It is not always a matter of time: Addressing the costs of metaphor and metonymy through a Speed-Accuracy Trade-off study

Valentina Bambini[1], Lewis Bott[2], Petra B. Schumacher[3]

[1] Center for Neurocognition, Epistemology and theoretical Syntax, University School for Advanced Studies IUSS Pavia, Pavia, Italy

[2] School of Psychology, Cardiff University, United Kingdom

[3] Department of German Language and Literature I, Linguistics, University of Cologne, Cologne, Germany

Acknowledgment: This work is based on a collaborative research initiative funded by the European Science Foundation's EURO-XPRAG Research Networking Program. We acknowledge the contribution of Dr. Marta Ghio in the initial phases of this study.

Author contribution: Idea of the study: VB, PS. Experimental design: VB, LB, PS. Investigation and formal analysis: VB, LB. First draft of the manuscript: VB. Manuscript writing: VB, LB, PS.

Abstract (max 250)

One of the most debated topics in figurative language studies is whether the access to non-literal meanings is direct or indirect. Although models that argue for longer processing times for figurative compared to literal meanings have been largely criticized, figurative language is often associated with increased cognitive work. We investigated whether such greater cognitive work is indicative of more time-consuming processes or rather lower availability of figurative meanings, and whether there are differences between figurative types. We used a multi-response Speed-Accuracy Trade-off paradigm, where a meaningfulness judgment task was combined with a response deadline procedure to estimate speed and accuracy independently for metaphorical (*Those dancers are butterflies*) and metonymic sentences (*That student reads Camilleri*), compared with literal equivalents. While both metaphors and metonymies showed lower asymptote, that is, they were judged less accurately than literal counterparts, only metonymies were associated with a processing delay. Moreover, the difference in asymptote with respect to the literal condition was greater for metaphor than for metonymy. These findings indicate that the process that derives metaphor and metonymy is more complex than the process that derives literal meanings, even more so for metaphor. The processing delay, however, is present only in the case of metonymies. Taken together, our study offers key findings that reconcile a lively debate on the time course of figurative language comprehension, showing that the cost of non-literal meaning is not always a matter of time, and depends also on the figurative type.

Keywords: metaphor; metonymy; figurative language; speed-accuracy trade-off

It is not always a matter of time: Addressing the costs of metaphor and metonymy through a Speed-Accuracy Trade-off study

Introduction

Broadly speaking, there are two classes of models to explain the time course of figurative language processing (Gibbs & Colston, 2012). The first is the *standard model*. Here, figurative language is assumed to be processed in serial stages, unfolding through the analysis of the literal meaning, its rejection as not appropriate in the context, and the construction of an alternative, non-literal, interpretation (Grice, 1975; Searle, 1979). From a processing perspective, the standard model predicts that, since the comprehension of figurative meaning requires more steps than the interpretation of literal meaning, the processing of the former should take more time than the processing of the latter (Janus & Bever, 1985). The alternative is that the figurative meaning is *directly accessed*. This model claims that figurative processing does not mandatorily require the full analysis of the literal meaning and can be completed in a single stage (Gibbs, 1994). Accordingly, processing figurative meanings does not necessarily take longer than processing literal meanings.

The general consensus is that the standard model is not correct (Gibbs & Colston, 2012; Glucksberg, 2003). A number of different studies argue that comprehenders do not always (and completely) process a literal meaning before deriving the figurative one and acknowledge the role of context, salience, familiarity and other features of the communicative situation to account for the modulation of (in)directness in accessing figurative interpretations (Ferretti, Katz, Schwint, Patterson, & Pradzynski, 2020; Giora, 2003; Katz & Ferretti, 2001). However, there is also evidence that, other things being equal, figurative language is more costly, as for instance reflected in longer response times observed in behavioral studies on different figures of speech (Columbus et al., 2014; Lowder & Gordon, 2013; Noveck, Bianco, & Castry, 2001). Neurophysiological methods, such as the Event-Related Brain Potential method (ERP), also show costs of generating figurative interpretations. When figurative and literal expressions are compared, context being equal, the figurative condition (metaphors, metonymies, irony or proverbs) typically elicits a greater brain response, visible in early phases and sometimes prolonged in time (Bambini, Bertini, Schaeken, Stella, & Di Russo, 2016; Schumacher, 2011, 2014; Spotorno, Cheylus, Van Der Henst, & Noveck, 2013). Importantly, ERP methodology reveals greater efforts in processing figurative as compared to literal meanings even when reading times are equivalent (Coulson & Van Petten, 2002; Ferretti, Schwint, & Katz, 2007). How can the evidence of higher costs for figurative language be reconciled with the idea that figurative processing does not require more time than the literal meaning?

