhat
(Intercept)	*712.3*	*14.13*	*684.34*	*740.05*	*1.00*
literal	-2.5	8.04	-18.29	13.31	1.00
non-literal	-18.88	7.96	-34.58	-3.4	1.00
nonword	35.04	8.31	18.71	51.48	1.00
idiom meto	17.25	12.03	-6.63	40.66	1.00
age	37.51	12.2	13.1	61.21	1.00
length	9.91	2.61	4.75	15.01	1.00
frequency	-11.41	2.65	-16.65	-6.2	1.00
swoemitt	3.35	4.89	-6.25	12.97	1.00
lit:idiom meto	-4.39	12.59	-29.3	20.27	1.00
nonlit:idiom meto	-17.63	12.26	-41.92	6.26	1.00
nonw:idiom meto	-13.72	12.25	-37.65	10.25	1.00

Data 2

variable	$\hat{\beta}$	SE	95% CI lower	95% CI upper	Rhat
(Intercept)	*884.25*	*14*	*856.75*	*911.95*	*1.00*
literal	-22.41	9.07	-40.23	-4.83	1.00
non-literal	-29.28	8.92	-46.91	-11.83	1.00
nonword	41.65	9.26	23.46	59.59	1.00
idiom meto	-1.05	12.9	-26.7	24.24	1.00
age	14.64	10.81	-6.46	35.86	1.00
length	9.79	2.95	4.07	15.58	1.00
frequency	-17.96	3	-23.9	-12.07	1.00
swoemitt	-8.03	5.41	-18.57	2.52	1.00
lit:idiom meto	-17.2	13.48	-43.62	9.35	1.00
nonlit:idiom meto	-14.71	13.24	-40.73	11.28	1.00
nonw:idiom meto	-30.78	13.4	-56.76	-4.25	1.00

RT are raw in ms.
CI = credible interval
Rhat = 1.00 indicates successful convergence

implied by the CIs. The BFs imply equally strong evidence for the null and the alternative (BF_{data1} =1) and slightly stronger evidence for the null (BF_{data2} = 1.39). Large SE's of 12.2 to 13.4ms in all interaction terms add uncertainty, so that none of these merely possible effects can be said to be clear and stable.

As for the difference between unrelated and literal words depending on idiom type, confidence was 33% that this difference was indeed larger (4ms) for metonymic than for metaphoric idioms. Given that this confidence level is rather low and the difference would be close to zero, responses in the non-literal condition were very likely equally fast for metonymic as for metaphoric idioms. This conclusion was strengthened by the large CI_{data1} [-20.25, 29.0] centering at close to 0, and by a BF of 0.49 (that is, the data were 2.04 times more likely under the null). In data 2, evidence was stronger for a larger difference between the unrelated and literal condition for metonymic idioms: the confidence was 71%, and the difference was estimated at 17ms. However, from the CI_{data2} of [-9.2, 43.5] which includes 0, an effect was possible but not definite. The BF of 2.42 (finding the data 2.42 times more likely under alternative hypothesis) offers weak evidence in favor of a difference over 10ms.

Regarding non-literalness ratings, no effect was found in data 1, and the BF of 33.33 offers strong evidence for the null and against an effect. While the BF of data 2 is much lower at 5.00, it is still positive evidence for the null and against the alternative hypothesis that higher levels of non-literalness of the idioms can be linked to faster responses in all conditions of the target strings.

The Bayesian additive models for data 1 and 2 were partly similar. In both datasets, a processing advantage was found for non-literal words, as expected (*data 1*: $\hat{\beta} = -24.9$, $\widehat{SE} = 6.9$, CI [−38.5 −11.4]; *data 2*: $\hat{\beta} = -33.9$, $SE = 7.8$, CI [−49.4, −18.6]). Neither dataset showed an effect for idiom type (*data 1*: $\hat{\beta} = 8.8$, $\widehat{SE} = 10.1$, CI [−11.2, 28.6]; *data 2*: $\hat{\beta} = -14.8$, $SE = 11.3$, CI [−37.0, 7.1]) or non-literalness (*data 1*: $\hat{\beta} = 3.6$, $\widehat{SE} = 4.8$, CI [−5.88, 13.1]; *data 2*: $\hat{\beta} = -7.2$, $\widehat{SE} = 5.5$, CI [−18.0, 3.5]). Only data 2 suggests that literal words are processed faster than unrelated words (*data 2*: $\hat{\beta} = -28$, $\widehat{SE} = 7.9$, CI [−43.6, −12.5]), while no processing advantage was found for literal words in

