available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

Comprehending conventional and novel metaphors: An ERP study

Vicky Tzuyin Lai[a,b,*], Tim Curran[a,c], Lise Menn[a,b]

[a]Institute of Cognitive Science, University of Colorado, 344 UCB, Boulder CO 80309, USA
[b]Department of Linguistics, University of Colorado, 295 UCB, Boulder CO 80309, USA
[c]Department of Psychology and Neuroscience, University of Colorado, 345 UCB, Boulder CO 80309, USA

ARTICLE INFO

Article history:
Accepted 26 May 2009
Available online 6 June 2009

Keywords:
Metaphor
Sentence processing
Figurative language
Semantic integration
N400

ABSTRACT

The neural mechanisms underlying the processing of conventional and novel conceptual metaphorical sentences were examined with event-related potentials (ERPs). Conventional metaphors were created based on the Contemporary Theory of Metaphor and were operationally defined as familiar and readily interpretable. Novel metaphors were unfamiliar and harder to interpret. Using a sensicality judgment task, we compared ERPs elicited by the same target word when it was used to end anomalous, novel metaphorical, conventional metaphorical and literal sentences. Amplitudes of the N400 ERP component (320–440 ms) were more negative for anomalous sentences, novel metaphors, and conventional metaphors compared with literal sentences. Within a later window (440–560 ms), ERPs associated with conventional metaphors converged to the same level as literal sentences while the novel metaphors stayed anomalous throughout. The reported results were compatible with models assuming an initial stage for metaphor mappings from one concept to another and that these mappings are cognitively taxing.

1. Introduction

How do people understand linguistic expressions such as "*His idea was half-baked*", "*That class gave me some food for thought*", and "*The teacher spoon-fed them the information*"? The Contemporary Theory of Metaphor (CTM, Lakoff and Johnson, 1980; Lakoff and Turner, 1987; Lakoff, 1993) suggests that these expressions are surface realizations of an underlying conceptual metaphor IDEAS ARE FOOD[1], and are understood via a cross-domain conceptual mapping between IDEAS and FOOD. The mapping consists of a fixed set of ontological correspondences, such as "thinking is preparing", "communication is feeding", and "understanding is digesting". When those linguistic expressions are used, the conceptual mapping is activated so that IDEAS can be reasoned about in terms of FOOD. Based on the CTM, conceptual metaphors are important because they reflect how abstract concepts may be structured, and how abstract and concrete concepts are organized and interrelated in our minds.

The metaphor IDEAS ARE FOOD is conventional, i.e., in the fixed part of the conceptual system, in English. There can also be newly-coined examples which are not part of the conventional patterns of mappings. Novel metaphors are possible new ways of thinking, for example, "THEORIES ARE FATHERS" (Lakoff and Johnson, 1980, p53). Sentences derived from such made-up metaphors, such as "classical theories are patriarchs

* *Corresponding author*. Department of Linguistics, University of Colorado, 295 UCB, Boulder CO 80309, USA. Fax: +1 303 492 4416.
E-mail address: laiv@colorado.edu (V.T. Lai).

[1] We follow the convention in Cognitive Linguistics of using upper-case letters (e.g. IDEAS) to refer to concept names, and lower-case letters (e.g. idea) to refer to lexical items.

0006-8993/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2009.05.088

who father many children" can be interpretable or anomalous, depending on whether readers are able to arrive at some creative interpretation.

Numerous behavioral studies have examined whether metaphors and literal meanings are processed differently, and most of them have contrasted indirect access models (Grice, 1975; Searle, 1979) with direct access models (Gibbs, 1994, 2001; Glucksberg, 2003). The indirect access models suggest that metaphorical meanings deviate from literal meanings and cannot be computed until literal meanings are refuted by the context. The direct access models suggest that metaphorical meanings are as readily available as literal meanings. Though a few studies found longer reading times for metaphors than for literal statements and supported the indirect access view (Ortony et al., 1978; Janus and Bever, 1985), the majority favored the direct access view and showed that metaphors in context were read as quickly as literal statements (Gerrig and Healy, 1983; Glucksberg et al., 1982; Keysar, 1989; Blasko and Connine, 1993).

Based on the CTM, metaphors are different from literal expressions because of the conceptual mappings. However, the classic debate of direct vs. indirect access views does not make clear predictions as to whether those hypothesized conceptual mappings are activated during processing. On the one hand, the CTM seems to be in line with the indirect access view, because the metaphorical meanings are computed by accessing literal meanings first, and mapping those meanings from one domain to another domain. On the other hand, the CTM is consistent with the direct access view, because Lakoff suggested that conventionalized conceptual mappings are "used with no noticeable effort" (Lakoff, 1993, p.245) and should show processing demands equivalent to those for understanding literal sentences. Therefore, we turn to psycholinguistic models of metaphors, and contrast models that require mappings (Gentner and Wolff, 1997; Gentner and Bowdle, 2001; Glucksberg et al., 1997; Coulson and Matlock, 2001) with models that require no mapping (Giora, 1997, 1999, 2003; Giora and Fein, 1999; Frisson and Pickering, 2001).