Here we consider two answers to this question. The first is that the costs observed for figurative meaning reflect lower accuracy rather than slower processing. Specifically, longer response times for figurative interpretations may reflect more errors in interpreting words in figurative contexts rather than additional time in retrieving figurative meanings (Bott, Bailey, & Grodner, 2012; McElree & Nordlie, 1999). The second is that different sorts of figurative language (metonymy, irony, metaphor) carry different time course profiles and when considered as an aggregated whole, inconsistencies arise across studies.

Distinguishing between speed and accuracy can be achieved using the Speed Accuracy Trade-Off procedure (SAT), a chronometric paradigm developed in the 70s (Reed, 1973). In SAT, the experimenter specifies the time at which a response has to be executed, by including either one or multiple response deadlines. Then a full time-course function is derived that measures how accuracy of interpretation (measured in d') varies with processing time. This function is characterized by three parameters: λ – related to the accuracy of processing, estimating the asymptotic level of performance; β – related to the speed of processing, estimating the rate at which accuracy grows from chance to asymptote; and δ – also related to processing speed, estimating the intercept of the function, or the point at which participants are first sensitive to the information necessary to make an accurate discrimination (i.e., d' departs from 0, chance performance).

Surprisingly, only two studies till now have used SAT in the field of figurative language. McElree & Nordlie (1999) investigated metaphor processing and showed that the SAT curves for metaphorical and literal sentences showed no differences in the speed parameters, although asymptotic level of performance was lower for metaphorical than for literal sentences. Bott, Rees, & Frisson (2016) focused on familiar metonymies and showed that these were associated with lower accuracy compared to literal sentences but there were no differences in the speed parameters. These two studies suggest that the literal meaning has no temporal priority over the figurative meaning (in contrast with the standard model), but they also indicate that the figurative meaning is less available than the literal meaning. This is consistent with observed differences in costs (e.g., reading times, ERP signals) between literal and figurative meanings in previous studies (e.g. participants may choose to delay responding in order to maximize accuracy, thereby leading to longer reaction times but without the additional processing stages required by the standard model). However, without a direct comparison between metaphor and metonymy, it is not possible to say whether there are differences between metaphor and metonymy in meaning availability.

Indeed, there are theoretical reasons to expect differences in the time course profile between metaphor and metonymy, which represent a classic dichotomy in literary and stylistic studies (Jakobson, 1956). On the theoretical side, in the context of Relevance Theory, metaphors are said to involve broadening and narrowing of lexically encoded concepts, whereas metonymies are seen as reference transfer (Wilson & Carston, 2007; but see Schumacher, 2013 on a typology of metonymy). Metaphors and literal meanings would therefore involve the same qualitative processes but applied to different degrees, suggesting parallel processes, but metonymy requires an additional qualitatively different process (reference transfer), suggesting an additional serial process. Earlier intercepts or steeper retrieval rates and/or higher accuracy might therefore be expected for metaphor compared to metonymy (relative to literal controls). On the empirical side, the studies that directly compared metaphor and metonymy are limited in number but they consistently find greater costs for metaphor compared to metonymy, as reflected in longer response times, more pronounced ERP effects, or later developmental trajectories (Bambini, Ghio, Moro, & Schumacher, 2013; Rundblad & Annaz, 2010; Weiland, Bambini, & Schumacher, 2014; but see Gibbs, 1990).

In the present study we used the SAT paradigm to shed light on the debate over the time course of figurative language comprehension. We asked whether the increased cognitive efforts observed in the literature are related to lower availability of non-literal meanings or to actual slowing down in time (or to both). Compared to the two previous SAT studies on figurative language, we used novel and tightly controlled sets of items, in a different language, and a different SAT technique. Moreover, we were interested in exploring further – for the first time in a single SAT study - possible differences between metaphor and metonymy.

Method

Participants

Eighteen university students, native speakers of Italian (right-handed, 10M/8F, mean age = 22.5, SD = 1.9), participated in the study and were reimbursed for their participation.

Materials

Materials included: a metaphor set (from Bambini et al., 2013), with 48 novel nominal metaphors (e.g., 'Those dancers are *butterflies*'), 48 literal (e.g., 'Those insects are *butterflies*') and 48 anomalous counterparts (e.g., 'Those bottles are *butterflies*'), plus 48 additional senseless filler items obtained by recombining words in the set (e.g., 'Those insects are *tables*');

a metonymy set (from Bambini et al., 2013), with 48 producer-for-product familiar metonymies (e.g., 'That student reads *Camilleri*'), 48 literal (e.g., 'That reporter interviews *Camilleri*') and 48 anomalous counterparts (e.g., 'That chef cooks *Camilleri*'), plus 48 additional senseless filler items obtained by recombining words in the set ('That student reads *Milan*'); a set of meaningful and meaningless filler sentences (N =768), with the same length in words of the experimental sentences and randomly interspersed with the other items.

In total, materials included 1170 sentences, with meaningful items being 58% of stimuli, and metaphor and metonymy 9%. Following the selection criteria of Bambini et al. (2013), the metaphor and the metonymy sets were restricted for the analysis to 42 triplets each. The obtained sets were balanced for length and frequency of the target words, and controlled for familiarity, meaningfulness, and difficulty (see the rating in Bambini et al., 2013).