Table 7
Hypotheses, mean effect sizes, posterior probabilities (PP, "confidence"), Bayes factors (BF)

null hypotheses	mean($\hat{\beta}$)	PP	BF	evidence for null?
literal: meto - no effect[5]	-4.4	55%	3.45	yes, positive
literal: meto - no effect	-17.2	27%	1.08	yes, weak
non-literal: meto - no effect	-17.6	25%	1	equal
non-literal: meto - no effect	-14.7	33%	1.39	yes, weak
nonlit. rating - no effect	3.4	n.A.[6]	33.33	yes, strong
nonlit. rating - no effect	-8.3	n.A.	5.00	yes, positive
alternative hypotheses				**evidence for alternative?**
meto unrel-lit vs. meta unrel-lit vs. > 10ms	4.1	33%	0.49	no, weak (2.04) for null
meto unrel-lit vs. meta unrel-lit vs. > 10ms	17.4	71%	2.42	yes, weak
literal vs. unrel < -10ms (main effect)	-2.5	18%	0.11	no, positive (9.09) for null
literal vs. unrel < -10ms (main effect)	-22.41	44%	3.62	yes, positive
non-literal vs. unrel < -20ms (main effect)[7]	-18.88	91%	2.33	yes, weak
non-literal vs. unrel < -20ms (main effect)	-29.28	85%	24.27	yes, strong

Data 1
Data 2
unrel = unrelated words
lit = literally related words
nonlit = non-literally related words

data 1 ($\hat{\beta} = -3.6$, $\widehat{SE} = 6.9$, CI [−17.2, 9.9]).

Unsurprisingly, effects of frequency and length are present in both datasets, confirming increasing length and decreasing frequency to lead to slower reaction times. An age effect seemed to be present in data 1, although the large SE (*data 1*: $\hat{\beta} = 37.5$, $\widehat{SE} = 12.2$) together with the lower bound of the 95% CI of 13.1 suggest that it could be close to nonexistent. In data 2, an age effect seemed theoretically possible (95% CI [−6.46, 36.86]), but rather unlikely and very weak (data 2: $\hat{\beta} = 14.6$, $\widehat{SE} = 10.8$).

Discussion

Overall, we can infer from the Bayesian results that all interactive and non-literalness effects of idioms are uncertain and small, or nonexistent. Unfortunately, credible intervals that cross 0 in or close to the tails make it challenging to unambiguously infer that effects either do or do not exist. Bayes factors provide additional evidence but are not always strong in favoring the null over the alternative hypothesis, or vice versa. In addition, findings from the two experimental data partly offer different findings. For the sake of clarity and orientation for the reader, we will assume a conservative view in discussing these findings.

Our findings suggest that the speed of reactions to non-literal words does not depend on whether their prime is a metonymic or a metaphoric idiom. It is also unlikely that reactions to literal words are faster when preceded by a metonymic idiom. Evidence for the null hypothesis that there is no effect is not compelling, but stronger than evidence for the alternative hypothesis (that there is a processing advantage for literal words with metonymic idioms). From this we conclude that there is no processing advantage. If there was a processing advantage for literal words tied to metonymic idioms, it would be very small.

Evidence against an effect from the non-literalness ratings is rather strong, as the tiny effect size with the comparably large SE, BF and comparably high confidence value suggest. We conclude that non-literalness ratings of idiom primes do not affect the processing of ensuing words.

Comparison between Findings from the Frequentist and Bayesian Analyses

In the frequentist paradigm, some conventions are particularly set to enable clearcut answers to whether effects exist or not. In contrast, Bayesian modelling is not designed to provide such dichotomous answers and often, results may rather show tendencies. The strength of the Bayesian analyses is that more differentiated results do better justice to reality and can provide evidence for the nonexistence of an effect.

Many findings from the frequentist analyses are in line with the Bayesian analyses. Two results of interest here are congruent across both datasets and frequentist and Bayesian models containing the interaction: Neither the literal nor the non-literal words are processed faster when preceded by a metonymic compared to a metaphoric idiom. While in the frequentist models, the effects are merely nonsignificant, the Bayesian models suggest possibilities that both literal and non-literal words are processed slightly faster

5 effect size between −10ms and 10ms

6 could not be calculated due to missing implementation in packages

7 well-documented effect and assumed to be larger than in literally related words

with a preceding metonymic idiom. But given their small size and low certainty, as well as some evidence for the absence of effects, we conclude there are no processing advantages. Even if effects existed, their size would likely be of no practical relevance.

For main effects and non-literal words, evidence is unanimous: they are clearly processed faster than unrelated words in both datasets. The frequentist models show clearly significant effects while the Bayesian models indicate a weaker, but present, advantage. This is backed by all additive models. For literal words, the frequentist and Bayesian model both suggest they are not processed faster than unrelated words in data 1, whereas in data 2, both models show or clearly indicate an effect. We interpret the conflicting evidence such that there is likely to be a small processing advantage for literal words as opposed to unrelated words. This is also found in all additive models for data 2. This indicates that the literal interpretation of idioms may be activated, even until after the idiom has been read, might spill onto the ensuing stimulus and facilitate it if it is literal.

