Affective and psycholinguistic norms for German conceptual metaphors (COMETA)

Francesca M.M. Citron[1], Mollie Lee[1], Nora Michaelis[2]

[1]Free University of Berlin, Germany

[2]Department of Psychology, Lancaster University, UK

Address correspondence to:
　　　　　　　　　Francesca Citron
　　　　　　　　　Department of Psychology
　　　　　　　　　Fylde College
　　　　　　　　　Lancaster University
　　　　　　　　　LA1 4YF
　　　　　　　　　Lancaster, UK

　　　　　　　　　Telephone: +44 1524594573
　　　　　　　　　E-mail: fmm.citron@gmail.com

Running head: Affective norms for metaphors

Abstract

Figurative expressions have been shown to play a special role in evoking affective responses compared to their literal counterparts. This study provides the first database of conceptual metaphors that includes ratings of affective, beyond psycholinguistic properties. To allow investigation of natural reading processes, 64 natural stories were created, half of which contained 2-3 conceptual metaphors relying on the same mapping, while the other half consisted in their literal counterparts. To allow tighter control and manipulation of different properties, instead, 120 isolated sentences were created, half of which contained one metaphorical word, replaced by its literal rendering in the other half. All stimuli were rated for emotional valence, arousal, imageability, metaphoricity, and pairs of metaphorical and literal stimuli for their similarity in meaning. A measure of complexity was determined and computed. Stories were also rated for naturalness and understandability, and sentences for familiarity. Differences between metaphorical and literal stimuli and relationships between affective and psycholinguistic variables were explored and are discussed in light of extant empirical research. In a nutshell, metaphorical stimuli are rated as higher in emotional arousal and are easier to imagine than their literal counterparts, thus confirming a role of metaphor in evoking emotion, and in activating sensorimotor representations. Affective variables show the typical U-shaped relationship consistently found in word databases, whereby increasingly positive and negative valence is associated with higher arousal. Finally, interesting differences between stories and sentences were observed.

Keywords: conceptual metaphor, story, emotion, imageability, figurative language norms

Introduction

Conceptual metaphors help us understand abstract concepts through the use of more concrete terms, for example, in the expressions *He's feeling up; He's down; Being on top of the clouds; Being in a depression,* happiness or good mood (abstract or *target domain*) is conceptualised as upper vertical position, while sadness or bad mood as down; here, vertical position represents the concrete *vehicle* or *source domain,* whose properties are ascribed to the target (Gibbs, 2011; Lakoff & Johnson, 1980). Neurophysiological research has shown that, indeed, sensorimotor representations are activated during comprehension of verbal metaphors linked to underlying conceptual metaphors, i.e., the brain cortices associated with perception in one of the five senses or with motor commands are recruited while readers understand expressions such as *She looked at him sweetly; He grasped the idea; She had a rough day; That idea stinks* (e.g., Boulenger, Hauk, & Pulvermueller, 2009; Cacciari et al., 2011; Citron & Goldberg, 2014; Desai, Conant, Binder, Park, & Seidenberg, 2013; Lacey, Stilla, & Sathian, 2012; Pomp et al., 2018). In addition to evoking bodily representations, metaphors are known to elicit richer semantic representations or multiple meanings, e.g., *sweet girl* may imply kind, nice, pretty, cute; in other words, a metaphor's meaning is not as strictly defined as the meaning of its literal rendering, i.e., *kind girl* (Gibbs & Colston, 2012; Prandi, 2010). The multiple meanings mentioned above are referred to as *emergent properties* by Relevance Theory accounts, i.e., properties of the vehicle that are not related to their literal meaning e.g., sugar, delicious, but to their superordinate category of sweet taste as kindness, e.g., pleasant (Glucksberg, 2008; Wilson & Carston, 2006).

Furthermore, in recent years neurophysiological research has provided empirical evidence for a special role of metaphors in engaging readers or listeners at

the emotional level, by showing enhanced brain activation of structures associated with the processing of evolutionary relevant or emotionally salient stimuli, i.e., the left amygdala in response to metaphorical formulations than literal ones (Bohrn, Altmann, & Jacobs, 2012; Citron & Goldberg, 2014; Forgács et al., 2012), and enhanced heart rate responses to metaphorical translations of English metaphors to Spanish than literal translations (Rojo, Ramos, & Valenzuela, 2014). This evidence confirms pioneering behavioural research showing more productive use of metaphors when describing one's own feelings during autobiographical events compared to when describing the event itself (Fainsilber & Ortony, 1987; Ortony & Fainsilber, 1987), and is further supported by more recent behavioural research showing enhanced empathy and theory of mind in response to stories containing metaphorical language (Horton, 2007, 2013).

More broadly, research on the effects of emotional content on language processing has flourished in the last decade. The two dimensions of *emotional valence* – the extent to which a stimulus is positive (flower) or negative (misfortune) – and *emotional arousal* – the degree of physiological activation in response to a stimulus, from not at all (pacifier) to extremely exciting/agitating (rollercoaster, bomb) - affect single word processing at early and late processing stages, and activate the emotion neural network beyond the language network (e.g., Citron, 2012; Hamann & Mao, 2002; Herbert et al., 2009; Kuperman, Estes, Brysbaert, & Warriner, 2014; Vinson, Ponari, & Vigliocco, 2014). Furthermore, emotional content affects semantic (and syntactic) processing of sentences (Delaney-Busch & Kuperberg, 2013; Diaz-Lago, Fraga, & Acuna-Farina, 2015; Lai, Willems, & Hagoort, 2015) and of longer texts (Ferstl, Rinck, & von Cramon, 2005; Hsu, Jacobs, Citron, & Conrad, 2015).

The enhanced emotional engagement in response to metaphorical formulations, e.g., *That was a <u>bitter</u> breakup,* may be due to the fact that readers or listeners evaluate such expressions as higher in degree of emotional content than their literal counterparts, i.e., *That was a <u>bad</u> breakup* (Citron & Goldberg, 2014). In fact, taste words may be particularly emotionally charged, similarly to smell words, and unlike other sensory domains (Winter, 2016). In their neuroimaging reading experiment, Citron and Goldberg (2014) employed a range of conventional (i.e., commonly used) taste metaphors embedded in short sentences, which were previously extensively rated for a range of properties: taste metaphors were rated as significantly higher in taste-reference and metaphoricity (to what extent is an expression metaphorical) than their literal counterparts, equal in emotional valence, arousal and imageability (how easy it is to imagine), highly similar in meaning, and slightly less familiar (how often does one come across a certain expression). Hence, the enhanced activation of the left amygdala during silent reading in the scanner, suggestive of stronger engagement in response to metaphors, cannot be due to an a-priori difference in degree of emotional content between the two types of stimuli, which was equal; in other words, equal affective content is perceived as more intense or stronger if formulated metaphorically (Citron & Goldberg, 2014; Forgács et al., 2012). Given that research on word recognition has recently shown a unique effect of emotional valence and arousal on lexical decision and naming, beyond more traditional variables such as length, frequency, imageability, etc. (e.g., Kousta, Vinson, & Vigliocco, 2009; Kuperman et al., 2014; Rodríguez-Ferreiro & Davies, 2019), extensive rating of affective beyond psycholinguistic properties is necessary in studies aimed to explore relationships between (figurative) language and emotion, followed by a thorough manipulation or matching of such properties between experimental conditions.

The present work provides the first dataset of conventional conceptual metaphors encompassing a wide range of concrete domains, along with ratings of affective beyond psycholinguistic properties. Our metaphors are embedded in natural short stories as well as in simple sentences, and highly similar literal versions are provided and extensively rated too. This dataset allows the investigation of metaphorical language processing with no restriction to a specific sensory domain as well as the extension of such investigation to more natural reading processes, i.e., the reading of short stories. Finally, our stimuli have been constructed so that mention of emotional states (angry, sad, happy) or their metaphorical renderings (pissed off, down, up) was avoided as much as possible; this is because any empirical evidence of an emotional advantage of metaphorical over literal language should be attributable to its figurativeness/metaphoricity rather than to the presence of highly emotive content. As a result, no explicit mention of emotions felt by the protagonists is contained in our sentences and stories. Nevertheless, these inevitably describe events that range in their affective content, i.e., positive, negative or affectively neutral.

Other metaphor datasets have not specifically focused on affective properties. For example, Cardillo, Schmidt, Kranjec and Chatterjee (2010; Cardillo, Watson, & Chatterjee, 2017) created English novel (non-conventional) metaphorical sentences involving the auditory and motor domains in which one word was used metaphorically, e.g., *His ugly car is a giggle*, and literal sentences with a different meaning but the same word used literally, i.e., *The child's answer was a giggle.* The authors collected ratings for metaphoricity, imageability, concreteness (to what extent can an expression be experienced with one of the 5 senses), naturalness, familiarity, and ease of interpretation (of an expression's meaning) for 560 stimuli (Cardillo et al., 2010) and 240 additional stimuli (Cardillo et al., 2017), half of which were

metaphorical; they also measured reaction times during a valence judgment task whereby participants were asked to categorise each sentence as either positive in valence or not - hence, either negative or neutral. Therefore, a valence index is available, which distinguishes positive sentences from others, but no detailed measures of emotional valence and arousal on Likert scales were collected. In both studies, novel metaphorical sentences were rated as more metaphorical and more difficult to interpret than literal sentences, and as less imageable, natural and familiar; in Cardillo et al. (2017), metaphorical sentences were also rated as less concrete.