Procedure

We used a multiple response SAT paradigm (Bott et al., 2016), where participants were asked to make a binary sensicality judgment at successive deadlines. Each trial was structured as displayed in Figure 1. Half of the participants started with the ''sense'' key as an initial undecided response and the other half started with the "nonsense" key. Participants were encouraged to modulate their responses if their judgment changed during the trial by switching key, as long as the response tones were still sounding.

The experiment consisted of a 1-h training session and three 1-h experimental sessions, completed over a period of two days. Each experimental session included 390 trials (128 experimental sentences, 262 fillers), with the triplets of the experimental sets split in different sessions and an equal number of figurative, literal, and anomalous sentences in each session.

Data analysis

The metaphor and the metonymy sets were analyzed separately, following the same procedure. We removed one participant who obtained an average d' of less than 1.0 on the longest time lag. Accuracy was measured in terms of d' corresponding to the difference between the z-score of proportion hits and the z-score of proportion false alarms. Hits were the “sense” responses to the figurative and literal sentences; false alarms were the “sense” responses to the anomalous sentences. Following Bott et al. (2012), proportion hits were calculated as the total number of correct figurative/literal trials plus 0.5, divided by the total number of figurative/literal trials plus 1. The same correction was applied for calculating the proportion false alarms. We calculated d's in fifteen 350 ms time bins surrounding the response cues (McElree, Pylkkänen, Pickering, & Traxler, 2006).

We derived SAT function by measuring accuracy (d' units) as a function of processing time (time of the response cue plus response latency measured in ms) with an exponential approach to limit. For parameter estimation we applied the maximum likelihood estimation (MLE) method which consists in seeking the parameter values that are most likely to have generated the data (Liu & Smith, 2009).

Following Liu & Smith (2009) and given that a close examination of our data revealed high subject variability, we performed the analyses on the individual participants’ data in order to avoid averaging artifacts in the shape of the functions. For the analysis of individual participants’ data we adopted the procedure described in Bott et al. (2012) and McElree & Nordlie (1999). For each participant we fitted the fully parameterized SAT model in which separate asymptote, beta, and intercept parameters were set for each condition. More specifically, this 2λ-2β-2δ model has six parameters: λ(figurative), β(figurative), δ(figurative), λ(literal), β(literal), δ(literal). Then, at the group level, we compared individual fitted parameter values between the figurative and the literal conditions by using Wilcoxon Signed Rank test. A similar procedure was employed for the item analysis. First, for each item separately we fitted the fully parameterized 2λ-2β-2δ model. Then, we compared parameter values between the figurative and the literal conditions by means of Wilcoxon Signed Rank test.

Finally, we compared the metaphorical and metonymical dynamic curves by using the parameter values obtained in the by subject analysis and performing the following comparisons through Wilcoxon Signed Rank test: (i) asymptote parameter: (λ metaphor – λ literal) vs. (λ metonymy – λ literal); (ii) rate parameter: (β metaphor – β literal) vs. (β metonymy – β literal); (iii) intercept parameter: (δ metaphor – δ literal) vs. (δ metonymy – δ literal).

Results

Metaphorical vs. literal sentences

To obtain an empirical measure of asymptotic recognition accuracy we averaged the d's for the last three time-lags (McElree et al., 2006). The asymptotes were higher for literal than for metaphorical sentences (median difference = 1.61, $V_1 = 0$, n = 17, $p < 0.001$; median difference = 1.50, $V_2 = 13.5$, n = 42, $p < 0.001$).

We fitted the six-parameter models (2λ-2β-2δ) to each individual (Figure 2A) and compared estimated parameter values at the group level (Figure 3A). Consistently with the analysis on d', by comparing estimated λ parameter values we found a difference between the literal and the metaphorical conditions ($V_1 = 9$, n = 17, $p = 0.0005$; $V_2 = 94$, n = 42, $p < 0.001$). This indicates that readers were less likely to compute a meaningful interpretation of the metaphorical sentences than of the literal sentences. As for the speed of processing, there were no significant differences in rate ($V_1 = 43$, n = 17, ns; $V_2 = 556$, n = 42, *ns*) nor in intercept ($V_1 = 73$, n = 17, *ns*; $V_2 = 590$, n = 42, *ns*).

For a subset of participants (n = 6) there was evidence of a late-intercept function ($\delta > 2$). Although we controlled the inter-individual variability by fitting the model on individual data, we ensured that the group-level result is not due to individual strategies by performing the analysis excluding the participants with the delay-strategy. Consistently with the results reported above, we found a difference in the asymptote (asymptote: $V_1 = 7$, n = 11, $p = 0.018$; $V_2 = 74$, n = 42, $p < 0.001$). No reliable differences in either the rate or the intercept were found between conditions (rate: $V_1 = 17$, n = 11, *ns*; $V_2 = 511$, n = 42, *ns*; intercept: $V_1 = 22$, n = 11, *ns*; $V_2 = 499$, n = 42, *ns*).