The Structure Mapping model (Gentner and Wolff, 1997) proposes that metaphors act to set up correspondences between conceptual structures of the target and base concepts. The model proposes an initial stage of structural alignment between concepts, followed by inference importations from one concept to the other. For example, in "*men are wolves*", the target concept MEN and the base concept WOLVES are aligned by the predicate "prey on", and "*men prey on women*" is understood as "*wolves prey on little animals*". Gentner and Bowdle (2001) and Bowdle and Gentner (2005) further proposes the Career of Metaphor model, which suggests that conventional and novel metaphors are processed differently. Novel metaphors such as "*science is a glacier*" are understood as comparisons. As metaphors become more and more conventionalized, there is a shift in type of processing from comparison to categorization, i.e., assertions of category membership such as "*a robin is a bird*". What happens is that through repeated use, the base concept acquires a domain-general category in addition to its original domain-specific sense category. For example, the base concept "jail" in "*my job is a jail*" can be a literal category or a name for an ad-hoc, metaphoric category "*situations that are extremely unpleasant, confining and difficult to escape from*". When mapped from the literal category, the metaphor processing remains a form of comparison. When mapped from the ad hoc, metaphoric category, the metaphor is understood via a categorization process. Based on these models, both conventional and novel metaphors should be somewhat cognitively taxing due to an initial stage for structural alignment (Gentner, Bowdle, Wolff, and Boronat, 2001) that is needed for mappings. But novel metaphors should be more difficult than conventional metaphors due to always having to compare concepts and generate mappings on-line.

Giora (2003) proposed the Gradient Salience model, which suggests that it is not the conceptual mappings or metaphoricity, but the "saliency" of the linguistic expressions that determines whether those expressions can be understood rapidly. Salient meanings in this model refer to meanings foremost in speakers' minds at time of speaking, and are characterized by conventionality, prototypicality, familiarity, and frequency. Conventional expressions such as "*step on someone's shoes*" are salient and should be processed instantaneously like salient literal expressions. Novel metaphors such as "*her wedding ring is a 'sorry we're closed' sign*" are non-salient, and are slowed in processing due to having to reject the literal meanings of the phrases first. The model suggests that the comprehension speeds for salient conventional metaphors and literal statements should be the same. But conventional metaphors should be comprehended faster than non-salient, novel metaphors.

Event-related potentials (ERPs) can be effective in measuring processing effort from conceptual mappings. ERPs are synaptic potentials recorded from the scalp, which are then amplified and time-correlated with the cognitive event of interest. ERPs can be more sensitive than reaction times, as equivalent processing time does not necessarily represent equivalent cognitive effort (Coulson and Van Petten, 2002; Kutas et al., 2006). In addition, qualitative differences in the amplitudes, latencies, and topographies of ERPs can inform us more about the underlying cognitive processes in conventional and novel metaphor processing. One of the most established language-related ERP components is the N400, a negative-going wave starting at around 200–250 ms post stimulus onset and peaking at around 400 ms. The N400 was first observed in semantically incongruent sentences, such as "He spread the warm bread with socks" (Kutas and Hillyard, 1980), and has subsequently been found in the processing of meaningful stimuli such as words, non-words, music, and pictures (Bentin et al., 1985; Rugg and Nagy, 1987, Besson and Macar, 1987; Barrett and Rugg, 1990). Recently, the N400 has also been reported in metaphor studies (Pynte et al., 1996; Tartter et al., 2002; Coulson and Van Petten, 2002; Iakimova, et al., 2005; Arzouan et al., 2007).

Pynte et al. (1996) compared familiar vs. unfamiliar nominal metaphors in French (e.g. "*ces combatants sont des lions* (Those fighters are lions)" vs. "*ces apprentis sont des cruches* (Those apprentices are jars)"). They found that both familiar and unfamiliar metaphors elicited larger N400s than literal categorical statements (e.g. "Those animals are lions"). But in subsequent experiments, they found that regardless of metaphor familiarity, contextually appropriate metaphors elicited an N400 smaller than the contextually inappropriate

Table 1 – Example sentences and their source and target domains.

Sentence type	Sentences	Source	Target
Literal control	Every soldier in the frontline was attacked	WAR	WAR
	The path turned in a new direction	ROAD	ROAD
	That was too much food to digest	FOOD	FOOD
	The coffee you drank was warm	FIRE	FIRE
Conventional metaphor	Every point in my argument was attacked	WAR	ARGUMENT
	Her life has a new direction	ROAD	LIFE
	That was too much info to digest	FOOD	IDEA
	The love she gave was warm	FIRE	LOVE
Novel metaphor	Every second of our time was attacked	WAR	TIME
	Their style has a new direction	ROAD	FASHION
	That was too much love to digest	FOOD	LOVE
	The anger he felt was warm	FIRE	ANGER
Anomalous	Every drop of rain was attacked	WAR	WEATHER
	The lawn has a new direction	ROAD	GARDEN
	That was too much wind to digest	FOOD	WEATHER
	The answer they gave was warm	FIRE	WORDS

ones. They concluded by arguing for a context-driven account, which has no distinct processing stages or conceptual mappings. The account seems consistent with the Gradient Salience model because contextual appropriateness can be viewed as saliency.

Tartter et al. (2002) examined ERPs to sentence-final words in novel metaphorical, literal, and anomalous sentences (e.g. "his face was contorted by an angry cloud/frown/map"). They found that the waveform for the literal condition diverged from the other two conditions at 160 ms. The waveforms for the novel metaphorical and the anomalous conditions diverged at 280 ms. It was concluded that novel metaphorical sentences were understood as anomalous expressions momentarily, but were then meaningfully resolved in the 300–500 ms window. Our concerns with Tartter et al.'s findings are that first, the sentence-final words (i.e. "cloud", "frown", and "map") differ across conditions, which may have introduced confounding factors such as word frequencies. Secondly, the stimuli were a mixture of sentential metaphors and nominal metaphors (e.g. "the camel is a desert taxi" as a nominal metaphor). The processing of sentential and nominal metaphors may differ in that understanding the conceptual category name (e.g. "taxi") and the sentence terminal lexical item (e.g. "cloud") might require different comprehension strategies. Lastly, Kutas et al. (2006) pointed out that Tartter et al.'s results are complicated by the higher cloze probabilities for their literal sentences, given that more expected words are known to elicit smaller N400s.