Bambini, Resta and Grimaldi (2014) instead selected a range of metaphors embedded in short sentences from Italian literary texts; thus, these stimuli also consist in essentially novel metaphors. The authors collected ratings of concreteness, familiarity, readability, cloze probability (how predictable is a metaphor given a sentential context), difficulty of comprehension, and meaningfulness (how much does the expression make sense) for 115 sentences, and the same ratings once again for a subset of 65 metaphors embedded in their original texts, hence within a larger context. Familiarity correlated positively with meaningfulness and concreteness, and negatively with difficulty – similarly to the positive correlations between imageability, naturalness, familiarity, and ease of interpretation found by Cardillo et al. (2010; Cardillo et al., 2017). Presenting a subset of literary metaphors in their original context decreased their rated concreteness, difficulty and familiarity while increasing meaningfulness, which was no longer correlated with familiarity, and weakened the pattern of correlations between variables overall; the authors suggested that literary context enhances a broader interpretative activity that renders metaphors more open to different interpretations rather than more familiar, and this is in line

with neurophysiological findings from the same group (Bambini, Bertini, Schaeken, Stella, & Di Russo, 2016; Bambini, Canal, Resta, & Grimaldi, 2019).

The correlation patterns above are in line with pioneering work by Katz, Paivio, Marschark and Clark (1988) on 204 literary and 260 non-literary metaphors, which both showed positive correlations between several measures including comprehensibility, ease of interpretation, familiarity, different measures of imageability, semantic relatedness (how similar in meaning are vehicle and target, e.g., _aching desire_ vs. _aching money_; example from Liu, 2018), and aptness (how well does the vehicle's metaphorical meaning describe an important feature of the target, e.g., _Memory is a warehouse_ may describe the target "memory" better than _A fisherman is a spider_ describes "fisherman"; Keysar & Glucksberg, 1990).

In addition, stemming from a large debate on the distinction between aptness and conventionality (common use or strength of association between a metaphor vehicle and its figurative meaning, e.g., _sunny disposition_ is conventional while _cloudy disposition_ is novel), and to what extent these variables can be reliably measured by asking naïve participants to rate them (Jones & Estes, 2006; Pierce & Chiappe, 2009; Roncero, de Almeida, Martin, & de Caro, 2016; Thibodeau & Durgin, 2011), Thibodeau, Sikos and Durgin (2017) conducted two experiments and re-analysed data from four extant databases of metaphors (Cardillo et al., 2010; Cardillo et al., 2017; Katz et al., 1988; Roncero et al., 2016); they showed that the variability of different properties characterising metaphors can be essentially reduced to two main components: 1) _processing fluency_ loads positively on variables such as comprehensibility, ease of interpretation, familiarity, aptness, conventionality, imageability, negatively on surprisingness (how surprising is the metaphoric word as it is currently used), and reflects how easily metaphors are understood; 2) while

figurativeness loads positively on degree of metaphoricity, surprisingness, imageability, and negatively on familiarity or conventionality. Finally, in their experiments Thibodeau et al. (2017) were able to show that a context that matches the metaphorical mapping used in the target metaphor facilitates comprehension and affects processing fluency, whereas figurativeness is a more stable component that is not affected by context; a finding that is in line with previous work (Gibbs & Gerrig, 1989; Ortony, Schallert, Reynolds, & Antos, 1978; P. Thibodeau & Durgin, 2008).

To our knowledge, the only dataset of figurative expressions that provides ratings for affective variables is PANIG (Citron et al., 2016a), which consists of 619 German idiomatic expressions rated on emotional valence, arousal, concreteness, familiarity, idiomaticity (to what extent is an expression idiomatic; similar to figurativeness or metaphoricity), semantic transparency (to what extent is the relationship between an idiom's meaning and the literal meaning of its constituent words intuitive/transparent), and confidence in meaning knowledge; in addition, idiom knowledge was measured through collection of full definitions from participants. Idioms differ from metaphors in that they represent highly conventionalised expressions, with a clearly defined, arbitrary and learned meaning (Cacciari, 2014; Glucksberg, 2001). Nevertheless, similarly to metaphors, idioms have also been shown to evoke stronger emotional responses in readers at the neural level (Citron, Cacciari, Funcke, Hsu, & Jacobs, 2019); they are preferred over literal expressions when formulating complaints, especially in the presence of a non-empathic interlocutor, and in topic transitions (Drew & Holt, 1988, 1998).

The PANIG database showed replication and generalisation to idioms of the well-known U-shaped and negative linear relationships between emotional valence and arousal previously reported for single words and pictures (e.g., Bradley & Lang,

1999; Lang, Bradley, & Cuthbert, 1999; Montefinese, Ambrosini, Fairfield, & Mammarella, 2013; Võ et al., 2009); in other words, increasingly positive and negative idioms are higher in arousal compared to emotionally neutral idioms, which are very low in arousal. In addition, negative idioms are higher in arousal than positive ones (Citron et al., 2016a). Furthermore, emotional arousal correlated positively with idiomaticity and concreteness, in line with the idea that the more idiomatic (or figurative) an expression, the more emotionally arousing it is perceived; in addition, idioms with more concrete meanings possibly lead to the activation of sensorimotor (bodily) representations, and are in turn associated with higher levels of arousal. Familiarity showed a positive correlation with emotional arousal, i.e., the more familiar an idiom the more arousing, and with emotional valence, i.e., the more familiar an idiom the more positively valenced. This finding may partly be due to the overall lower number of positive idioms, which may be used more frequently compared to the large variety of negative idioms to choose from when expressing something unpleasant. Finally, idiomaticity showed a negative correlation with concreteness, i.e., the more idiomatic the more abstract, and knowledge showed a positive correlation with familiarity, i.e., the better known the more familiar.

The present study

Thirty-two short stories containing about 2-3 conventional German metaphors that relied on the same mapping were created, along with their 32 literal counterparts. In addition, 60 sentences containing one conventional metaphor each and their 60 literal renderings, which most often differed for only one word, were created. All 64 stories and 120 sentences were rated for:

- emotional valence and arousal, imageability, metaphoricity;

- pairs of metaphorical and literal stories and sentences were rated for how similar their meanings are;

- a measure of objective complexity was determined and calculated.

In addition, the stories were rated for:

- naturalness and understandability.

Instead, the sentences were rated for:

- familiarity.

Imageability was preferred over concreteness because the former seems to better represent emotionally-laden verbal stimuli (excitement), which tend to be rated as abstract (excitement cannot be directly experienced with one of the 5 senses) but nevertheless quite imageable compared to emotionally neutral abstract stimuli (thought), which are rated as abstract *and* difficult to imagine (Altarriba & Bauer, 2004). Furthermore, imageability tends to have a more continuous distribution than concreteness, which is more bimodal (Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011). The two variables share a large portion of their variance, and thus we deemed unnecessary to rate both.

Based on the literature reviewed above, and the link of conceptual metaphors with sensorimotor representations, we expect metaphorical versions of our stimuli to be rated as higher in metaphoricity, imageability, and emotional arousal than their literal counterparts. Despite all our metaphors being conventional, they vary in familiarity. Therefore, our metaphorical sentences may be rated as less familiar overall than their literal counterparts, while our metaphorical stories may be rated as less natural and less understandable than their literal versions, given the relatively less common use of metaphors relying on a same mapping within a short paragraph or discourse.

For all stories and all sentences, we predict a U-shaped and a linear negative relationship between valence and arousal whereby increasingly positive and negative stories and sentences will be rated as increasingly more arousing than neutral ones, and negative stories and sentences will be rated as more arousing than positive ones; if these predictions are met, current findings from extant databases of single words in several languages and from the only idiom database that includes affective variables would be generalised to metaphorical and literal natural stories and sentences. Based on previous databases and the literature, we have no reason to predict differences between literal and metaphorical versions of stories and sentences in the relationships between affective variables. In addition, for both stories and sentences we predict positive correlations between degree of metaphoricity, imageability, and emotional arousal, in line with empirical findings showing that increasingly metaphorical verbal stimuli are perceived as increasingly emotionally arousing and easier to imagine. Similarly, direct contrasts between metaphorical and literal versions of stories and sentences will show higher arousal level and imageability for the metaphorical versions. Note that the latter prediction is in contrast with findings by Cardillo et al. (2010; 2017), whose literal sentences were rated as easier to imagine than the metaphorical ones; this is because the authors employed literal sentences with different words and a different meaning than the metaphorical sentences, e.g., *The birds were yellow canaries* versus *The sweethearts were two canaries*; the only common word between the two was concrete and used literally in the literal sentences, thus being necessarily easier to imagine than the same word used metaphorically. Our study remains exploratory with respect to possible correlations between other non-affective variables as we used a different list of variables compared to extant metaphor databases. Finally, in terms of differences in patterns of

correlations between the metaphorical and the literal subsets, we expect the correlations predicted above to be stronger or hold within metaphorical stimuli compared to literal stimuli; this is because literal renderings with highly similar meaning do not entail a link to sensorimotor representations as metaphors do, they have been shown to evoke less of an affective response in the brain than the metaphorical formulations (Citron, Güsten, Michaelis, & Goldberg, 2016b), and their use in discourse is not specifically associated with expression of emotion, theory of mind or perceived intimacy (Fainsilber & Ortony, 1987; Horton, 2007).