Metonymical vs. literal sentences

We computed the average accuracy (d' units) across the three last time-lags in order to provide an empirical measure of asymptotic performance. The asymptotes were higher for literal than for metonymical sentences (median difference = 0.44, $V_1 = 13$, n = 17, $p = 0.001$; median difference = 0.14, $V_2 = 239$, n = 42, $p = 0.01$).

We fitted the six-parameter models (2λ-2β-2δ) to each individual (Figure 2B) and compared estimated parameter values at the group level (Figure 3B). Consistently with the d' measure, we found a difference in the estimated λ parameters between the literal and the metonymical conditions (V_1= 15, n = 17, $p = 0.002$). As for speed processing, readers were slower in interpreting metonymical as compared to literal sentences. Although there were no significant differences in the rate estimated values ($V_1 = 65$, n = 17, ns), significant differences in the intercept estimated values were found ($V_1 = 130$, n = 17, $p = 0.009$). The same results were obtained by excluding participants (n = 6) who adopted a delay strategy in responding (asymptote: $V_1 = 4$, n = 11, $p = 0.006$; rate: $V_1 = 38$, n = 11, *ns*; intercept: $V_1 = 65$, n = 11, $p = 0.001$). However, we did not find significant differences in the items analysis (asymptote: $V_2 = 365$, n = 42, *ns*; rate: $V_2 = 487$, n = 42, *ns*; intercept: $V_2 = 491$, n = 42, *ns*), possibly due to insufficient data to accurately model individual curves for each item (17 data points per cell for each item vs. 42 per cell for each subject).

Metaphorical vs. Metonymical sentences

The comparison between the parameter values describing the dynamic functions of metaphorical and metonymical sentences revealed a significant effect with respect to the accuracy parameter (V = 18, n = 17, $p = 0.003$), with the difference between the metaphorical and the literal conditions being greater than the difference between the metonymical and the literal conditions. As for the speed parameters, no significant results were obtained (rate: V = 52, n = 17, *ns*; intercept: V = 45, n = 17, *ns*).

Discussion

The present study aimed at investigating the time course of two different types of figurative language, metaphor and metonymy, by applying an experimental paradigm specifically suitable for disentangling speed and accuracy of processing. The results can be summarized in three points: i) metaphors were less accurate (lower asymptote) than their literal counterparts, while they were processed at the same speed (same intercept and same rate); ii) metonymies too were less accurate and additionally they showed dynamics differences (a later intercept in the by participant data); iii) metaphors were less accurate with respect to literal sentences than metonymies.

Starting with metaphor, we showed that the pattern of lower asymptote yet equal speed compared to literal language described in McElree & Nordlie (1999) holds also with a different set of items, in a different language (Italian vs. English), and with another deadline technique (multiple response SAT vs. standard SAT). More generally, our findings support the view that literal language has no temporal priority over metaphorical meanings. We also found a lower asymptotic accuracy associated with metaphors (i.e., a lower probability of successfully retrieving the correct meaning). There are a number of possible explanations for this.

Previous literature reported lower accuracy (but equal speed) when participants were asked to recover less typical semantic relations (e.g., Corbett & Wickelgren, 1978). It is possible that metaphors involve less constrained meanings and their lower accuracy is due to failure in retrieving the key semantic properties of the metaphor's vehicles (e.g., *butterflies*) and applying them to the topic (e.g., *dancers*; McElree & Nordlie, 1999). Another explanation can be found in Relevance Theory and the continuity view (Wilson & Carston, 2007). For Relevance Theory, the pragmatic process of conceptual adjustment of lexically encoded concepts does not differ between literal and figurative senses, yet such a process is more radical in the case of metaphor compared to literal expressions. Our results can be explained by assuming that the degree of broadening necessary for metaphor (e.g., broadening BUTTERFLIES so as to include dancers) is much greater than for literal meanings and so carries a greater risk of error (e.g., not broadening sufficiently to include dancers, context not sufficiently specified to direct the broadening in the right direction). Crucially, greater broadening does not add to the time necessary to complete the process. Broadening cannot therefore be a serial process in which each additional broadening feature is considered sequentially, but a parallel process in which many possible broadenings are considered at the same time. It is also possible that metaphorical expressions might carry an array of potential interpretations, especially in minimal contexts as in the case of the stimuli used here. Considering this multitude of weakly implied meanings might contribute to lower the probability of reaching a final, correct interpretation.

The reduced availability of figurative meaning is consistent with a number of studies that showed a more pronounced cognitive response for non-literal language (Bambini et al., 2016; Schumacher, 2011), especially without differences found in reading times (Ferretti et al., 2007). Moreover, our results are compatible with a series of important findings concerning metaphorical processing. For instance, the absence of time difference we reported can accommodate the metaphor interference effect (Glucksberg, 2003) as well as evidence of early availability of figurative meaning coming from priming studies (Blasko & Connine, 1993), since people can automatically and immediately activate the available metaphorical meanings.