Coulson and Van Petten (2002) proposed that both metaphor and literal sentences require the same process of evaluating and selecting properties in the concepts involved. For example, to understand the metaphorical sentence "after giving it some thought, I realized the new idea was a gem", the concepts "idea" and "gem" needed to be analyzed in more detail to achieve interpretation. To understand the literal match of the metaphorical sentence, such as "the ring was made of tin, with a pebble instead of a gem", the mention of "pebble" caused the concept "gem" to be analyzed more fully, which required deep analytical processing similar to that in metaphor processing. But to understand a general literal sentence, such as "that stone we saw in the natural history museum is a gem", the concept "gem" did not need to undergo as much analysis. Coulson and Van Petten's proposal was supported by their findings that metaphorical sentences elicited an N400 more negative than the literal-matched ones, and that both had N400s more negative than the literal ones. Of crucial importance for the present study, some cognitive effort was needed for understanding metaphors compared with literal sentences, possibly for structural alignment and property importations as proposed in the Structure Mapping model.

However, after examining the 10 stimulus items they provided (Table 1, Coulson and Van Petten ,2002), we found that 3 of 10 examples had double metaphors and 4 of 10 were unconventional. Double metaphors coming from two very different metaphor domains might have made their metaphor condition more difficult. For example, in the sentence "The independent prosecutor thought he was a bulldog, but he was really more of a flea", there are "bulldog" and "flea" personifications. Personifications were actually treated as novel metaphors in CTM (Lakoff 1993). Secondly, the stimulus items varied in conventionality. For example, "He knows that power is a strong intoxicant" is conventional to native ears while "My crazy uncle says jokes are conversation's cayenne" is unconventional for describing "jokes". Unconventional items might increase the N400 amplitude while conventional ones might reduce the amplitude in their metaphor condition. Therefore, the multiple metaphors and the variations in conventionality might have confounded the observed N400.

Iakimova et al. (2005) examined highly conventionalized 'dictionary metaphors' in French (e.g. "*Il est parti dans les nuages* (he is away in the clouds)") in people with schizophrenia and non-patients. They found that these dictionary metaphors were not more difficult to process than the literals in non-patients. Arzouan et al. (2007) examined conventional metaphorical word pairs (e.g. "lucid mind") and novel ones (e.g., "conscience storm") in Hebrew. They found that novel metaphorical word pairs were more difficult to process than conventional ones, and both were more difficult than related word pairs (e.g. "burning fire"). They suggested that novel and conventional metaphors appeared to be accessed similarly, but differ in terms of processing difficulty. Both Iakimova et al.'s and Arzouan et al.'s results are consistent with the Gradient Salience model.

In summary, ERP studies on metaphors have produced inconsistent results. Some studies found that metaphors

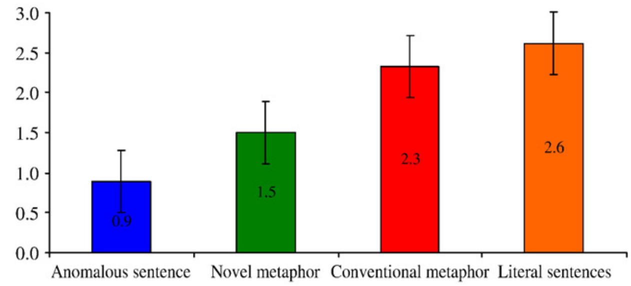

Fig. 1 – Mean sensicality judgments (perfect sense = 3, some sense = 2, little sense = 1, and no sense = 0) for anomalous sentences, novel metaphors, conventional metaphors, and literal sentences. Error bars indicate the standard errors of the condition difference.

elicited brainwaves indexing more cognitive effort than the literals (Tartter et al., 2002, Coulson and Van Petten, 2002, Arzouan et al., 2007). Others found that metaphors were no more difficult to process than the literals (Pynte et al., 1996, Iakimova et al., 2005). We argued that the inconsistency may have resulted from a failure to keep the linguistic forms of metaphor stimuli consistent (e.g. nominal vs. sentential metaphors) and to distinguish among metaphors with varying degrees of conventionality (e.g. unfamiliar metaphors, dictionary metaphors).

The present study creates conventional and novel metaphorical sentences based on the CTM and with reference to the examples available on the Conceptual Metaphor Home Page[2]. For example, we extracted a conventional metaphorical sentence "every point in my argument was attacked" from the conventional metaphor "ARGUMENT IS WAR". We created novel metaphors such as "TIME IS WAR", and then created examples that matched the syntactic structure and the final word of the conventional one, such as "every second of my time was attacked". Anomalous sentences and literal sentences were created e.g., "every drop of rain was attacked" and "every soldier in the frontline was attacked" for comparison. To ensure that conventional metaphorical expressions are familiar and interpretable, and that novel ones are unfamiliar but interpretable, we conducted pretests of familiarity and interpretability. The example stimuli are illustrated in Table 1. See Experimental procedures section for pretest details.

Our predictions are as follows. If the Gradient Salience model holds, there should be no N400 difference between the conventional and literal conditions because they are equally salient. The N400s should be more negative in the novel metaphor condition than the conventional one, because novel metaphors are less salient than conventional ones. If the Structure Mapping/Career of Metaphor models hold, then both conventional and novel metaphorical expressions should show some N400 due to an initial stage of structural alignment for conceptual mappings.

[2] Conceptual Metaphor Home Page (http://cogsci.berkeley.edu/lakoff/).