Method

This study was approved by the Ethics Committee of the Cluster of Excellence "Languages of Emotion", Freie Universität Berlin, Germany, and its conduction is in line with the guidelines of the American Psychological Association and the Declaration of Helsinki. The full database is available as Supplementary Material and can also be freely accessed at: LINK TO BE PROVIDED.

Participants

Four hundred and fifteen native speakers of German took part in the rating study, including 243 women (59%) and 172 men (41%). Their age ranged from 18 to 78 years old (*mean = 35, SD* = 14). Participants were mostly university students (49%) or workers (44%), some were retired (4%), unemployed (2%), or unknown (1%). Participants gave their consent to participate through the online survey link and they were paid 5€ for each completed survey. While for stories each participant would rate half of the sample - 32 out of 64 stories - for one property, for sentences each participant would rate the full set of 120 sentences. Similarity in meaning was

similarly rated but stimuli were presented in pairs. Table 1 shows mean age and gender proportion broken down by type of stimulus, property and stimulus samples. Although female participants were more numerous than male participants, no significant differences in gender proportion were found either between stories and sentences, $X^2(1) = 2.45$, *ns*, or between properties in stories in sample 1, $X^2(6) = 1.39$, *ns*, and sample 2, $X^2(6) = 1.30$, *ns*, or between properties in sentences, $X^2(5) = 2.85$, *ns*. No significant differences in age were found between stories and sentences, $t(413) = 0.44$, *ns,* or between properties and between samples in stories, $F(6, 289 = 1.33, ns$; $F(1, 289) = 0.08$, *ns,* except a marginally significant interaction between property and sample, $F(6, 289) = 2.05$, $p = 0.06$; age did not differ significantly between properties in sentences either $F(5, 120) = 0.26$, *ns.* On average, each story or sentence was rated for a single property by 20-23 participants, in a between-participants design, given that different properties of the same stimuli were rated by different participants samples. Each participant could complete as many surveys at they liked but couldn't complete the same survey twice.

Table 1. Descriptive statistics of age and gender proportion for each distinct participant sample.

Variables	N stimuli rated	Participant sample	Mean age (SD)	N (women, men)
Stories				
Emotional valence	32	sample 1	31.60 (13.76)	20 (11, 9)
	32	sample 2	30.86 (12.14)	21 (12, 9)
Emotional arousal	32	sample 1	36.5 (15.75)	20 (11, 9)
	32	sample 2	30.62 (12.42)	21 (14, 7)
Imageability	32	sample 1	31.25 (12.32)	20 (11, 9)
	32	sample 2	41.33 (16.85)	21 (10, 11)
Metaphoricity	32	sample 1	39.14 (16.48)	21 (12, 9)
	32	sample 2	36.50 (15.75)	20 (11, 9)
Naturalness	32	sample 1	30.95 (12.65)	20 (13, 7)
	32	sample 2	31.60 (13.76)	20 (11, 9)
Understandability	32	sample 1	32.48 (12.79)	21 (11, 10)
	32	sample 2	39.14 (16.48)	21 (12, 9)
Similarity in meaning	16 pairs	sample 3	39.55 (15.12)	20 (10, 10)
	16 pairs	sample 4	32.22 (12.27)	23 (13, 10)
Sentences				
Emotional valence	120	only sample	37.71 (12.82)	21 (11, 10)
Emotional arousal	120	only sample	34.80 (10.95)	20 (13, 7)
Imageability	120	only sample	35.33 (13.07)	21 (15, 6)
Metaphoricity	120	only sample	33.43 (11.74)	21 (12, 9)
Familiarity	120	only sample	34.76 (13.06)	21 (15, 6)
Similarity in meaning	60 pairs	only sample	35.18 (13.49)	22 (15, 7)

Materials

Sixty metaphorical sentences were created by the experimenters by using conventional German metaphors, e.g., *light life*, which is based on the mapping DIFFICULTY is WEIGHT, and embedding them within short sentences, i.e., *Ach, was für ein leichtes Leben hatten wir während der Schulferien!* (*Oh, what a light life we had during the school holidays!*). The metaphors came to mind spontaneously to the experimenter who is a native speaker of German, and they are all conventional

15

expressions, some used more often than others but all very well known to German native speakers. The conceptual mappings were then deducted (all listed in the dataset). Some of them may sound familiar in English and others not at all, but they are all existing in German. Their 60 literal counterparts were created by replacing the metaphorical word, e.g., "light", with its literal counterpart, i.e., "easy". For most sentences the replacement consisted in one word, however for some sentences more than one word had to be changed, or a different preposition was associated with the literal verb, or a change of case marking occurred (see the dataset for details). A similar procedure was used to create the 32 metaphorical stories, which contained between 2 and 3 conceptual metaphors that relied on the same mapping (see Table 2). Stories and sentences contained different metaphorical mappings. To make the 32 literal stories flow naturally, however, the literal versions do not always consist in a one-word replacement for each metaphor; thus, a less tight control over structure and single words used is possible for stories than sentences. As already explained in the introduction, explicit mention of emotional reactions (happy, sad, or their metaphorical renderings) was avoided. Nevertheless, our stimuli varied in emotional content. In all our stories and sentences, metaphorical and literal words appeared at different points in the story/sentence and could belong do a range of different grammatical classes (adjective, noun, verb, etc.). This variability contributes to the ecological validity of our stimuli, allowing investigation of natural reading processes, however, this variability may be counterproductive for experiments employing time-sensitive measures such as EEG, especially in the case of our short stories.

Table 2. Example of metaphorical and literal story with three metaphors, all based on the mapping SOLVING MATHS PROBLEMS is DIGESTING, which is a common conceptual mapping in German.

German metaphorical story	German literal story
Lisa saß im Physikunterricht und verdaute noch den Stoff aus der letzten Stunde, als der Lehrer eine Textaufgabe zum Zähneausbeißen ankündigte. Lisa stöhnte. Sie konnte nichts mehr aufnehmen, war vollgestopft mit anderen schulischen Inhalten. Aber natürlich sah sie sich die Aufgabe trotzdem an.	Clara saß im Physikunterricht und verarbeitete noch den Inhalt der letzten Stunde, als der Lehrer eine sehr schwierige Textaufgabe ankündigte. Clara stöhnte. Sie konnte sich nicht mehr konzentrieren, die Themen der anderen Fächer lenkten sie von der jetzigen Aufgabe ab. Aber natürlich sah sie sich die Aufgabe trotzdem an.
English translation, metaphorical story	English translation, literal story
Lisa was sitting in her physics class and was still digesting the stuff from the lesson before when her teacher announced a task to bite your teeth out on. Lisa moaned. She couldn't take anything in anymore, she was stuffed with school contents. But, of course, she still had a look at the task.	Lisa was sitting in her physics class and was still processing the stuff from the lesson before when her teacher announced a really difficult task. Lisa moaned. She couldn't think anymore, the topics of the other subjects distracted her from the current task. But, of course, she still looked at the task.

Procedure

Online surveys were created using SurveyMonkey. Instructions for each variable to be rated were first provided, with a description of the 7-point Likert scale used for each variable, and examples of stimuli that were high or low on the scale. Full instructions and their translation are provided in the Supplementary Materials, and can be also freely accessed at: LINK TO BE PROVIDED. Here, a short description of each scale is provided. For all stimuli, emotional valence was defined as the extent to which an event (or series of events) is positive or negative; it goes from -3 (very negative) through 0 (neutral) to +3 (very positive); emotional arousal was defined as the extent to which an event (or series of events) is emotionally stimulating, from 1 (not at all intense) to 7 (very intense); imageability was defined as

the ease with which a story/sentence or its elements can evoke images of the senses/be imagined, from 1 (not at all imageable) to 7 (very imageable); metaphoricity was defined as the degree of figurativeness (Bildersprachlichkeit in German), from 1 (literal) to 7 (very metaphorical); similarity in meaning was defined as how similar the meaning of the literal and metaphorical story/sentence of each pair is, and ranged from 1 (not at all similar) to 7 (very similar). For the latter, participants were instructed that they would read story or sentence pairs, each including a metaphorical and a literal story or sentence, and asked to rate how similar in meaning they found them. Only for stories, naturalness was defined as how normal and daily a story or its parts sound and goes from 1 (very unnatural) to 7 (very natural), while understandability was defined as how well one can grasp what goes on in a story, from 1 (very difficult to understand) to 7 (very easy to understand). Only for sentences, familiarity was defined as how common a sentence is, i.e., how often one hears or reads a sentence, from 1 (not at all common/used) to 7 (very common/very often used). Furthermore, a measure of objective complexity was determined and computed for each story and sentence by adding the number of subordinate clauses, relative clauses, passive forms, compound nouns, new referents, adverbs and adverbial phrases, conjunctive forms, analytically-formed tenses or infinitive constructions, and marked or low-frequency sentence structures. Finally, length in words was calculated for each stimulus.