For metonymy, we reported a lower asymptote compared to literal expressions, as in the work of Bott et al. (2016), but, differing from it, we also observed dynamics differences, with metonymical sentences being processed with a later intercept than literal sentences. Bott et al. (2016) explained the lower asymptote by arguing that the retrieved metonymic interpretations were less plausible than the literal ones. Participants were less able to find an interpretation in which context and target word meaning were consistent for sentences in the metonymic condition than the literal condition. However, in our study, sentences were matched across conditions for meaningfulness and difficulty. It is therefore unlikely that global plausibility could account for the accuracy difference. Instead, some process specifically related to the metonymic mechanism must be particularly error prone, such as the reference substitution process of Relevance (Wilson & Carston, 2007).

Is this lower accuracy accompanied by delayed processing? Bott et al. (2016) did not observe significant dynamics differences. Their data therefore supports the direct access view of metonymic meanings. In contrast, we found a later intercept for metonymic meanings than literal meanings, albeit only in the participants analysis. Intercept differences can be explained by longer processes in one condition, or by additional processes computed serially in one condition compared to the other, such as the need to reject a literal meaning before deriving an implied meaning (Bott et al., 2012) or reanalysis of the grammatical structure (Bornkessel, McElree, Schlesewsky, & Friederici, 2004). The standard model of figurative language (Grice, 1975) is therefore consistent with our findings, in which the literal meaning is first rejected before a metonymic meaning is derived. Other accounts are also possible. The reference substitution process of Relevance (Wilson & Carston, 2007) could operate serially, so that literal references are computed before metonymic processes, especially in cases where the metonymic reading is not very familiar. Grammatical reanalysis might also have occurred with our stimuli because an animacy violation had to be accommodated (as it would for many types of metonymy). For example, there is an animacy mismatch in the metonymic use "reading Camilleri", but not in the literal use "interviewing Camilleri" (but see Weiland-Breckle & Schumacher, 2017 for no neurophysiological correlates at the proper name after a minimally licensing context). It should be noted, however, that the intercept differences were limited to the participant analysis (not the item analysis). Moreover, McElree et al. (2006) argued that serial models typically predict larger dynamics effects than those observed here (intercept differences = 5 ms). Therefore, the difference in speed must be taken with caution.

A novel aspect of our study was the direct comparison between metaphor and metonymy. The differences centered around the asymptotic accuracy, with metaphors departing more greatly from literal sentences compared to metonymies. In other words, participants were less accurate in finding an appropriate metaphoric interpretation than a metonymic interpretation. While different words were used across conditions, the sentences were controlled for many psycholinguistic variables such as frequency and familiarity, and thus it seems unlikely that this result is due to structural differences or general plausibility. The result is more likely to reflect a genuine difference in meaning availability between metaphor and metonymy. Some component of metaphor processing is more error-prone than metonymy processing, independent of the specifics of the stimuli. For example, Relevance Theory describes the adjustment of the lexically encoded concept as more radical in the case of metaphor than in the case of metonymy (Wilson & Carston, 2007): while the former involves the broadening of the concept through the drop of some of its definitory characteristics (e.g., dropping the feature ‘insect’ for the concept BUTTERFLY when used in *Those dancers are butterflies*), the latter requires the search for real world associations (e.g., ‘book’ for CAMILLERI when used in *That student reads Camilleri*).

To conclude, this study offers key findings that reconcile a lively debate on the time course of figurative language comprehension. We make two specific contributions (i) The process that derives metaphor and metonymy is more complex than the process that derives literal meanings, even more so in the case of metaphor. (ii) For metonymy, there is a processing delay associated with its interpretation (cf. Bott et al., 2016), at least for some items, consistent with the standard model and other theories that predict time differences, but for metaphor, no evidence of such a difference exists (consistent with McElree & Nordlie, 1999). Future work should concentrate on why availability differences exist between figurative and literal meanings and why metonymy incurs processing delays. More generally our study illustrates the advantages of separating speed from accuracy and metaphor from metonymy.

References

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

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

Blasko, D. G., & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19*(2), 295–308. https://doi.org/10.1037//0278-7393.19.2.295

Bornkessel, I., McElree, B., Schlesewsky, M., & Friederici, A. D. (2004). Multi-dimensional contributions to garden path strength: Dissociating phrase structure from case marking. *Journal of Memory and Language*, *51*(4), 495–522. https://doi.org/10.1016/j.jml.2004.06.011

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language*, *66*(1), 123–142. https://doi.org/10.1016/j.jml.2011.09.005

Bott, L., Rees, A., & Frisson, S. (2016). The time course of familiar metonymy. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(7), 1160–1170. https://doi.org/10.1037/xlm0000218