An effect for idiom type is very unlikely, with and without non-literalness ratings included. Only the additive frequentist model for data 2 suggests a significant effect of idiom type. However, evidence for its absence in the other models is stronger, as indicated by effect size close to zero, credible intervals, confidence levels, and Bayes factors of the Bayesian models. There is also no evidence for an effect from the frequentist models for data 1. Similarly, all interaction models and three out of four additive models found either no effect of non-literalness rating, or evidence against it. Once again, only the additive frequentist model for data 2 indicated a significant effect. In both cases, for idiom type and rating, evidence for null effects or missing evidence for present effects is much stronger. We thus conclude that truly existing processing advantages driven by idiom type or non-literalness rating are very unlikely.

General Discussion

In sum, the absence of effects found of different degrees and kinds of non-literalness on the processing speed in lexical decisions suggests there are indeed no such effects in reality. We also conclude true nonexistence for small effects.

Non-literally related words were processed fastest, literally related words had the tendency to be processed second fastest, while unrelated words were processed the most slowly. These results indicate that idioms – here as primes – are activated and processed as whole, independent, semantic entities. As there was no difference found between metaphoric and metonymic idioms, neither the kind nor the degree of non-literalness of idioms seems to affect their activation, at least in automatic processing. This in turn indicates that idiomaticity is the more economic and faster property in semantic processing. In other words, recognizing an idiom as familiar immediately leads to its processing and activation of relevant related words.

Non-literalness of idioms, on the other hand, seems to be a property that does not affect the processing or analysis of an idiom at the early stage of lexical access. Consequently, the metonymic and metaphoric structures do not affect idiom processing at the automatic processing stage. The immediate activation of the entire idiomatic meanings causes spreading activation to relevant words and does not differentiate between more or less non-literal structures or degrees. Of note, it was found in other primed lexical decisions that figurative priming emerged only at an ISI of 300ms (Cacciari & Tabossi, 1988), that figurative priming was stronger after 500ms than after 0 or 350ms (Caillies & Butcher, 2007), and that the activation of an idiom's figurative meaning only slowly develops and peaks at an ISI of 1000ms (Titone & Libben, 2014). Yet, these experiments all used neutral prime sentence that did not bias an idiomatic interpretation. For a setting such as ours, where idiomatic interpretations are biased, figurative interpretations become available much sooner, as our results show and as is also found in a similar experiment by Cacciari and Tabossi (1988), and by Fanari, Cacciari, and Tabossi (2010).

Also, literally related words showed a tendency to be relevant to idiom processing, which indicates that literal meanings of idioms are at least partially activated as well by spreading activation. Naturally, literally related words are less relevant to the idiomatic interpretation than the non-literally related words whose activation is clearly stronger. Our results indicate that the advantage for literally related word is small compared to unrelated words, as a pronounced effect could only be found in our second experiment with a larger sample. Yet clearly at this early processing stage, literal interpretations are activated, while the non-literal interpretations are dominant.

The lexical decision task seems to be prone to large variances (see also Bambini et al., 2013; Findlay & Carrol, 2018). In our experiments, there is not only very large between-subjects variance in reaction times, but also considerable within-subjects variance. It is possible that the large standard errors are anchored within the task itself (Tabossi, 1996). Large standard errors, however, might cloud small effects that actually exist at this early processing stage. However, an alternative interpretation seems more likely: It is reasonable to assume that automatic processing is primarily guided by the recognition of an idiom as familiar and its correct interpretation. Even if processing an idiom as more or less non-literal were a step in automatic processing, recognition and correct interpretation would be the more economic and more goal-oriented steps for the processing system to execute first, at least in any task that demands fast comprehension and fast reactions that are assessed as correct or incorrect and are not opinion-based.

It is also possible that effects of non-literalness are rather deeply rooted within analytical abilities and can be measured best in offline cognitive analytical tasks. Our findings from the rating study showed a very clear difference in how literal metonymic idioms were rated compared to metaphoric idioms. The experimental method of rating on Likert scales is entirely different from a lexical decision task. In the rating study, participants were asked to reflect and judge and had a range of possible answers. A cross-modal primed lexical decision, on the other hand, demands strictly focused attention, listening and reading, and highly time-sensitive binary reactions that can be correct or incorrect. Its cognitive demands are in stark contrast to those of a rating study. It is likely that such contrary tasks produce very different results. At any rate, these results indicate that processing an idiom as rather literal or rather non-literal is not an automatic process.