After the instructions, a list of stories or sentences was provided. One survey consisted in the rating of a single variable for half the stories (32); each half would contain a mixture of metaphorical and literal stories, although two versions of the same story (metaphorical, literal) would appear in different halves. Alternatively, 16 story pairs were presented after the meaning similarity instructions. For sentences, the

full set of 120 was presented after the rating instruction for one variable, or 60 sentence pairs were presented after meaning similarity. This subdivision ensured that one survey would take between 1 and 1,5 hours to complete, to avoid fatigue. In each survey, the stimuli were presented in randomised order, and metaphorical and literal stimuli were intermixed. In surveys with sentences, participants would rate metaphorical and literal versions of each sentence pair. To avoid repetition effects as much as possible, sentences were divided in two blocks and within each block only one version of each sentence pair would appear. Therefore, different versions of the same sentence pair would be very distant from one another. Surveys with sentences were organised in such a way that Participants received a URL for each survey; they could complete it at their own pace and take short breaks in between.

Data analysis

For stories we calculated mean, standard deviation, median, minimum and maximum scores of each story for ten variables: emotional arousal, emotional valence, imageability, metaphoricity, naturalness, understandability, complexity, length in words, similarity in meaning as well as an additional variable of emotional valence squared that was calculated by squaring emotional valence. The same scores were created for sentences with the variables emotional arousal, emotional valence, imageability, metaphoricity, familiarity, complexity, length in words, similarity in meaning, and emotional valence squared. These were calculated separately for metaphorical and literal versions of the stories and sentences apart from similarity in meaning, for which there was a single set of scores. When kurtosis and skewness were examined, all variables were deemed normally distributed. We then ran t-tests

on these variables for stories and sentences separately to assess differences between the metaphorical and literal versions of each.

To examine linear and quadratic relationships between emotional arousal and emotional valence, a quadratic stepwise forward regression analysis was conducted with emotional arousal as dependent variable, separately for stories and sentences. In a first step, we controlled for the effects of all non-affective variables (except similarity in meaning) and in a second step we entered emotional valence and emotional valence squared as predictors. To explore linear relationships between affective and non-affective variables as well as between non-affective variables only, we conducted partial correlations separately for stories and sentences. When significant, partial correlations up to ± .1 are considered "small", between ± .1 and ± .3 are "moderate" and above ± .3 are "large". The same regressions and partial correlations were then conducted for metaphorical and literal subsets of stories and sentences.

Reliability analysis. An analysis of internal consistency was conducted using Cronbach's alpha. The ratings from each single participant were used as variables and the single stories or sentences as cases. One alpha value was computed for each rated property and participant sample, separately for stories and sentences.

Results

Descriptive statistics for all variables can be found in Table 3.

Table 3. Descriptive statistics for all rated and calculated variables for metaphorical and literal versions of stories and sentences. Similarity in meaning was rated for each

pair of stories or sentences. We report the statistics under "Metaphorical" but they actually refer to pairs of metaphorical and literal versions.

	Stories									
	Metaphorical					Literal				
Variables	Mean	SD	Median	Minimum	Maximum	Mean	SD	Median	Minimum	Maximum
Emotional arousal	4,55	0,87	4,61	2,67	6,15	3,82	0,75	3,68	2,25	5,48
Emotional valence	0,37	1,37	0,43	-2,7	2,58	0,40	1,23	0,25	-2,38	2,37
Valence squared	1,94	1,99	1,37	0	7,29	1,63	1,77	1,08	0	5,66
Imageability	4,58	0,80	4,5	3	6,15	3,66	0,73	3,58	2,42	5,05
Metaphoricity	4,66	1,09	4,81	2,32	6,48	2,23	0,51	2,19	1,5	3,33
Naturalness	4,38	0,48	4,4	2,95	5,37	4,92	0,45	4,95	3,7	6
Understandability	3,20	0,69	3,29	1,95	4,9	3,15	0,66	3,17	1,95	4,52
Complexity	19,72	5,00	19,5	9	29	19,56	5,02	20	10	29
Length in words	58,59	16,94	54,5	33	102	57,38	16,27	20	34	93
Similarity in meaning	5,52	0,64	5,51	3,8	6,78	-	-	-	-	-
	Sentences									
	Metaphorical					Literal				
Variables	Mean	SD	Median	Minimum	Maximum	Mean	SD	Median	Minimum	Maximum
Emotional arousal	4,01	0,85	4,23	1,95	5,6	3,74	0,92	3,88	1,85	5,55
Emotional valence	-0,06	1,22	-0,19	-2,43	2,38	0,06	1,14	-0,07	-2,1	2
Valence squared	1,48	1,31	1,16	0	5,9	1,28	1,12	1,05	0	4,41
Imageability	4,38	0,95	4,33	2,57	6,55	4,04	0,96	4	2,43	6,2
Familiarity	4,70	0,65	4,86	3,05	5,76	4,92	0,51	5	3,67	5,81
Metaphoricity	4,02	1,03	3,98	1,86	5,71	1,99	0,63	1,84	1,29	3,76
Complexity	3,28	0,96	3	1	5	3,28	0,96	3	1	5
Length in words	8,37	1,60	8	5	11	8,32	1,67	8	5	12
Similarity in meaning	5,60	0,72	5,7	3,41	6,77	-	-	-	-	-

Differences between metaphorical and literal versions

Metaphorical stories were rated as significantly higher in emotional arousal, $t(62) = 3.62$, $p < .001$, imageability, $t(62) = 4.76$, $p < .001$, and metaphoricity, $t(62) = 11.38$, $p < .001$, than literal stories, and as significantly less natural, $t(62) = -4.60$, $p < .001$. No significant differences in emotional valence, understandability, complexity, or length in words were found, all $ts(62) < 0.65$, *ns*.

Metaphorical sentences were rated as significantly higher in metaphoricity, than literal sentences, $t(118) = 13.02$, $p < .001$, and only marginally higher in emotional arousal, $t(118) = 1.69$, $p = .94$, and imageability, $t(118) = 1.95$, $p = .054$; metaphorical sentences were also rated as significantly less familiar $t(118) = -2.05$, $p < .05$ than literal ones. No significant differences in emotional valence, complexity, or length in words were found, all $ts(118) < 0.89$, *ns*.

Relationships between affective variables

A graphical representation of the relationship between emotional valence and arousal for all stories and all sentences, with metaphorical and literal versions differently marked, can be found in Figure 1.

Stories. In the quadratic regression predicting emotional arousal from valence ratings for all stories, the first model, which included only non-affective variables, had imageability as unique significant predictor and accounted for 34% of the variance ($r^2 = .34$, $r = .58$; $F(1, 62) = 31.61$, $p < .001$); the second model, which included emotional valence and valence squared, had only the latter as significant predictor and accounted for an additional 7% of variance ($r^2 = .41$, $r = .64$, $F(2, 61) = 21.12$, $p < .01$). The regression line is: Arousal = 1.85 + 0.51 x Imageability + 0.13 x Emotional valence squared.

For the subset of metaphorical stories, the first model included only imageability as significant predictor and accounted for 47% of the variance ($r^2 = .47$, $r = .68$; $F(1, 30) = 26.43$, $p < .001$); the second model included only emotional valence squared as significant predictor and accounted for an additional 14% of the variance ($r^2 = .61$, $r = .78$; $F(2, 29) = 22.83$, $p < .01$). Regression line: Arousal = 1.36 + 0.63 x Imageability + 0.17 x Emotional valence squared.

In contrast, the same regression analysis for the subset of literal stories showed no significant predictors.

Sentences. For all sentences, the first regression model included only imageability as a significant predictor amongst all non-affective variables, accounting for 20% of the variance ($r^2 = .20$, $r = .45$; $F(1, 118) = 29.20$, $p < .001$). A second model included emotional valence as significant predictor, accounting for an additional 13% of the variance ($r^2 = .33$, $r = .57$; $F(2, 117) = 28.60$, $p < .001$). Finally,

22

a third model[1] also included valence squared as a significant predictor, accounting for an additional 9% of the variance ($r^2 = .42$, $r = .65$; $F(3, 116) = 28.01$, $p < .001$). Regression line: Emotional arousal = 2.23 + 0.31 x Imageability - 0.28 x Emotional valence + 0.24 x Emotional valence squared.

For the subset of metaphorical sentences, the first model containing all non-affective variables also contained only imageability as a significant predictor, accounting for 28% of the variance ($r^2 = .28$, $r = .53$; $F(1, 58) = 27.73$, $p < .001$). A second model had emotional valence squared as a significant predictor, accounting for an additional 11% of the variance ($r^2 = .39$, $r = .62$; $F(2, 57) = 18.12$, $p < .01$), while a third model also included emotional valence, accounting for an additional 9% of the variance ($r^2 = .48$, $r = .69$; $F(3, 56) = 17.14$, $p < .01$. Regression line: Emotional arousal = 2.07 + 0.37 x Imageability + 0.22 x Emotional valence squared – 0.21 x Emotional valence.