Columbus, G., Sheikh, N. A., Côté-Lecaldare, M., Häuser, K., Baum, S. R., & Titone, D. (2014). Individual differences in executive control relate to metaphor processing: an eye movement study of sentence reading. *Frontiers in Human Neuroscience*, *8*, 1057. https://doi.org/10.3389/fnhum.2014.01057

Corbett, A. T., & Wickelgren, W. A. (1978). Semantic memory retrieval: analysis by speed accuracy tradeoff functions. *The Quarterly Journal of Experimental Psychology*, *30*(1), 1–15. https://doi.org/10.1080/14640747808400648

Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: an event-related potential study. *Memory & Cognition*, *30*(6), 958–968. https://doi.org/10.3758/bf03195780

Ferretti, T. R., Katz, A. N., Schwint, C. A., Patterson, C., & Pradzynski, D. (2020). How discourse constraints influence neurolinguistic mechanisms during the comprehension of proverbs. *Cognitive, Affective and Behavioral Neuroscience*, *20*(3), 604–623. https://doi.org/10.3758/s13415-020-00790-9

Ferretti, T. R., Schwint, C. A., & Katz, A. N. (2007). Electrophysiological and behavioral measures of the influence of literal and figurative contextual constraints on proverb comprehension. *Brain and Language*, *101*(1), 38–49. https://doi.org/10.1016/j.bandl.2006.07.002

Gibbs, R. W. (1990). Comprehending figurative referential descriptions. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *16*(1), 56–66. https://doi.org/10.1037/0278-7393.16.1.56

Gibbs, R. W. (1994). *The Poetics of Mind: Figurative Thought, Language, and Understanding*. New York: Cambridge University Press.

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

Giora, R. (2003). *On our mind: Salience, context and figurative language*. New York: Cambridge University Press.

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences*, *7*(2), 92–96. https://doi.org/10.1016/S1364-6613(02)00040-2

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics, Vol. 3, Speech Acts* (pp. 41–58). New York: Academic Press.

Jakobson, R. (1956). Two aspects of language and two types of aphasic disturbance. In R. Jakobson & M. Halle (Eds.), *Fundamentals of language* (pp. 53–83). S-Gravenhage: Mouton.

Janus, R. A., & Bever, T. G. (1985). Processing of metaphoric language: An investigation of the three-stage model of metaphor comprehension. *Journal of Psycholinguistic Research*, *14*(5), 473–487. https://doi.org/10.1007/BF01666722

Katz, A. N., & Ferretti, T. R. (2001). Moment-by-moment reading of proverbs in literal and nonliteral contexts. *Metaphor and Symbol*, *16*(3–4), 193–221. https://doi.org/10.1080/10926488.2001.9678895

Liu, C. C., & Smith, P. P. L. (2009). Comparing time-accuracy curves: Beyond goodness-of-fit measures. *Psychonomic Bulletin and Review*, *16*(1), 190–203. https://doi.org/10.3758/PBR.16.1.190

Lowder, M. W., & Gordon, P. C. (2013). It's hard to offend the college: Effects of sentence structure on figurative-language processing. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(4), 993–1011. https://doi.org/10.1037/a0031671

McElree, B., & Nordlie, J. (1999). Literal and figurative interpretations. *Psychonomic Bulletin*, *6*(3), 486–494. https://doi.org/10.3758/BF03210839

McElree, B., Pylkkänen, L., Pickering, M. J., & Traxler, M. J. (2006). A time course analysis of enriched composition. *Psychonomic Bulletin & Review*, *13*(1), 53–59. https://doi.org/10.3758/BF03193812

Noveck, I. A., Bianco, M., & Castry, A. (2001). The costs and benefits of metaphor. *Metaphor and Symbol*, *16*(1–2), 109–121.

https://doi.org/10.1080/10926488.2001.9678889

Reed, A. V. (1973). Speed-accuracy trade-off in recognition memory. *Science*, *181*(4099), 574–576. https://doi.org/10.1126/science.181.4099.574

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

Schumacher, P. B. (2011). The hepatitis called ...: Electrophysiological evidence for enriched composition. In J. Meibauer & M. Steinbach (Eds.), *Experimental Pragmatics/Semantics* (pp. 199–219). Amsterdam: John Benjamins.

Schumacher, P. B. (2013). When combinatorial processing results in reconceptualization: toward a new approach of compositionality. *Frontiers in Psychology, 4*, 677. https://doi.org/10.3389/fpsyg.2013.00677

Schumacher, P. B. (2014). Content and context in incremental processing: "the ham sandwich" revisited. *Philosophical Studies*, *168*(1), 151–165. https://doi.org/10.1007/s11098-013-0179-6

Searle, J. (1979). Metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 83–111). Cambridge: Cambridge University Press.