2. Results

2.1. Behavioral results

The sensicality ratings showed that during real-time comprehension, subjects found that conventional metaphors made more sense than novel metaphors, and both made more sense than the anomalous condition. There is a main effect between sentence types [$F(3,69) = 336.1$, $p < .0005$] (see Fig. 1). Pairwise comparisons confirmed that each condition differed from all others (conventional vs. literal [$F(1,69) = 21.948$, $p < .0005$], conventional vs. anomalous [$F(1,69) = 563.592$, $p < .0005$], novel vs. literal [$F(1,69) = 335.319$, $p < .0005$], novel vs. anomalous [$F(1,69) = 102.277$, $p < .0005$], literal vs. anomalous [$F(1,69) = 807.978$, $p < .0005$], conventional vs. novel [$F(1,69) = 185.692$, $p < .0005$]).

The reaction times for the sensicality rating task were noticeably longer than those in other metaphor research, because participants were specifically told to delay their responses until after 700 ms (200 ms word presentation +500 ms dark screen) to suspend movement. Data were analyzed with repeated measures of ANOVA and a main effect between conditions was found [$F(3, 69) = 35.5$, $p < .0005$] (see Fig. 2). Pairwise comparisons indicated no differences between the anomalous (1615 ms) and novel (1620 ms) conditions, which were each slower than conventional (1487 ms) (conventional vs. anomalous [$F(1, 69) = 21.491$, $p < .0005$], conventional vs. novel [$F(1, 69) = 23.437$, $p < .0005$]) and literal (1377 ms) (literal vs. anomalous [$F(1, 69) = 24.336$, $p < .0005$], literal vs. novel [$F(1, 69) = 25.479$, $p < .0005$]) conditions, which differed from each other (conventional vs. literal [$F(1, 69) = 15.825$, $p < .0005$]).

2.2. Event-related potentials

The mean for the numbers of trials per condition per block per subject included in the ERP analysis was 25 (range: 19–26; median: 25), resulting in approximately 400 out of the 416 stimulus sentences (i.e. 100 sentences for each one of the 4 conditions) in the ERP analysis. Grand averaged waveforms for the sentence-final words in each of the four conditions at 63 electrode sites grouped into 9 groups (left anterior, anterior midline, right anterior, left central, central midline, right

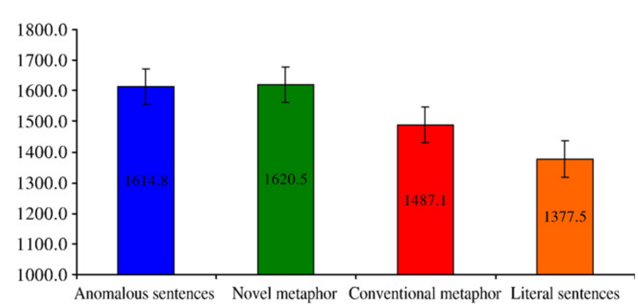

Fig. 2 – Mean reaction times (from the onset of the target word to the time when a sensicality judgment was made) for anomalous sentences, novel metaphors, conventional metaphors, and literal sentences. Error bars indicate the standard errors of the condition difference.

central, left posterior, posterior midline, and right posterior) are displayed in Fig. 3. The early perceptual components of N100 and P200 could be observed clearly at the frontal and central sites; they were followed by a negative deflection starting at 320 ms, peaking at around 370 ms at all scalp sites, which was identified as the N400.

Visual inspection indicated that the waveform for the conventional metaphor condition diverged from the anomalous and novel metaphor conditions at around 440 ms and then converged with the literal condition at around 560 ms. Waveforms for the novel metaphors stay in line with the anomalous condition throughout. Therefore, we considered the activities over the 440–560 ms window separately from the 320–440 ms window.

Mean amplitudes from the 63 electrode sites for each condition in each block for both time windows were entered into a 2 time × 4 condition × 4 block × 3 left/middle/right location × 3 anterior/central/posterior location repeated measures of ANOVA. The Greenhouse–Geisser (Greenhouse and Geisser, 1959) sphericity correction was applied. There is no significant interaction between condition × block [$F(9, 207) = .727$, $p = .63$] or between time × condition × block [$F(9,207) = 1.802$, $p = .11$], so repeating the target words across blocks had little effect. Results confirmed that the patterns of conditions in the two time windows differ significantly: time × condition interaction [$F(3,69) = 9.738$, $p < .0005$]. Furthermore, the time × condition × left/middle/right location interaction [$F(6, 138) = 6.529$, $p < .0005$] indicated that the topographic distribution of the condition effects differed across the two times. The scalp distributions of the effects in the two time windows are displayed in Fig. 4 by subtracting the literal condition from each of the other conditions. In the 320–440 ms window, anomalous, novel, and conventional differences were all of similar magnitude and similarly distributed over the central scalp. In the later window, only the anomalous and novel conditions differed from literal, as statistically supported below.

Because of the time × condition × location interaction, separate ANOVAs were conducted within each time window to better understand each of these effects separately. Only effects and interactions involving the condition factor are of interest, so only these are reported here.

Results from the earlier window (Fig. 5, left) yielded a significant effect of condition [$F(3,69) = 8.677$, $p < .0005$] that interacted with left/middle/right location [$F(6,138) = 3.104$, $p < .05$] and with anterior/central/posterior location [$F(6,138) =$