Similarly, for the subset of literal sentences, when all non-affective variables were entered into the first step of the regression, only imageability was a significant predictor, accounting for 12% of the variance ($r^2 = .12$, $r = .34$; $F(1, 58) = 7.56$, $p < .01$). In a second model with affective variables added, emotional valence was a significant predictor, explaining an additional 17% of variance ($r^2 = .29$, $r = .54$; $F(2, 57) = 11.62$, $p < .001$). Finally, a third model also included emotional valence squared, accounting for an additional 9% of variance ($r^2 = .38$, $r = .61$; $F(3, 56) = 11.29$, $p < .01$). Regression line: Emotional arousal = 2.43 + 0.25 x Imageability – 0.37 x Emotional valence + 0.26 x Emotional valence squared.

---

[1] In the stepwise regression, every time a significant predictor is included, the model is retested for its additional unique contribution to the variance. Therefore we have three models despite having entered all our variables using only two steps.
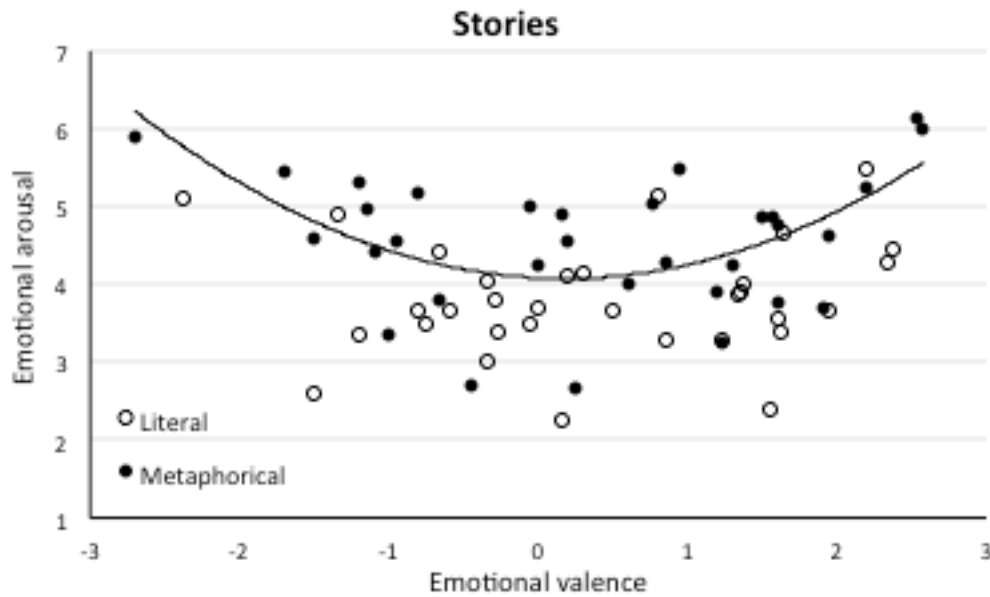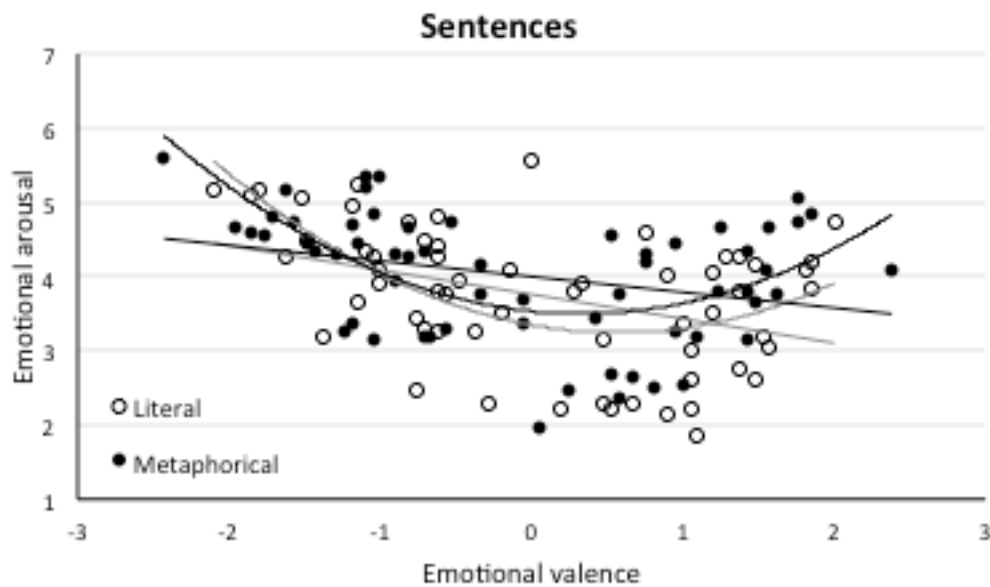
Figure 1a.

Figure 1b.

Figure 1. Scatter plot showing the relationship between emotional valence and arousal for (a) stories, where a quadratic relationship is clearly observable. The quadratic trend refers to the metaphorical stories only since the quadratic trend was significant only in this subset as well as in all stories; and for (b) sentences, where a negative linear trend is observable beyond a quadratic relationship. The black trends refer to the metaphorical and the grey trends to the literal sentences.

Partial correlations between affective and non-affective variables

All correlation terms and significance levels are reported in Table 4.

Stories. A large positive partial correlation between emotional arousal and imageability was found; this was the case for all stories analysed together as well as metaphorical and literal stories analysed separately. A large negative partial correlation between emotional valence and understandability was also found, i.e., the more negatively valenced a story the easier to understand; this was found in stories analysed overall and literal stories analysed separately but not for metaphorical stories only. No other significant partial correlations were found.

Sentences. A large positive partial correlation between emotional arousal and imageability was found for both sentences overall and metaphorical sentences, whilst a moderate positive partial correlation was found for literal sentences. No other significant partial correlations were found.

Table 4. Partial correlations (Person's r) between affective and non-affective variables in stories and sentences.

| | Stories | | | | | |
| | Emotional valence | | | Emotional arousal | | |
	All stories	Metaphorical	Literal	All stories	Metaphorical	Literal
Imageability	-.02	.13	-.20	.30*	.45*	.40*
Metaphoricity	-.18	-.22	.07	.12	.13	-.38
Naturalness	-.24	-.30	-.19	-.001	-.16	.05
Understandability	-.33*	-.31	-.43*	.13	.04	.36
Complexity	.11	.07	.27	-.06	.14	-.27
Length in words	.21	.16	.14	-.05	-.24	.60
	Sentences					
	Emotional valence			Emotional arousal		
	All sentences	Metaphorical	Literal	All sentences	Metaphorical	Literal
Imageability	.18	.18	.12	.34***	.41**	.28*
Familiarity	.02	-.11	.15	.08	.06	.18
Metaphoricity	-.08	-.13	.06	.10	.17	.02
Complexity	.03	-.02	.18	.04	-.12	.13
Length in words	-.04	-.01	.08	.09	.21	-.002

*$p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Partial correlations between non-affective variables

All correlation terms and significance levels are reported in Table 5 for stories and Table 6 for sentences.

Stories. A large positive partial correlation between imageability and metaphoricity was found for all stories as well as for metaphorical and literal subsets of stories (see Table 5). A large negative partial correlation between imageability and understandability was found for all stories as well as for metaphorical and literal subsets, i.e., stories that were easier to imagine were also rated as less easy to understand. A large positive partial correlation between imageability and complexity was also found, but this was the case only for stories analysed as a whole and for the literal subset; no significant correlation was found between the two variables in the metaphorical subset of stories. In the analysis of all stories, a large negative partial correlation was found between metaphoricity and naturalness suggesting that the more metaphorical a story was rated the less natural it was perceived; this large negative partial correlation was also found in the metaphorical subset but not in the literal one. In addition, two correlations were only found when all stories were analysed together and did not appear when either metaphorical or literal stories were analysed separately: a large positive partial correlation between imageability and naturalness, suggesting the easier to imagine the more natural a story was rated, and a moderate negative partial correlation between metaphoricity and complexity, i.e., increasingly metaphorical stories tended to be less complex. Finally, length in words correlated positively with both complexity and understandability, i.e., the longer a story the more complex but also the easier to understand.

Table 5. Partial correlations (Person's r) between non-affective variables in stories.

	All stories					
	Imageability	Metaphoricity	Naturalness	Complexity	Understandability	Length in words
Imageability	1					
Metaphoricity	.69***	1				
Naturalness	.35**	-.65***	1			
Complexity	.36**	-.29*	-.25	1		
Understandability	-.51***	.16	-.21	.12	1	
Length in words	-.04	.21	.36**	.65***	.41**	1
	Metaphorical stories					
	Imageability	Metaphoricity	Naturalness	Complexity	Understandability	Length in words
Imageability	1					
Metaphoricity	.53**	1				
Naturalness	.35	-.53**	1			
Complexity	.20	-.20	-.17	1		
Understandability	-.49*	.03	-.26	.04	1	
Length in words	.12	.17	.26	.68***	.42*	1
	Literal stories					
	Imageability	Metaphoricity	Naturalness	Complexity	Understandability	Length in words
Imageability	1					
Metaphoricity	.61**	1				
Naturalness	.37	-.34	1			
Complexity	.48*	-.26	-.30	1		
Understandability	-.59**	.34	-.10	.25	1	
Length in words	-.11	.12	.38	.61**	.37	1

*p < .05 ** p < .01 *** p < .001

Sentences. A moderate partial positive correlation between imageability and metaphoricity was found for sentences overall and literal sentences only, but not for metaphorical sentences (see Table 6). A large negative partial correlation between metaphoricity and familiarity was found for all sentences as well as both subsets of metaphorical and literal sentences, suggesting that the more metaphorical the sentences the less familiar they were rated. In addition, metaphorical sentences showed a significant large negative partial correlation between metaphoricity and length in words, and literal sentences showed a significant moderate negative partial correlation between familiarity and complexity, i.e., the more familiar a sentence the less complex. Finally, a large positive partial correlation between complexity and length was found in all analyses of sentences.