Spotorno, N., Cheylus, A., Van Der Henst, J.-B., & Noveck, I. A. (2013). What's behind a P600? Integration operations during irony processing. *PLoS ONE*, *8*(6), e66839. https://doi.org/10.1371/journal.pone.0066839

Weiland-Breckle, H., & Schumacher, P. (2017). Artist-for-work metonymy: Type clash or underspecification? *Mental Lexicon*, *12*(2), 219–233. https://doi.org/10.1075/ml.16014.wei

Weiland, H., Bambini, V., & Schumacher, P. (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

Wilson, D., & Carston, R. (2007). A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In N. Burton-Roberts (Ed.), *Pragmatics* (pp. 230–259). Basingstoke: Palgrave Macmillan.

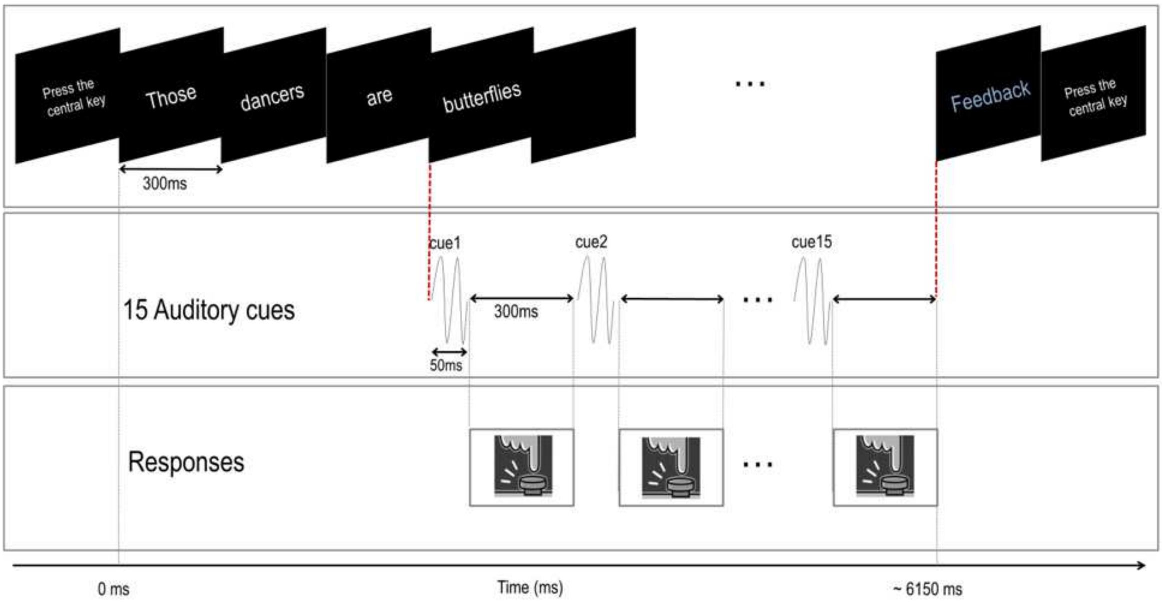

Figure 1. Schematic representation of an experimental trial.

Each sentence was presented word by word at a fix rate (300 ms) with the onset of the last word initiating a sequence of 15 tones (50 ms, 1000 Hz, intercue interval: 350 ms) which served as the signal to cue participants' response.