An alternative, related reason for the absence of effects could be that effects of non-literalness are only linked to controlled processing, not to automatic processing that we tested here. In that case, non-literalness effects would be linked to expectancy-based strategies or contextual effects, rather than spreading activation, and thus should only arise at an ISI around 400ms (Carter et al., 2011). As Moss et al. (1995) point out, a 200ms ISI is "approximately the shortest interval at which participants can keep up with the rate of presentation" (Moss et al., 1995, p. 20). Since it seems that at the stage of automatic processing, idiomaticity takes the lead, a potential non-literalness effect might be too slow or too strongly linked to analytical abilities to emerge at only 200ms. In addition, at this early stage in processing the high level of engagement necessary in primed lexical decision might only benefit the most economic and effective cognitive effort, which is recognition and activation of the idiom as an idiom. This would lead to the postponement of a non-literalness effect.

Our findings of absent effects are also backed by very similar priming effects of literal and non-literal meanings at an ISI of 0ms found by Colombo (1993). On the other hand, studies that did not focus on immediate, automatic processing have found effects of non-literalness: self-paced reading studies with mostly the same idioms as used here (Michl, 2019b), as well as the sensicality judgment study by Bambini et al. (2013) suggest that non-literalness is a factor during controlled processing. For this reason, it seems likely that the effects of non-literalness would follow the automatic processing phase focused on in the present study. In our reading experiments, participants were asked to read as fast as possible, but also for correct comprehension, and could choose their own speed (Michl, 2019b). In the study by Bambini et al. (2013), which is comparable to a lexical decision without priming, participants also had to read metaphoric and metonymic sentences for comprehension and judge whether they made sense or not. Our current experiments are in contrast to both other studies in that participants had primes and did not have time to "think about" or check their understanding of the sentence immediately after playback, but were directly presented with the ensuing part of the trial, namely the target word that required a fast reaction.

As pointed out by one anonymous reviewer, a potential impact by the specific language investigated on the findings should be considered. The phenomena of metonymy, metaphor, and idiomatic expressions are likely pervasive throughout most, if not all, languages. At least in studies investigating idioms in various Germanic or Romance languages, idioms are commonly defined as having a structure of at least 'verb + noun phrase' (such as 'break the ice') to, maximally, 'verb + (preposition +) noun phrase + (preposition +) noun phrase)' (such as 'stand with one's back to the wall'). These structures are also typical of German idioms. Moreover, whenever an immediate response to the idiom is the object of interest in idiom processing studies, it is common practice to build idiomatic prime sentences such that they end in the idiom predicate because then the last word renders the idiom complete and recognizable. This experimental practice was also applied in the current experiments. Due to these consistencies for idioms across languages and available idiom research, it is unlikely that any idiosyncrasies of the German language impacted our results. Thus, our findings should not be restricted to German, but apply to metonymy and metaphor in idioms as defined here in general.

Whether non-literalness is reflected in a predefined time window of controlled processing, is a different question with different implications altogether, and we did not pursue it here. To test it in a primed lexical decision as here, we recommend implementing a longer ISI of at least 400ms. This might also lead to larger differences between non-literal and literal conditions, and to larger differences between related and unrelated conditions (Groot, 1984).

Conclusion

In summary, results indicate that there is no difference to how fast words are processed, whether they are primed by a metonymic or metaphoric idiom. This in turn suggests that differences in non-literal structure do not have an effect in the automatic processing of idioms. Processing speed is also not determined by how strongly non-literal the idiom is.

It is clear that non-literally related words are processed faster than unrelated and literal words because they are most closely related to the meaning of the preceding idiom. It is possible that literal words are in fact processed faster than unrelated words and this may indicate that literal meanings of idioms are activated until after the idiom has been processed. Furthermore, processing an idiom as literal or non-literal is likely linked to controlled processing, which we did not test here.

References

Acha, J., & Perea, M. (2008). The effect of neighborhood frequency in reading: Evidence with transposed-letter neighbors. *Cognition*, *108*(1), 290–300. https://doi.org/10.1016/j.cognition.2008.02.006

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, *17*(9), 814–823. https://doi.org/10.1111/j.1467-9280.2006.01787.x

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 802–814. https://doi.org/10.1037/0278-7393.15.5.802

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, *4*(4), 439–461. https://doi.org/10.3758/BF03214334

Annaz, D., van Herwegen, J., Thomas, M., Fishman, R., Karmiloff-Smith, A., & Rundblad, G. (2009). Comprehension of metaphor and metonymy in children with Williams syndrome. *International Journal of Language & Communication Disorders*, *44*(6), 962–978. https://doi.org/10.3109/13682820802525005