Table 6. Partial correlations (Person's r) between non-affective variables in sentences.

	All sentences				
	Imageability	Metaphoricity	Familiarity	Complexity	Length in words
Imageability	1				
Metaphoricity	.23*	1			
Familiarity	-.08	-.33***	1		
Complexity	.08	-.001	-.12	1	
Length in words	.09	-.09	-.09	.37***	1
	Metaphorical sentences				
	Imageability	Metaphoricity	Familiarity	Complexity	Length in words
Imageability	1				
Metaphoricity	.14	1			
Familiarity	-.12	-.33*	1		
Complexity	.14	.16	.001	1	
Length in words	.08	-.3*	-.15	.38**	1
	Literal sentences				
	Imageability	Metaphoricity	Familiarity	Complexity	Length in words
Imageability	1				
Metaphoricity	.36**	1			
Familiarity	.03	-.36**	1		
Complexity	.03	-.06	-.29*	1	
Length in words	-.003	.24	.03	.37**	1

*p < .05 ** p < .01 *** p < .001

Reliability analysis

The results showed high reliability of all variables and within each sample for stories except naturalness, which showed low variability in one sample (see Table 7). Similarly, for sentences all variables showed high reliability.

Table 7. Cronbach's alpha calculated for each rated variable and participant sample.

Variables	Participant sample	Cronbach's α	αs if each participant deleted	N of participants
Stories				
Emotional valence	sample 1	.97	all > .97	20
	sample 2	.97	all > .96	21
Emotional arousal	sample 1	.91	all > .90	20
	sample 2	.86	all > .84	21
Imageability	sample 1	.89	all > .87	19*
	sample 2	.89	all > .87	21
Metaphoricity	sample 1	.98	all > .97	21
	sample 2	.95	all > .95	20
Naturalness	sample 1	.77	all > .73	20
	sample 2	.57	all > .49	20
Understandability	sample 1	.85	all > .82	21
	sample 2	.90	all > .89	21
Similarity in meaning	sample 3	.86	all > .84	20
	sample 4	.85	all > .83	23
Sentences				
Emotional valence	only sample	.95	all > .95	21
Emotional arousal	only sample	.91	all > .90	20
Imageability	only sample	.89	all > .89	21
Metaphoricity	only sample	.96	all > .95	21
Familiarity	only sample	.77	all > .74	21
Similarity in meaning	only sample	.82	all > .79	22

Discussion

The aims of this study were to provide researchers with extensively rated stories and sentences that can be used to investigate questions related to the relationship between (figurative) language and emotion, the processing of (figurative) language, and many more, by controlling and/or manipulating properties that are known to affect language processing; and to further investigate and explore relationships between affective and psycholinguistic properties of (figurative) verbal materials.

To address the first aim, this study provides the first dataset of conventional conceptual metaphors embedded in natural short stories as well as in isolated sentences, along with their literal counterparts, which are highly similar in meaning, almost identical in grammatical structure, and which only differ from the metaphorical stimuli for a few words in stories and only one word in most sentences. Crucially, this is also the first dataset of metaphors that includes ratings for affective properties – emotional valence and arousal – beyond a range of other psycholinguistic properties, namely imageability, metaphoricity, similarity in meaning between metaphorical and literal versions, naturalness and understandability for stories only, and familiarity for sentences only. An objective measure of complexity and length in words are also included for all stimuli. The use of stories allows the investigation of more natural and ecologically valid reading processes while the use of sentences allows a tighter control over extraneous sources of variability such as larger variation in the structure and the words used (as it is the case in stories). Our materials can be used for experiments measuring reading times, physiological responses (e.g., heart rate, skin conductance), brain activations through functional magnetic resonance imaging (fMRI), and perturbation of brain activity through transcranial magnetic

stimulation (TMS). Because our materials vary greatly in the position in which our metaphorical vs. literal key words appear and in their grammatical class, they may be less suitable for methods that are highly time-sensitive such as eye movements, pupil dilation, and electroencephalogram/magnetoencephalogram (EEG/MEG). Nevertheless, most of our metaphorical and literal versions of sentences are identical but for one word. Hence, no variability in the position of the word between the two versions may allow employment of our sentences in eye-movement, pupil dilation, and EEG/MEG studies.

To address the second aim and test the hypotheses formulated, differences between metaphorical and literal versions on the rated variables and correlations among them were explored. We predicted that 1) metaphorical versions of stories and sentences will show higher arousal level and imageability than the literal versions; 2) a U-shaped and a linear negative relationship between emotional valence and arousal will be observed in both metaphorical and literal stories and sentences; 3) positive correlations between degree of metaphoricity, imageability, and emotional arousal in both metaphorical and literal stories and sentences will be observed. Finally, 4) our study was exploratory with respect to correlations between other variables as these differ from the variables included in previous metaphor databases. In line with our first prediction, metaphorical stories were higher in emotional arousal, imageability and metaphoricity than their literal counterparts, while metaphorical sentences showed the same pattern but only metaphoricity reached the set significance threshold. This pattern is in line with a large body of theoretical and empirical research that posited and showed that during comprehension of conceptual metaphors sensorimotor representations are activated (Desai et al., 2013; Gibbs, 2011; Lacey et al., 2012; Lakoff & Johnson, 1980; Pomp et al., 2018), and with research showing

that metaphorical formulations are perceived as more emotionally arousing (Bohrn et al., 2012; Citron & Goldberg, 2014; Citron et al., 2016b; Rojo et al., 2014), beyond confirming that our metaphorical stimuli are indeed deemed as more metaphorical. However, isolated sentences did not apparently provide enough material or context for the metaphorical formulations to be rated as significantly easier to imagine and more emotionally arousing. Remember that all our stimuli were devoid of any explicit mention of emotional reactions (happy, sad, or their metaphorical renderings); while this may have affected ratings of sentences, stories may have been relatively unaffected given the fact that stories are more engaging overall, no matter whether literal or metaphorical (Citron et al., 2016b). Existing studies on the effect of context on metaphor comprehension show that the presence of context in the interpretation of literary metaphors facilitates comprehension as metaphors are rated as less difficult to understand and more meaningful (Bambini et al., 2014). Context also facilitates metaphor prediction as shown by cloze probability measures (participants are asked to type in endings to incomplete sentences) (Bambini et al., 2014) as well as by ERPs: metaphors preceded by supportive context are more successfully predicted in that the N400 component disappears compared to when the same metaphors are preceded by a minimal context (Bambini et al., 2016). These findings suggest that context may facilitate metaphor comprehension and interpretation, and, with regards to our findings, this facilitation may in turn lead to stronger simulation/imaginative processes and perceived emotional arousal (Miall & Kuiken, 1994b).

In addition, metaphorical stories were rated as less natural and metaphorical sentences as less familiar, than their literal counterparts. The former finding may be due to the fact that our short stories contained two or three metaphors that all rely on the same conceptual mapping, which is relatively unusual within a short discourse or

text, whereas the latter finding may be due to the fact that, despite all metaphors being conventional and well known by native speakers of German, some of them are used less frequently than others, and may therefore be perceived as less familiar.

In line with our second hypothesis, affective variables showed a positive quadratic relationship in all stories and in all sentences, and in metaphorical and literal subsets (except for literal stories) whereby increasingly valenced (positive and negative) stimuli were also higher in arousal. However, this relationship was asymmetrical only for sentences, including all sentences and both metaphorical and literal subsets, i.e., increasingly negative sentences were higher in arousal than positive ones (negative linear relationship between valence and arousal), whereas stories showed no significant linear trend. An example of negative sentence high in arousal is "*Karina bekam fast jeden Tag eine gesalzene Ohrfeige von ihrer Mutter*" *(Karina got a salty slap on the face by her mother almost every day)* whereas a positive sentence high in arousal is "*In Wien habe ich einen himmlischen Kaffee getrunken*" *(In Vienna, I drank a heavenly coffee)*. This pattern of results replicates and extends to metaphors previous findings on single words in several languages (Bradley & Lang, 1999; Citron, Weekes, & Ferstl, 2014; Hinojosa et al., 2015; Montefinese et al., 2013; Schmidtke, Schröder, Jacobs, & Conrad, 2014; Võ et al., 2009), on pictures (Lang et al., 1999), and on idioms (Citron et al., 2016a). The difference between stories and sentences may be due to the fact that stories tend to vary more in emotional valence, i.e., despite each of our stories being overall positive or negative in valence, its unfolding entails a mixture of events, impressions and reactions that may vary in emotional valence (see for example Hsu et al., 2015; Miall & Kuiken, 1994); therefore, this variable may contribute less to the overall emotional arousal of the story. Instead, isolated sentences contain less elements and tend to be

more unambiguously perceived as either positive or negative, therefore valence contributes to a sentence's arousal.