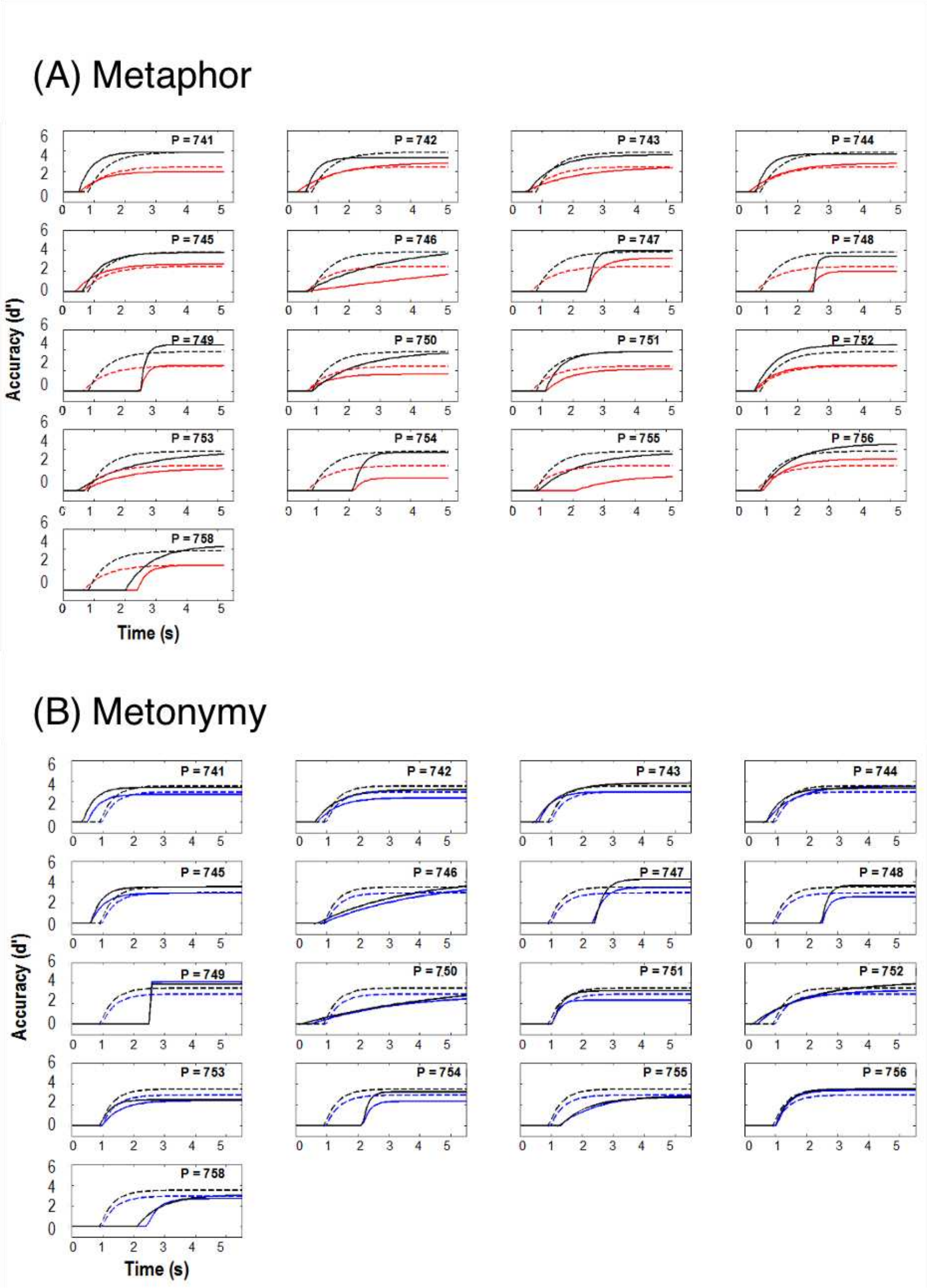

Figure 2. SAT functions for metaphor and metonymy.

For each Participant (P = participant's ID), the best fitting SAT functions (continuous lines) for the figurative and literal conditions are illustrated (in panel A: red = metaphorical condition; black = literal condition; in panel B: blue = metonymic condition; black = literal condition). The dotted lines represent the best fitting SAT functions for the median data (n participants = 17).

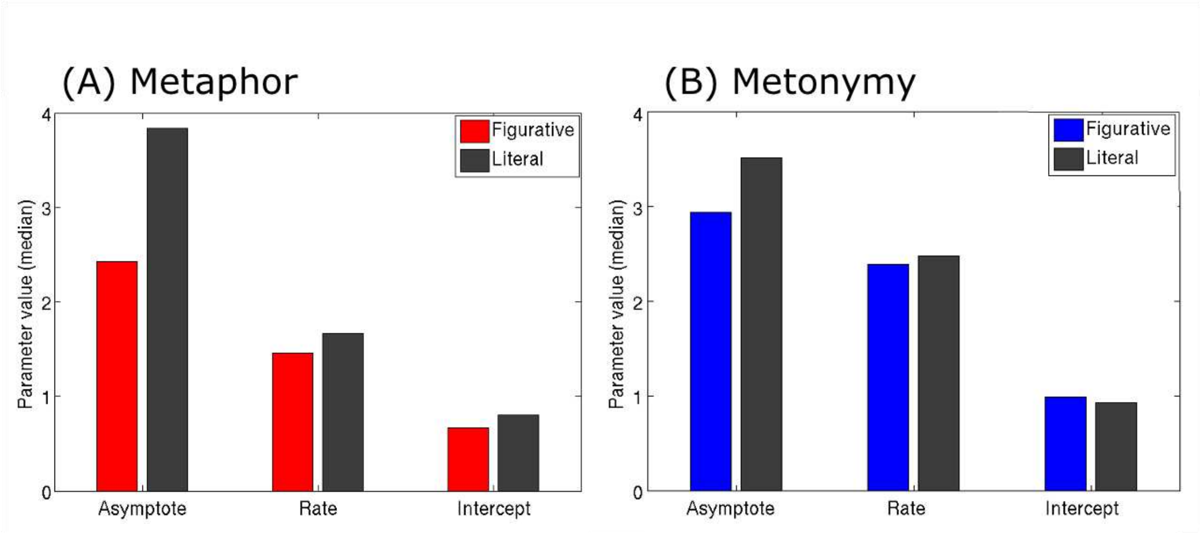

Figure 3. Exponential parameter estimates for metaphor and metonymy

Median exponential parameter estimates (asymptote, rate, intercept) of the fully parameterized model are represented (in panel A: red = metaphorical condition; gray = literal condition; in panel B: blue = metonymic condition; black = literal condition).