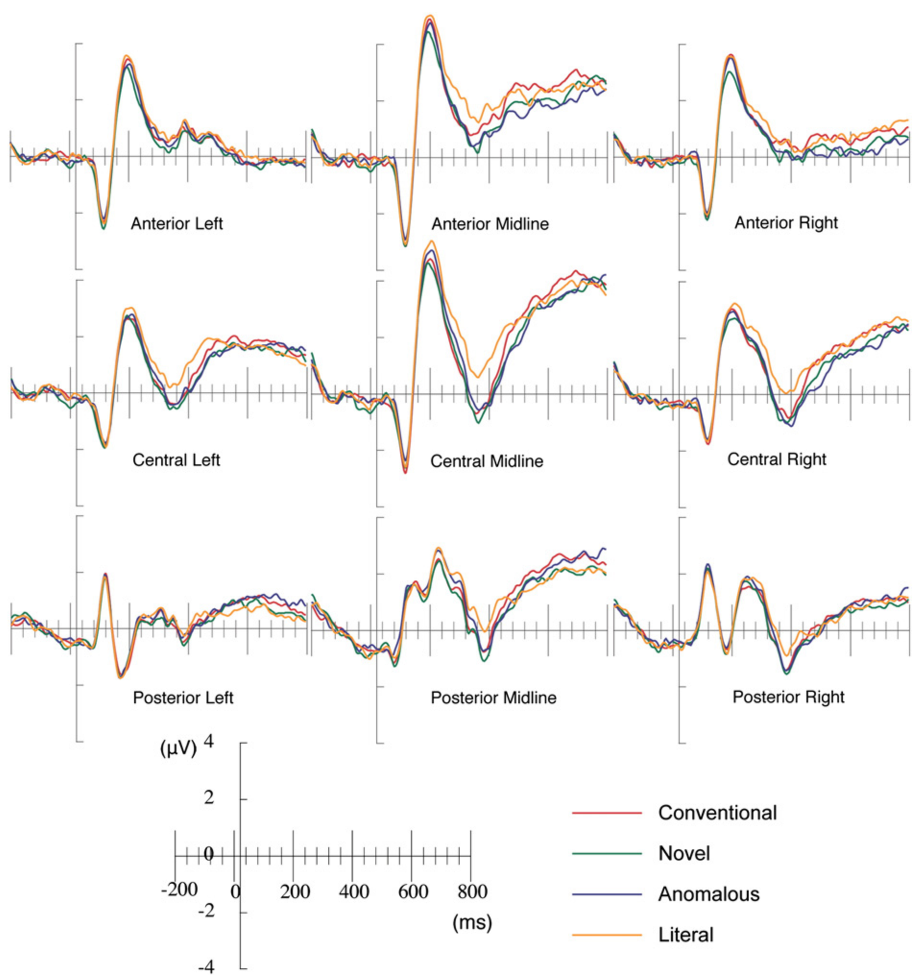

Fig. 3 – Grand average ERP waveforms recorded at 63 electrode sites grouped into 9 groups (left anterior, anterior midline, right anterior, left central, central midline, right central, left posterior, posterior midline, and right posterior sites).

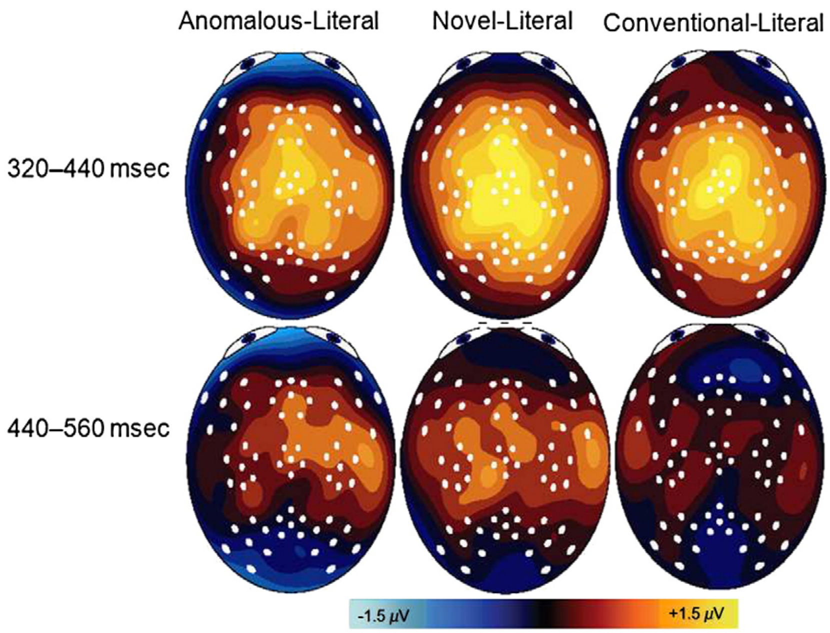

Fig. 4 – Topographic plots of anomalous-literal, conventional-literal, and novel-literal differences for the earlier (320-440 ms) window (top row) and the later (440-560 ms) window (bottom row). The white dots are the 9 groups of electrodes for statistical analyses. Note that cluster sizes are approximately equal, but inferior clusters appear larger in the maps because of the 3D to 2D projection.

3.228, $p<.05$]. Post hoc mean comparisons showed that amplitudes for conventional and novel metaphors do not differ from the anomalous condition [$F(1, 23)=1.27$, $p=.26$; $F(1, 23)=.57$, $p=.44$]; and that conventional and novel metaphors do not differ from each other [$F(1,23)=3.542$, $p=.07$]. The novel, conventional, and anomalous conditions each differed from the literal conditions (novel vs. literal: $F(1,23)=22.536$, $p<.0005$; conventional vs. literal: $F(1,23)=8.209$, $p<.01$; anomalous vs. literal: $F(1,23)=15.937$, $p<.0005$;) The condition by location interaction indicates that differences were greatest near midline centroparietal locations, as is typical of the N400 (see Figs. 3–4).

Results from the later window (Fig. 5, right) also yielded a main effect between conditions ($F(3,69)=5.003$, $p<.01$), but no condition by location interaction (condition vs. left/middle/right location: $F(6,138)=1.948$, $p=.11$; condition vs. location: $F(6,138)=1.922$, $p=.13$). Unlike the earlier window, post hoc mean comparisons between conditions showed that the waveforms for conventional metaphors and literal sentences have converged [$F(1,23)=1.257$, $p=.27$] while the waveform for novel metaphors remained equivalent to anomalous sentences [$F(1,23)=1.027$, $p=.31$]. Conventional and novel metaphors differ significantly [$F(1,23)=6.636$, $p<.05$].