As already evident by the fact that the regression models predicting arousal contained imageability as a first strong predictor, and in line with our third hypothesis, these two variables showed a large positive partial correlation in all stories, all sentences and in their metaphorical and literal subsets, i.e., the easier to imagine the more arousing. This finding suggests that the ability to imagine a story or a sentence evokes stronger affective associations, possibly because of the activation of bodily representations; it is in line with the positive correlations between concreteness and arousal found for idioms (Citron et al., 2016a) and for Spanish words (Hinojosa et al., 2015), and between imageability and arousal for English words (Citron et al., 2014). However, other studies showed a different pattern (Montefinese et al., 2013), and most norming studies of single words that included affective variables did not explore correlations between these and either concreteness or imageability (Eilola & Havelka, 2010; Redondo, Fraga, Padrón, & Comesaña, 2007; Võ et al., 2009). Hence, more work on the relationship between imageability, concreteness and affective variables is needed, and it would be particularly informative with respect to figurative language processing. In fact, relevance theoretic accounts of figurative language processing point toward a special role of mental images in the comprehension of verbal metaphors and a specific function of verbal metaphors in evoking sensory representations (e.g., characterising the drunk at the bar as a "wheezing bagpipe") (Carston, 2018). In line with this account, in our study imageability was also positively correlated with metaphoricity for all stimuli and subsets except literal sentences. This confirms the difference between metaphorical and literal materials discussed above and is in line with findings from metaphor

databases (see Thibodeau et al., 2017 for an overview), although most of them do not include literal items (Bambini et al., 2014; Katz et al., 1988) or the latter are not paraphrases of the metaphors (Cardillo et al., 2010; Cardillo et al., 2017). Finally, the PANIG database showed a negative correlation between concreteness and idiomaticity, which is at odds with our finding. However, in that study participants were asked to rate how concrete they found the idiomatic meaning: this may have led them to try and ignore as much as possible the literal meaning and thus to consider less abstract idioms as less idiomatic; note also that the majority of their idioms had abstract meanings (Citron et al., 2016a).

The other affective dimension, emotional valence, showed instead a large negative partial correlation with understandability, in all stories and the literal subset (but not in the metaphorical subset), i.e., more negative stories were easier to understand. This is unlikely due to a sample bias, given that no correlation between complexity and valence was found. Possibly, negative stories may capture and withhold attention more effectively than positive or neutral stories (Nasrallah, Carmel, & Lavie, 2009) and readers engage with them more deeply, trying to make sense out of misfortune and understand its causes (Altmann, Bohrn, Lubrich, Menninghaus, & Jacobs, 2012), with the result that they find them more understandable. This is a tentative interpretation of rating data that would clearly need further empirical investigation.

Imageability also showed a large negative partial correlation with understandability for all stories and their subsets, i.e., the more understandable a story is perceived the less easy it is to imagine. If we combine this non-immediately intuitive result with the large positive partial correlation found between complexity and imageability (for all stories and literal stories only) - i.e., the more structurally

complex a story, the easier it is to imagine - we may postulate that less understandable or more complex stories need more effort to be processed; mental imagery may be used to support this process, more so than in the case of more understandable or simple stories, which are understood quickly without much need for the construction of a mental image. Also, more complex or stories are richer and provide more elements to create a mental image. Clearly, one needs to bear in mind that we are talking about ratings of stimuli and not about actual reading processes. Nevertheless, these interpretations are in line with data on reading processes which show that imagery facilitates comprehension of short narratives as assessed by literal and inferential questions (Joffe, Cain, & Marić, 2009).

Metaphoricity showed large partial negative correlations with naturalness in all stories and in the metaphorical subset, and with familiarity in all sentences and both metaphorical and literal subsets, thus replicating and confirming the results of the comparison between metaphorical and literal materials.

Furthermore, for all stories (but for neither of the two metaphorical and literal subsets), the easier to imagine the more natural, and the longer the more complex but also easier to understand. In addition, the more metaphorical a story the less structurally complex: this may be simply due to our less complex stimuli containing expressions deemed as higher in metaphoricity. Finally, all sentences and their subsets became increasingly more complex with increasing length. Metaphorical sentences with increasing metaphoricity were increasingly shorter, while literal sentences of increasing complexity were increasingly less familiar.

With the present work, we have provided an overview of the relationships between affective and psycholinguistic properties of metaphorical and literal materials, extending previous work on idioms to conceptual metaphors, and providing

helpful materials for the investigation of (figurative) language processing during natural reading processes as well as during reading of simple sentences. The limited number of stimuli created and the correlational nature of the analyses conducted prevent generalisation of our findings to metaphors, figurative language more generally, or to actual reading processes. Nevertheless, they provide initial insights that may be backed up by future databases and further empirical research on (figurative) language processing. For example, the relationship between metaphoricity and imageability, which in turns leads higher perceived levels or emotional arousal can be further explored, especially in literary texts that contain metaphors as well as other stylistic devices. The typical U-shaped and negative linear relationships between emotional valence and arousal previously shown for words, picture and idioms have been replicated and extended to conceptual metaphors, although natural short stories did not show any significant linear trend. Finally, other differences between stories and sentences highlight the need for the investigation of more natural reading processes, as results on highly de-contextualised and simplified materials may not reflect people's performance and cognitive processes in real-life settings.

Acknowledgments

Author contributions

Creation of the stimuli and the instructions was mostly conducted by NM, with input from FMMC. Preparation of online surveys and data collection were entirely conducted by NM. Data analysis was conducted by ML, with guidance from FMMC. ML wrote up most of the data analysis sub-section and the results section, with some contribution from FMMC. FMMC wrote up the rest of the manuscript.

References

Altarriba, J., & Bauer, L. M. (2004). The distinctiveness of emotion concepts: A comparison between emotion, abstract and concrete words. *American Journal of Psychology, 117*, 389-410.

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

Bambini, V., Bertini, C., Schaeken, W., Stella, A., & Di Russo, F. (2016). Disentangling metaphor from context: An ERP study. *Frontiers in Psychology, 7*, 559. doi:10.3389/fpsyg.2016.00559

Bambini, V., Canal, P., Resta, D., & Grimaldi, M. (2019). Time course and neurophysiological underpinnings of metaphor in literary context. *Discourse Processes, 56*, 77-97. doi:10.1080/0163853X.2017.1401876

Bambini, V., Resta, D., & Grimaldi, M. (2014). A dataset of metaphors from the Italian literature: Exploring psycholinguistic variables and the role of context. *PLoS ONE, 9*, e105634.

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

Boulenger, V., Hauk, O., & Pulvermueller, F. (2009). Grasping ideas with the motor system: Semantic somatotopy in idiom comprehension. *Cerebral Cortex, 19*, 1905-1914. doi:10-1093/cercor/bhn217

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Simuli, instruction manual and affective ratings* (C-1). Retrieved from Gainesville, FL:

Cacciari, C. (2014). Processing multiword idiomatic strings: many words in one? *The Mental Lexicon, 9*, 267-293.

Cacciari, C., Bolognini, N., Senna, I., Pellicciari, M. C., Miniussi, C., & Papagno, C. (2011). Literal, fictive and metaphorical motion sentences preserve the motion component of the verb: A TMS study. *Brain and Language, 119*, 149-157. doi:10.1016/j.bandl.2011.05.004

Cardillo, E. R., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2010). Stimulus design in and obstacle course: 560 matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavior Research Methods, 42*, 651-664. doi:10.3758/BRM.42.3.651

Cardillo, E. R., Watson, C., & Chatterjee, A. (2017). Stimulus needs ar a moving target: 240 additional matched literal and metaphorical sentences for testing

neural hypotheses about metaphor. *Behavior Research Methods, 49*, 471-483. doi:10.3758/s13428-016-0717-1

Carston, R. (2018). Figurative language, mental imagery, and pragmatics. *Metaphor and Symbol, 33*, 198-217. doi:10.1080/10926488.2018.1481257

Citron, F. M. M. (2012). Neural correlates of written emotion word processing: A review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language, 122*, 211-226. doi:10.1016/j.bandl.2011.12.007

Citron, F. M. M., Cacciari, C., Funcke, J., Hsu, C.-T., & Jacobs, A. M. (2019). Idiomatic expressions evoke stronger emotional responses in the brain than literal sentences. *Neuropsychologia, 131*, 233-248. doi:10.1016/j.neuropsychologia.2019.05.020

Citron, F. M. M., Cacciari, C., Kucharski, M., Beck, L., Conrad, M., & Jacobs, A. M. (2016a). When emotions are expressed figuratively: Psycholinguistic and affective norms of 619 idioms for German (PANIG). *Behavior Research Methods, 48*, 91-111. doi:10.3758/s13428-015-0581-4

Citron, F. M. M., & Goldberg, A. E. (2014). Metaphorical sentences are more emotionally engaging than their literal counterparts. *Journal of Cognitive Neuroscience, 26*, 2585-2595. doi:10.1162/jocn_a_00654

Citron, F. M. M., Güsten, J., Michaelis, N., & Goldberg, A. E. (2016b). Conventional metaphors in longer passages evoke affective brain response. *NeuroImage, 139*, 218-230. doi:10.1016/j.neuroimage.2016.06.020

Citron, F. M. M., Weekes, B. S., & Ferstl, E. C. (2014). How are affective word ratings related to lexico-semantic properties? Evidence from the Sussex Affective Word List (SAWL). *Applied Psycholinguistics, 35*, 313-331. doi:10.1017/S0142716412000409

Delaney-Busch, N., & Kuperberg, G. R. (2013). Friendly drug dealers and terrifying puppies: affective primacy can attenuate the N400 effect in emotional discourse contexts. *Cognitive Affective and Behavioural Neuroscience, 13*, 473-490. doi:10.3758/s13415-013-0159-5

Desai, R. H., Conant, L. L., Binder, J. R., Park, H., & Seidenberg, M. S. (2013). A piece of action: Modulation of sensory-motor regions by action idioms and metaphors. *NeuroImage, 83*, 862-869. doi:10.1016/j.neuroimage.2013.07.044

Diaz-Lago, M., Fraga, I., & Acuna-Farina, C. (2015). Time course of gender agreement violations containing emotional words. *Journal of Neurolinguistics, 36*, 79-93.