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Special Issue: Emerging Data Analysis*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bambini, V., Ghio, M., Moro, A., & Schumacher, P. B. (2013). Differentiating among pragmatic uses of words through timed sensicality judgments. *Frontiers in Psychology*, *4*, 938. https://doi.org/10.3389/fpsyg.2013.00938

Barton, K. (2019). R package 'MuMin' (Version 1.43.6) [Computer software]. Retrieved from Retrieved from <https://cran.r-project.org>.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2019). R package 'lme4' (Version 1.1-21) [Computer software]. Retrieved from Retrieved from <https://cran.r-project.org>

Bowles, N. L., & Poon, L. W. (1988). Age and context effects in lexical decision: An age by context interaction. *Experimental Aging Research*, *14*(4), 201–205. https://doi.org/10.1080/03610738808259748

Bürkner, P.-C. (2019). R package 'brms' (Version 2.10.0) [Computer software]. Retrieved from Retrieved from <https://cran.r-project.org>

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, *27*, 668–683.

Caillies, S., & Butcher, K. (2007). Processing of Idiomatic Expressions: Evidence for a New Hybrid View. *Metaphor and Symbol*, *22*(1), 79–108. https://doi.org/10.1080/10926480709336754

Canal, P., Pesciarelli, F., Molinaro, N., Vespignani, F., & Cacciari, C. (2015). Electrophysiological correlates of idiom comprehension: Semantic composition does not follow lexical retrieval. In Ceur (Chair), *NetWordS Final Conference on Word Knowledge and Word Usage: Representations and Processes in the Mental Lexicon (NetWordS 2015)*. Retrieved from https://iris.unimore.it/bitstream/11380/1066269/2/paper21.pdf

Canal, P., Pesciarelli, F., Vespignani, F., Molinaro, N., & Cacciari, C. (2017). Basic composition and enriched integration in idiom processing: An EEG study. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(6), 928–943. https://doi.org/10.1037/xlm0000351

Caramazza, A. (1997). How Many Levels of Processing Are There in Lexical Access? *Cognitive Neuropsychology*, *14*(1), 177–208. https://doi.org/10.1080/026432997381664

Carreiras, M., Mechelli, A., & Price, C. J. (2006). Effect of word and syllable frequency on activation during lexical decision and reading aloud. *Human Brain Mapping*, *27*(12), 963–972. https://doi.org/10.1002/hbm.20236

Carter, M. D., Hough, M. S., Stuart, A., & Rastatter, M. P. (2011). The effects of inter-stimulus interval and prime modality in a semantic priming task. *Aphasiology*, *25*(6-7), 761–773. https://doi.org/10.1080/02687038.2010.539697

Chumbley, J. I., & Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory & Cognition*, *12*(6), 590–606. https://doi.org/10.3758/BF03213348

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Colombo, L. (1993). The comprehension of ambiguous idioms in context. In C. Cacciari & P. Tabossi (Eds.), *Idioms: Processing, structure, and interpretation* (pp. 163–200). New York, New York, Hove, England: Psychology Press.

Conklin, K., & Schmitt, N. (2008). Formulaic Sequences: Are They Processed More Quickly than Nonformulaic Language by Native and Nonnative Speakers? *Applied Linguistics*, *29*(1), 72–89. https://doi.org/10.1093/applin/amm022

Conklin, K., & Schmitt, N. (2012). The Processing of Formulaic Language. *Annual Review of Applied Linguistics*, *32*, 45–61. https://doi.org/10.1017/S0267190512000074

Cronk, B., & Schweigert, W. A. (1992). The comprehension of idioms: The effects of familiarity, literalness, and usage. *Applied Psycholinguistics*, *13*(2), 131–146.

Cutler, A. (1981). The reliability of speech error data. *Linguistics*, *19*(7-8), 561–582. https://doi.org/10.1515/ling.1981.19.7-8.561

Diependaele, K., Brysbaert, M., & Neri, P. (2012). How Noisy is Lexical Decision? *Frontiers in Psychology*, *3*, 348. https://doi.org/10.3389/fpsyg.2012.00348

Drummond, A. (2013). Ibex (Version 0.3.7) [Computer software]. Retrieved from Accessible at <http://spellout.net/ibexfarm>

Fanari, R., Cacciari, C., & Tabossi, P. (2010). The role of idiom length and context in spoken idiom comprehension. *European Journal of Cognitive Psychology*, *22*(3), 321–334. https://doi.org/10.1080/09541440902843866

Fauconnier, G., & Turner, M. (1996). Blending as a central process of grammar. *Conceptual Structure, Discourse, and Language*, *113*, 130–153.

Field, A. P., Miles, J., & Field, Z. (2014). *Discovering statistics using R* (Repr). London: Sage.