An analysis of cloze probability was carried out to demonstrate that the effect was not driven by cloze probabilities of the items. 146 additional participants were asked to do a cloze test by completing each one of the 416 sentence frames with the first word that came to mind. The literal sentences had higher cloze probabilities (0.09) than the conventional metaphors (0.04), the novel metaphor metaphors (0.02) and anomalous

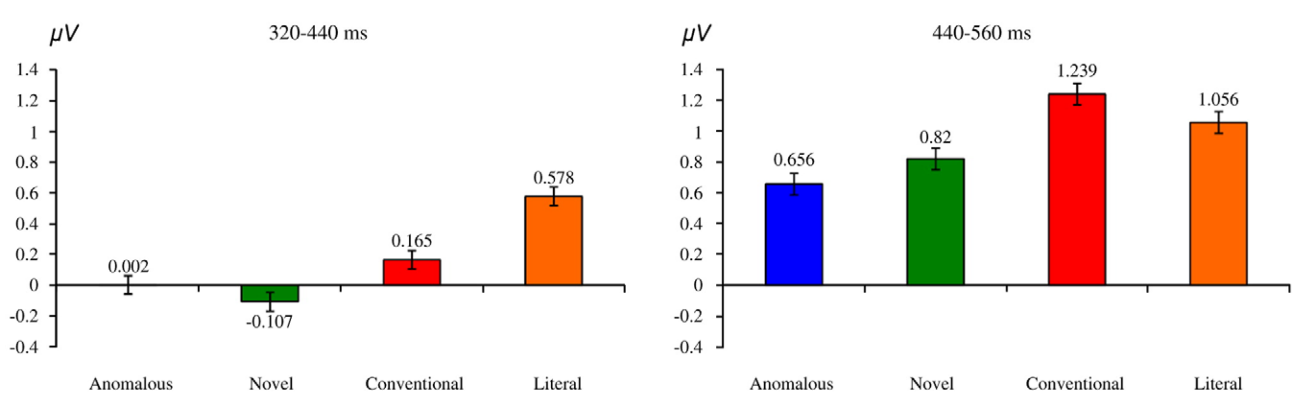

Fig. 5 – Comparison of means for the earlier window (320–440 ms) (left) and the later window (440–560 ms) (right).

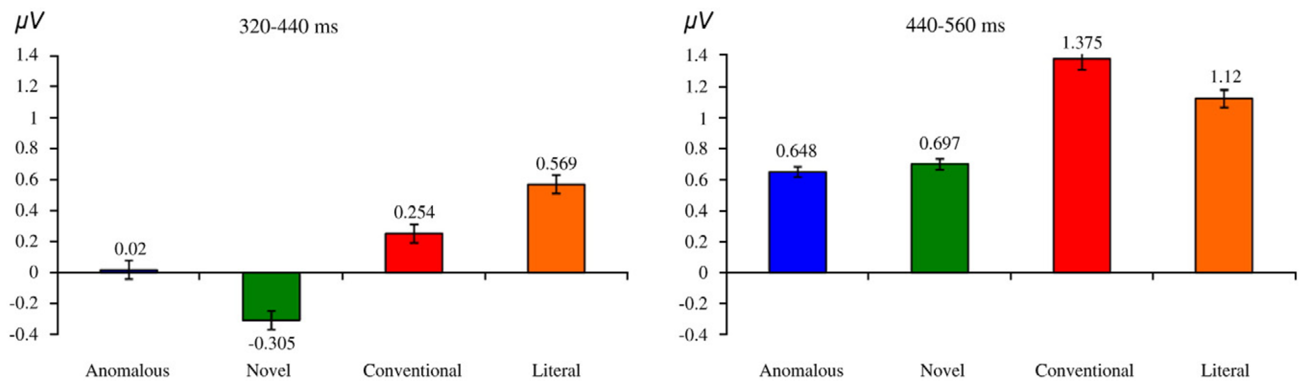

Fig. 6 – Comparison of means after controlling for cloze probabilities for the earlier window (320–440 ms) (left) and the later window (440–560 ms) (right).

conditions (0.01). To verify whether this difference was reflected in the N400 analysis, we excluded items that received higher cloze probabilities and conducted the same analysis previously comparing early vs. late windows. The exclusion criterion was that if more than one of the participants completed the sentence frames with the target sentence-final words, then that sentence was excluded. The remaining items (368 sentences) were submitted for N400 analysis. The results showed exactly the same pattern. Fig. 6 is almost identical with Fig. 5: in the early window, the anomalous, novel metaphorical, and conventional metaphorical conditions were similar to each other, and all differed from the literal condition. In the late window, the conventional metaphors converged with the literal condition, while the novel metaphors remained the same as the anomalous sentences.

As suggested by a reviewer, we sorted the ERPs according to subjects' sensicality ratings. We contrasted novel metaphors that received lower sensicality ratings (0 and 1 on a scale from 0 to 3) and those that received higher (2 and 3). It was found that the mean amplitudes of the two sets do not differ statistically from each other in either the entire 320–560 ms window [$F(1, 23) = .598$, $p = .45$], the early 320–440 ms window [$F(1, 23) = .197$, $p = .66$], or the late 440–560 ms [$F(1, 23) = .888$, $p = .36$]. To be cautious, we excluded two subjects who did not rate many novel metaphors as making some sense or making perfect sense, and recomputed the analyses in the three time windows. The results showed that still, the two groups did not differ statistically from each other in either the entire 320–560 ms window [$F(1, 21) = .627$, $p = .44$], the early 320–440 ms window [$F(1, 21) = .712$, $p = .41$], or the late 440–560 ms window [$F(1, 21) = 1.798$, $p = .19$].