Drew, P., & Holt, E. (1988). Complainable matters: The use of idiomatic expressions in making complaints. *Social Problems, 35*, 398-417.

Drew, P., & Holt, E. (1998). Figures of speech: Figurative expressions and the management of topic transition in conversation. *Language in Society, 27*, 495-522.

Eilola, T. M., & Havelka, J. (2010). Affective norms for 210 British English and Finnish nouns. *Behavior Research Methods, 42*, 134-140. doi:10.3758/BRM.42.1.134

Fainsilber, L., & Ortony, A. (1987). Metaphorical uses of language in the expression of emotions. *Metaphor and symbolic activity, 2*, 239-250.

Ferstl, E. C., Rinck, M., & von Cramon, D. Y. (2005). Emotional and temporal aspects of situation model processing during text comprehension: An event-related fMRI study. *Journal of Cognitive Neuroscience, 17*, 724-739.

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

Gibbs, R. W. (2011). Evaluating conceptual metaphor theory. *Discourse Processes, 48*, 529-562.

Gibbs, R. W., & Colston, H. L. (2012). *Interpreting figurative meaning*: Cambridge University Press.

Gibbs, R. W., & Gerrig, R. J. (1989). How context makes metaphor comprehension seem "special". *Metaphor and Symbol, 4*(145-158).

Glucksberg, S. (2001). *Understanding figurative language: From metaphors to idioms*. New York, USA: Oxford University Press.

Glucksberg, S. (2008). How metaphors creat categories - quickly. In R. W. Gibbs (Ed.), *The Cambridge handbook of metaphor and thought* (pp. 67-83): Cambridge University Press.

Hamann, S., & Mao, H. (2002). Positive and negative emotional verbal stimuli elicit activity in the left amygdala. *NeuroReport, 13*, 15-19.

Herbert, C., Ethofer, T., Anders, S., Junghofer, M., Wildgruber, D., Grodd, W., & Kissler, J. (2009). Amygdala activation during reading of emotional adjectives - an advantage for pleasant content. *Social Cognitive and Affective Neuroscience, 4*, 35-49. doi:10.1093/scan/nsn027

Hinojosa, J. A., Martínez-García, N., Villalba-García, C., Fernández-Folgueiras, U., Sánchez-Carmona, A., Pozo, M. A., & Montoro, P. R. (2015). Affective norms of 875 Spanish words for five discrete emotional categories and two emotional dimensions. *Behavior Research Methods*. doi:10.3758/s13428-015-0572-5

Horton, W. S. (2007). Metaphor and reader's attribution of intimacy. *Memory & Cognition, 35*, 87-94.

Horton, W. S. (2013). Character intimacy influences the processing of metaphoric utterances during narrative comprehension. *Metaphor and Symbol, 28*, 148-166.

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

Joffe, V. L., Cain, K., & Marić, N. (2009). Comprehension problems in children with specific language impairment: does mental imagery training help? *International Journal of Language and Communication Disorders, 42*, 648-664. doi:10.1080/13682820601084402

Jones, L. L., & Estes, Z. (2006). Roosters, robins, and alarm clocks: Aptness and conventionality in metaphor comprehension. *Journal of Memory and Language, 55*, 18-32. doi:10.1016/j.jml.2006.02.004

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*, 191-214.

Keysar, B., & Glucksberg, S. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review, 97*, 3-18. doi:10.1037/0033-295X.97.1.3

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General, 1*, 14-34. doi:10.1037/a0021446

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition, 112*, 473-481. doi:10.1016/j.cognition.2009.06.007

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General, 143*, 1065-1081. doi:10.1037/a0035669

Lacey, S., Stilla, R., & Sathian, K. (2012). Metaphorically feeling: Comprehending textural metaphors activates somatosensory cortex. *Brain and Language, 120*, 416-421. doi:10.1016/j.bandl.2011.12.016

Lai, V. T., Willems, R. M., & Hagoort, P. (2015). Feel between the lines: Implied emotion in sentence comprehension. *Journal of Cognitive Neuroscience*. doi:10.1162/jocn_a_00798

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

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). *International affective picture system (IAPS): Instruction manual and affective ratings* (A-4). Retrieved from Gainsville, FL:

Miall, D. S., & Kuiken, D. (1994b). Foregrounding, defamiliarization, and affect: Response to literary stories. *Poetics, 22*, 389-407.

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2013). The adaptation of the affective norms for English words (ANEW) for Italian. *Behavior Research Methods, 46*, 887-903. doi:10.3758/s13428-013-0405-3

Nasrallah, M., Carmel, D., & Lavie, N. (2009). Murder, she wrote: Enhanced sensitivity to negative word valence. *Emotion, 9*(5), 609-618. doi:10.1037/a0016350

Ortony, A., & Fainsilber, L. (1987). *The role of metaphors in descriptions of emotions.* Paper presented at the 1987 workshop on theoretical issues in natural language processing.

Ortony, A., Schallert, D., Reynolds, R., & Antos, S. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behaviour, 17*, 465-477.

Pierce, R. S., & Chiappe, D. L. (2009). The roles of aptness, conventionality, and working memory in the production of metaphors and similes. *Metaphor and Symbol, 24*, 1-19. doi:10.1080/10926480802568422

Pomp, J., Bestgen, A.-K., Schulze, P., Müller, C. J., Citron, F. M. M., Suchan, B., & Kuchinke, L. (2018). Lexical olfaction recruits olfactory orbitofrontal cortex in metaphorical and literal contexts. *Brain and Language, 179*, 11-21. doi:10.1016/j.bandl.2018.02.001

Prandi, M. (2010). Typology of metaphors: Implications for translation. *Mutatis Mutandis, 3*, 304-332.

Redondo, J., Fraga, I., Padrón, I., & Comesaña, M. (2007). The Spanish adaptation of ANEW (Affective Norms for English Words). *Behavior Research Methods, 39*, 600-605. doi:10.3758/BF03193031

Rodríguez-Ferreiro, J., & Davies, R. A. I. (2019). The graded effect of valence on word recognition in Spanish. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*, 851-868.

Rojo, A., Ramos, M., & Valenzuela, J. (2014). The emotional impact of translation: A heart rate study. *Journal of Pragmatics, 71*, 31-44. doi:10.1016/j.pragma.2014.07.006

Roncero, C., de Almeida, R. G., Martin, D. C., & de Caro, M. (2016). Aptness predicts metaphor preference in the lab and on the internet. *Metaphor and Symbol, 31*, 31-46. doi:10.1080/10926488.2016.1116908

Schmidtke, D., Schröder, T., Jacobs, A. M., & Conrad, M. (2014). ANGST: Affective norms for German sentiment terms, derived from the affective norms of English words. *Behavior Research Methods*. doi:10.3758/s13428-013-0426-y

Thibodeau, P., & Durgin, F. H. (2008). Productive figurative communication: Conventional metaphors facilitate the comprehension of related nover metaphors. *Journal of Memory and Language, 58*, 521-540. doi:10.1016/j.jml.2007.05.001

Thibodeau, P. H., & Durgin, F. H. (2011). Metaphor aptness and conventionality: A processing fluency account. *Metaphor and Symbol, 26*, 206-226. doi:10.1080/10926488.2011.583196

Thibodeau, P. H., Sikos, L., & Durgin, F. H. (2017). Are subjective ratings of metaphors a red herring? The big two dimensions of metaphoric sentences. *Behavior Research Methods*, 1-14. doi:10.3758/s13428-017-0903-9

Vinson, D. P., Ponari, M., & Vigliocco, G. (2014). How does emotional content affect lexical processing? *Cognition and Emotion, 28*, 737-746. doi:10.1080/02699931.2013.851068

Võ, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods, 41*, 534-538. doi:10.3758/BRM.41.2.534

Wilson, D., & Carston, R. (2006). Metaphor, relevance and the 'emergent property' issue. *Mind & Language, 21*, 404-433.

Winter, B. (2016). Taste and smell words form and affectively loaded and emotionally flexible part of the English lexicon. *Language, Cognition and Neuroscience*. doi:10.1080/23273798.2016.1193619