Findlay, H., & Carrol, G. (2018). Contributions of semantic richness to the processing of idioms. *The Mental Lexicon*, *13*(3), 311–332. https://doi.org/10.1075/ml.18014.fin

Forster, J., & Forster, K. (2003). DMASTR (Version 5.1.3.3) [Computer software].

Forster, K., & Chambers, S. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 627–635. https://doi.org/10.1016/S0022-5371(73)80042-8

Fox, J., Weisberg, S., & Price, B. (2019). R package 'car' (Version 3.0-3) [Computer software]. Retrieved from Retrieved from <https://cran.r-project.org>

Gernsbacher, M., & Robertson, R. (1999). The role of suppression in figurative language comprehension. *Journal of Pragmatics*, *31*(12), 1619–1630. https://doi.org/10.1016/S0378-2166(99)00007-7

Gibbs, R. W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition*, *8*(2), 149–156. https://doi.org/10.3758/BF03213418

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, *8*(3), 183–206. https://doi.org/10.1515/cogl.1997.8.3.183

Groot, A. de (1984). Primed Lexical Decision: Combined Effects of the Proportion of Related Prime-Target Pairs and the Stimulus-Onset Asynchrony of Prime and Target. *The Quarterly Journal of Experimental Psychology Section a*, *36*(2), 253–280. https://doi.org/10.1080/14640748408402158

Groot, A. de, Thomasson, A., & Hudson, P. (1986). Primed-lexical decision: The effect of varying the stimulus-onset asynchrony of prime and target. *Acta Psychologica*, *61*(1), 17–36. https://doi.org/10.1016/0001-6918(86)90019-3

Hutchison, K. A., Balota, D. A., Neely, J. H., Cortese, M. J., Cohen-Shikora, E. R., Tse, C.-S., . . . Buchanan, E. (2013). The semantic priming project. *Behavior Research Methods*, *45*(4), 1099–1114. https://doi.org/10.3758/s13428-012-0304-z

Johnson, R. L. (2009). The quiet clam is quite calm: Transposed-letter neighborhood effects on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 943–969. https://doi.org/10.1037/a0015572

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*(2), 228–238. https://doi.org/10.1037/0096-3445.111.2.228

Katz, A. N., Paivio, A., Marschark, M., & Clark, J. M. (1988). Norms for 204 Literary and 260 Nonliterary Metaphors on 10 Psychological

Dimensions. *Metaphor and Symbolic Activity*, *3*(4), 191–214. https://doi.org/10.1207/s15327868ms0304_1

Klepousniotou, E. (2002). The Processing of Lexical Ambiguity: Homonymy and Polysemy in the Mental Lexicon. *Brain and Language*, *81*(1-3), 205–223. https://doi.org/10.1006/brln.2001.2518

Kuznetsova, A., Brockhoff, P., & Haubo Bojesen Christensen, R. (2019). R package 'lmerTest' (Version 3.1-0) [Computer software]. Retrieved from Retrieved from <https://cran.r-project.org>.

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

Laurent, J.-P., Denhières, G., Passerieux, C., Iakimova, G., & Hardy-Baylé, M.-C. (2006). On understanding idiomatic language: The salience hypothesis assessed by ERPs. *Brain Research*, *1068*(1), 151–160. https://doi.org/10.1016/j.brainres.2005.10.076

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*. (10), 707–710.

Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory & Cognition*, *36*(6), 1103–1121. https://doi.org/10.3758/MC.36.6.1103

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171. https://doi.org/10.3389/fpsyg.2015.01171

Loy, A. (2015). R package 'HLMdiag' (Version 0.3.1) [Computer software]. Retrieved from Retrieved from <https://cran.r-project.org>

Macdonald, G. (2013). Aging and Semantic Processing. *Electronic Theses and Dissertations*. Retrieved from https://scholar.uwindsor.ca/etd/4738

Makowski, D., Lüdecke, D., Ben-Shachar, M. S., Wilson Michael D., Bürkner, P.-C., Mahr, T., … Gronau, Q. F. (2019). bayestestR (Version 0.4.0) [Computer software]. Retrieved from Retrieved from <https://cran.r-project.org>

Malsburg, T. von der (2018). *Inversing is the most adequate transformation for the nature of reaction times*. Potsdam.

Malsburg, T. von der, & Angele, B. (2017). False Positives and Other Statistical Errors in Standard Analyses of Eye Movements in Reading. *Special Issue: Emerging Data Analysis*, *94*, 119–133. https://doi.org/10.1016/j.jml.2016.10.003

Marusch, T., Jäger, L. A., Neiß, L., Burchert, F., & Nickels, L. (2019). Overt language production of German past participles: investigating (ir-)regularity. *Language, Cognition and Neuroscience*, *34*(3), 289–308. https://doi.org/10.1080/23273798.2018.1527936

Mendoza Ibánez, F. de (2003). The role of mappings and domains in understanding metonymy. In A. Barcelona (Ed.), *Metaphor and metonymy at the crossroads: A cognitive perspective* (pp. 109–132). Berlin/New York: Mouton de Gruyter.