3. General discussion

The main finding of the study was that while both conventional metaphors and literal sentences were rated similarly as familiar and interpretable, the ERP results showed that the conventional metaphors required a short burst of additional processing effort when compared with literal sentences. Novel metaphors, which were rated unfamiliar and less interpretable, required a more sustained effort, similar to the effort observed in anomalous sentences, which were rated unfamiliar and least interpretable.

We intended to use the N400 as an index of semantic processing in metaphor comprehension, but identification of the N400 component was complicated by changes in the pattern of experimental effects and topographic distribution across the 300–500 ms interval typically assigned to the N400. In the earlier window, ERP amplitudes were more negative for anomalous sentences, novel metaphors, and conventional metaphors as compared with literal sentences. In the later window, conventional metaphors converged to the same negativity level as literal sentences. We can confidently associate the earlier window to the N400 because it shows the standard difference between anomalous and literal conditions, the window encompasses the peak of the negativity to anomalous sentences, and the topography of the effects conform to the typically observed centro-parietal distribution.

The identification of the later window is a more open question. The later window may reflect the activity of a distinct, later occurring process — as would be consistent with the different pattern of effects and significantly different topographic distributions. Alternatively, the later window may reflect a continuation of the processes underlying the N400, with condition differences reflecting the accrual of additional information over time and topographic differences reflecting summation with other later-occurring processes. Critically, the interpretation of our results does not depend on distinguishing these alternatives.

Our findings replicated Coulson and Van Petten (2002), who found that metaphors (novel or conventional) elicited N400 more negative than literal sentences. We partially replicated Tartter et al. (2002) in that their novel metaphors were perceived as being anomalous momentarily, but our novel metaphors continued to be perceived as being anomalous. We also partially replicated Azouan et al. (2008) in that we showed an effort for metaphors with varying degrees of conventionality in terms of the early vs. late difference, whereas they showed such difference in the gradient of their N400 amplitudes. In contrast to Iakimova et al. (2005), who found no N400 effect for the dictionary metaphors, we found enhanced N400s for our conventional metaphors. We interpret this by suggesting that dictionary metaphors may well be like the dead metaphors of Gentner and Bowdle (2001): "any expression using the metaphoric sense of the base term is a dead metaphor and will not seem metaphoric". In other words, the dictionary can be viewed as the graveyard of metaphors (Boroditsky, personal communication). Lastly, we found that conventionality matters, whereas Pynte et al. (1996) found that the familiarity, which in our view is one of the variables for characterizing conventionality, does not influence N400 amplitudes.

The current study seems to be consistent with the indirect access view, at first sight. The conventional metaphorical and literal conditions differ in their N400 amplitude, indicating a difference in processing effort. However, this difference could be interpreted in two ways: Either the conventional metaphorical sentences were difficult because the system was busy rejecting the first available literal meaning and retrieving the appropriate metaphorical meanings, or because it was selecting among multiple meanings that were all retrieved at the same time. In the former case, the literal meaning was available first, and hence would be supporting the indirect access view. In the latter case, all multiple meanings are available as in the classic exhaustive access view (Onifer and Swinney, 1981, Seidenberg, Tanenhaus, Leiman, and Bienkowski, 1982), which would be consistent with the direct access view.

Most crucially, are there conceptual mappings as hypothesized by the CTM and are those mappings cognitively taxing? Our data do not support models that require no conceptual mappings. The Gradient Salience model is not supported, because conventional metaphors that participants had judged to be as interpretable and as familiar as literal sentences, and which were therefore operationally "salient", still elicited the same size N400 in the earlier window as novel metaphors which participants had judged to be non-salient. Novel metaphors that participants had judged to be more interpretable and sensible than the anomalous sentences, and therefore more salient than the anomalous, still showed the same size N400 as the anomalous throughout the early and late windows. These results support models in which conceptual mappings are in use to some extent during metaphor comprehension. The Structure Mapping model is supported, because our conventional metaphors still needed more effort to process than literal sentences, consistent with Gentner et al.'s (2001) claim that even highly conventionalized metaphors required an initial stage for structural alignment. In addition, consistent with the Career of Metaphor model, understanding novel metaphors is harder than understanding conventional metaphors, because novel ways of thinking require comparing the concepts and creating conceptual mappings on the spot.

The Career of Metaphor model also proposes a comparison process for novel metaphors and a comparison/categorization process for conventional metaphors. The current finding can only establish that there are differences between conventional and novel metaphor processing, and between literal and metaphor processing. Our results cannot clearly distinguish which type of process underlies those differences, but the finding that the conventional metaphors differ from the literal sentences implies a comparison process for conventional metaphors, which requires mapping to the target category from a literal base category rather than from a metaphoric one.

The difference in N400 amplitude between literal sentences and conventional metaphors addresses an additional issue of what it is that the N400 indexes (see Lau et al., 2008 for a detailed review). Some researchers have suggested that N400 reflects an ease of lexical meaning retrieval from memory (Van Petten et al., 1999; Kutas and Federmeier, 2000; Van Berkum, 2008, in press; Coulson and Federmeier, in press) while others have suggested that the N400 reflects the integration of lexical meaning with context at a post-lexical stage (Brown and Hagoort, 1993; Chwilla et al., 1995; Hagoort et al., 2004). If we define the earlier window as the pure N400, then our results are inconsistent with the integration account, because conventional metaphors should not require substantial effort to integrate with the sentential context. However, if we define the earlier and the later window together as the N400 window, then the waveform for conventional metaphors converges to that of the literal sentences, as would be expected by the integration account.

Another way to view our results is that if the earlier window is representative of the meaning retrieval process, and the later window, the meaning integration process, then our results would suggest that at the retrieval stage, it requires effort to retrieve both conventional and novel metaphorical meanings. In the integration stage, only novel metaphorical meaning is difficult while conventional metaphorical meaning is easily integrated with the rest of the context. Further studies and elaborations of processing models will be needed in order to determine the best interpretation of these data.