Michl, D. (2019a). Metonymies are more literal than metaphors: Evidence from ratings of German idioms. *Language and Cognition, 11*(1), 98–124. doi:10.1017/langcog.2019.7

Michl, D. (2019b). Speedy Metonymy, Tricky Metaphor, Irrelevant Compositionality: How Nonliteralness Affects Idioms in Reading and Rating. *Journal of Psycholinguistic Research* 48, 1285–1310. https://doi.org/10.1007/s10936-019-09658-7

Moss, H. E., Tyler, L. K., Hodges, J. R., & Patterson, K. (1995). Exploring the loss of semantic memory in semantic dementia: Evidence from a primed monitoring study. *Neuropsychology*, *9*(1), 16–26. https://doi.org/10.1037/0894-4105.9.1.16

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, *13*(1), 45–52. https://doi.org/10.3758/BF03193811

Nippold, M. A., & Taylor, C. L. (2002). Judgments of Idiom Familiarity and Transparency. *Journal of Speech Language and Hearing Research*, *45*(2), 384–391. https://doi.org/10.1044/1092-4388(2002/030)

Omazić, M. (2008). Processing of idioms and idiom modifications: A view from cognitive linguistics. In F. Meunier & S. Granger (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 67–80). Amsterdam, Philadelphia: John Benjamins Pub.

Posner, M. I., & Snyder, C. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 55–85). New York: Wiley & Sons.

R Core team (2019). R (Version 3.6.1) [Computer software]. Retrieved from Retrieved from <https://cran.r-project.org>

R Studio team (2019). RStudio (Version 1.2.1335) [Computer software]. Retrieved from Retrieved from <https://cran.r-project.org>

Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, *25*, 111. https://doi.org/10.2307/271063

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532. https://doi.org/10.1037/0033-2909.114.3.510

Rouder, J. N., & Morey, R. D. (2012). Default Bayes Factors for Model Selection in Regression. *Multivariate Behavioral Research*, *47*(6), 877–903. https://doi.org/10.1080/00273171.2012.734737

Rundblad, G., & Annaz, D. (2010). Development of metaphor and metonymy comprehension: Receptive vocabulary and conceptual knowledge. *British Journal of Developmental Psychology*, *28*(3), 547–563. https://doi.org/10.1348/026151009X454373

Sailer, M. (2013). Idiom and Phraseology. Oxford Bibliographies Online Datasets.

Schemann, H. (2011). *Deutsche Idiomatik*. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110217896

Schilling, H. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition*, *26*(6), 1270–1281. https://doi.org/10.3758/bf03201199

Schmitt, N., & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt (Ed.), *Language learning and language teaching: Vol. 9. Formulaic sequences: Acquisition, processing and use* (pp. 173–190). Amsterdam: Benjamins.

Schweigert, W. A. (1986). The comprehension of familiar and less familiar idioms. *Journal of Psycholinguistic Research*, *15*(1), 33–45. https://doi.org/10.1007/BF01067390

Smolka, E., Rabanus, S., & Rösler, F. (2007). Processing Verbs in German Idioms: Evidence Against the Configuration Hypothesis. *Metaphor and Symbol*, *22*(3), 213–231. https://doi.org/10.1080/10926480701357638

Sorensen, T., Hohenstein, S., & Vasishth, S. (2016). Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *The Quantitative Methods for Psychology*, *12*(3), 175–200. https://doi.org/10.20982/tqmp.12.3.p175

Spencer, K. A., & Wiley, E. (2008). Response priming patterns differ with interstimulus interval duration. *Clinical Linguistics & Phonetics*, *22*(6), 475–490. https://doi.org/10.1080/02699200801896406

Spieß, C., & Köpcke, K.-M. (Eds.) (2015). *Metapher und Metonymie: Theoretische, methodische und empirische Zugänge*: De Gruyter. Retrieved from https://books.google.de/books?id=kQJfCAAAQBAJ

Sprenger, S., Levelt, W., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, *54*(2), 161–184. https://doi.org/10.1016/j.jml.2005.11.001

Stefanowitsch, A., & Gries, S. T. (2007). *Corpus-based approaches to metaphor and metonymy*. *Mouton select*. Berlin, New York: Mouton de Gruyter.

Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, *18*(5), 523–534. https://doi.org/10.1016/S0022-5371(79)90284-6

Tabossi, P. (1996). Cross-Modal Semantic Priming. *Language and Cognitive Processes*, *11*(6), 569–576. https://doi.org/10.1080/016909696386953

Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast? *Memory & Cognition*, *37*(4), 529–540. https://doi.org/10.3758/MC.37.4.529

Titone, D. A., & Connine, C. M. (1994). Descriptive Norms for 171 Idiomatic Expressions: Familiarity, Compositionality, Predictability, and Literality. *Metaphor and Symbolic Activity*, *9*(4), 247–270. https://doi.org/10.1207/s15327868ms0904_1

Titone, D. A., Holzman, P. S., & Levy, D. L. (2002). Idiom processing in schizophrenia: Literal implausibility saves the day for idiom priming. *Journal of Abnormal Psychology*, *111*(2), 313–320. https://doi.org/10.1037//0021-843X.111.2.313

Titone, D. A., & Libben, M. R. (2014). Time-dependent effects of decomposability, familiarity and literal plausibility on idiom priming: A cross-modal priming investigation. *The Mental Lexicon*, *9*(3), 473–496. https://doi.org/10.1075/ml.9.3.05tit

Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General*, *123*(1), 34–80.

Van de Voort, M., & Vonk, W. (1995). You Don't Die lmmediately When You Kick an Empty Bucket: A Processing View on Semantic and Syntactic Characteristics of ldioms. In M. Everaert (Ed.), *Idioms: Structural and psychological perspectives*. New York: Psychology Press.

Vergara-Martínez, M., & Swaab, T. Y. (2012). Orthographic neighborhood effects as a function of word frequency: An event-related potential study. *Psychophysiology*, *49*(9), 1277–1289. https://doi.org/10.1111/j.1469-8986.2012.01410.x

Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, *22*(8), 1682–1700. https://doi.org/10.1162/jocn.2009.21293

Weiland, H., Bambini, V., & Schumacher, P. B. (2014). The role of literal meaning in figurative language comprehension: Evidence from masked priming ERP. *Frontiers in Human Neuroscience*, *8*, 583. https://doi.org/10.3389/fnhum.2014.00583

(Appendix follows)

Appendix: Examples of Materials

		auditory prime sentence		visual target		
Item no.	Idiom type	Beginning	Idiom	Target	translation	condition
1)	metonymic	Bei der Nachricht	macht Michael einen Luftsprung.	freudig	joyous	Nonliteral
		At the news	makes Michael an air jump.	hüpfend	jumping	Literal
			jump for joy	flüchtig	transient	Unrelated
				knosslich		Nonword
2)	meton.	Seit diesem Jahr	sind alle Töchter aus dem Haus.	erwachsen	grown-up	N
		Since this year	are all daughters out of the house.	auswärts	outside	L
				universell	universal	U
				darbolich		NW
3)	meton.	In der Lehrzeit	hat Marie ihre eigenen vier Wände.	eigenständig	independent	N
		During her apprenticeship	has Marie her own four walls.	räumlich	spatial	L
			live in one's own four walls	gefühllos	insensitive	U
				wopelhaft		NW
4)	meton.	Viele Themen	bespricht das Paar unter vier Augen.	vertraulich	confidential	N
		Many topics	discusses the couple under four eyes.	blickend	gazing	L
			discuss sth. in private	juristisch	juridical	U
				schrull		NW
5)	meton.	Zu dieser Feier	kommt Katrin mit leeren Händen.	knauserig	stingy	N
		To this celebration	comes Katrin with empty hands.	eintreffend	arriving	L
			come empty-handed	besinnlich	tranquil	U
				trägelich		NW
6)	metaphoric	Wieder einmal	sprengt Simons Vorhaben den Rahmen.	ausufernd	sprawling	N
		Once again	bursts Simon's project the frame.	entzwei	broken	L
			go beyond the scope	genüsslich	relishing	U
				beiterlich		NW
7)	meta.	Mit ihrer Behauptung	begibt sich Julia auf Glatteis.	riskant	risky	N
		With her statement	goes Julia on slippery ice.	rutschig	slippery	L
			roughly: walk on thin ice	sympathisch	likeable	U
				keiterreich		NW
8)	meta.	Dieses Argument	rückt Johannes ins rechte Licht.	berichtigend	correcting	N
		This argument	puts John in the right light.	blendend	blinding	L
			put sth. in perspective	gescheit	clever	U
				woderlich		NW
9)	meta.	Für die Freundin	ist Julia ein Fels in der Brandung.	unerschütterlich	steadfast	N
		For the girlfriend	is Julia a rock in the surge.	gewaltig	mighty	L
			be a tower of strength	elegant	elegant	U
				tackelhaft		NW

Received: 1.30.2020
Revised: 4.18.2020
Accepted: 4.18.2020