In conclusion, we observed several ERP differences in the processing dynamics between literal, conventional and novel metaphorical sentences. Our findings differentiating conventional and novel metaphors supported an indirect access processing model compatible with the Career of Metaphor Theory.

4. Experimental procedures

4.1. Participants

Twenty-nine right-handed native English speakers (19 men, 10 women, average age 19.7) in the University of Colorado in Boulder participated in this experiment for course credit. None had any neurological disorder or major head injury that was diagnosed as having a long-term side effect. Data were discarded from 5 subjects who had less than 15 acceptable trials per condition per block due to blinking.

4.2. Stimuli

416 sentences (104 quadrates with 4 sentence types in each quadrate) were created by two linguists (the first and third authors) with reference to the CTM as described in the Introduction. Conventional metaphors were more familiar and interpretable. Novel metaphors were less familiar, but still interpretable. Anomalous sentences were unfamiliar and least interpretable. Within a given quadrate (see Table 1 for examples), the same target word was used in all conventional,

novel, anomalous, and literal conditions. Across the 104 quadrates, the target words were comprised of 42 verbs, 41 adjectives, and 21 nouns. The mean length of sentences was 6.7 words with a standard deviation of 1.4.

To check whether the stimuli were familiar and interpretable, two pretests were conducted (Blasko and Connine, 1993; Titone and Connine, 1994; Budiu and Anderson, 2002). Thirty-eight native speakers of English from two undergraduate classes in the Department of Linguistics at the University of Colorado, Boulder, volunteered for participation. Using Latin Square design, the 416 sentences (4 conditions/sentence frames × 104 target words) were divided into 4 blocks so that each target word appeared only once in each block. As a result, each block contained 26 conventional metaphors, 26 novel metaphors, 26 anomalous sentences, and 26 literal sentences. Sentences in each block were then randomized. A given participant saw all four blocks during the course of the pre-test.

After participants signed the informed consents, they were instructed to rate each sentences on two scales from 0 to 3, first for the familiarity, and then for interpretability. The instructions for the familiarity scale were: "If you have heard similar expressions *frequently* before and feel that the meaning is *highly familiar*, give it a 3. If you have heard similar expressions *occasionally* before and feel that the meaning is *somewhat familiar*, give it a 2. If you have heard similar expressions *once or twice* before and feel the meaning is *somewhat unfamiliar*, give it a 1. If you have *never* heard it before and feel that the meaning is *unfamiliar*, give it 0." For the interpretability scale, the instructions were "If you feel that the sentence is *easily interpretable*, give it a 3. If the sentence *takes you a while to come up with an interpretation*, give it a 2. If the sentence *takes you a long while to come up with an interpretation*, give it a 1. If you *couldn't think of an interpretation* that would make sense of the sentence, it's 0."

Repeated measures analysis of variance (ANOVA) confirmed that familiarity ratings were significantly different between sentence types [$F(3,18) = 256.3$, $p < .0005$] (see Fig. 7, left). Pairwise comparisons indicated no difference between conventional metaphors and literal sentences (conventional vs. literal [$F(1,18) = 3.7$, $p = 0.1$]), which were both more familiar than the novel metaphors (conventional vs. novel [$F(1,18) = 71.3$, $p < .0005$], literal vs. novel [$F(1,18) = 74.8$, $p < .0005$]), and the anomalous sentences (conventional vs. anomalous [$F(1,18) = 234.3$, $p < .0005$], literal vs. anomalous [$F(1,18) = 240.5$, $p < .0005$]), which differed from each other (novel vs. anomalous [$F(1,18) = 61.3$, $p < .0005$]). The interpretability ratings were significantly different between sentence types [$F(3,18) = 111.3$, $p < .0005$] (see Fig. 7, right). Similar to the familiarity results, pairwise comparisons of the interpretability results indicated no difference between conventional metaphors and literal sentences (conventional vs. literal [$F(1,18) = .04$, $p = 0.8$]), which were both more interpretable than novel metaphors (conventional vs. novel [$F(1,18) = 71.3$, $p < .0005$], literal vs. novel [$F(1,18) = 74.8$, $p < .0005$]) and the anomalous sentences (conventional vs. anomalous [$F(1,18) = 234.3$, $p < .0005$], literal vs. anomalous [$F(1,18) = 240.5$, $p < .0005$]), which differed from each other (novel vs. anomalous [$F(1,18) = 47.1$, $p < .001$]).

In the ERP experiment, each participant saw 104 anomalous sentences, 104 novel metaphors, 104 conventional metaphors, and 104 literal sentences, split into 4 blocks. As in the pre-tests, a Latin Square design was employed and a particular sentence frame/condition and target word appeared together only once in each block. As a result, each block contained 26 conventional metaphors, 26 novel metaphors, 26 anomalous sentences, and 26 literal sentences. Sentences in each block were then randomized.

4.3. *Procedure*

Participants first completed a consent form, followed by Sensor Net setup, a brief practice session, and then the main experiment in a quiet room with white noise in the background and dim light. Sensor Net setup took about 20–30 min, including placing the net on the subject's head, positioning sensors, and adjusting/rewetting sensors to reach desired impedance levels of less than 40 kΩ.

We partially replicated the paradigm of Coulson and Van Petten (2002) in stimulus presentation. Each word in each sentence was presented for 200 ms with a length-dependent interword interval: 100 ms plus an additional 37 ms for each character in the previous word. At the offset of the sentence-final target word, a dark screen was presented for 500 ms before a question mark "?" appeared. Upon seeing the question mark, participants were to judge how much sense the sentences make in English by pressing either one of the

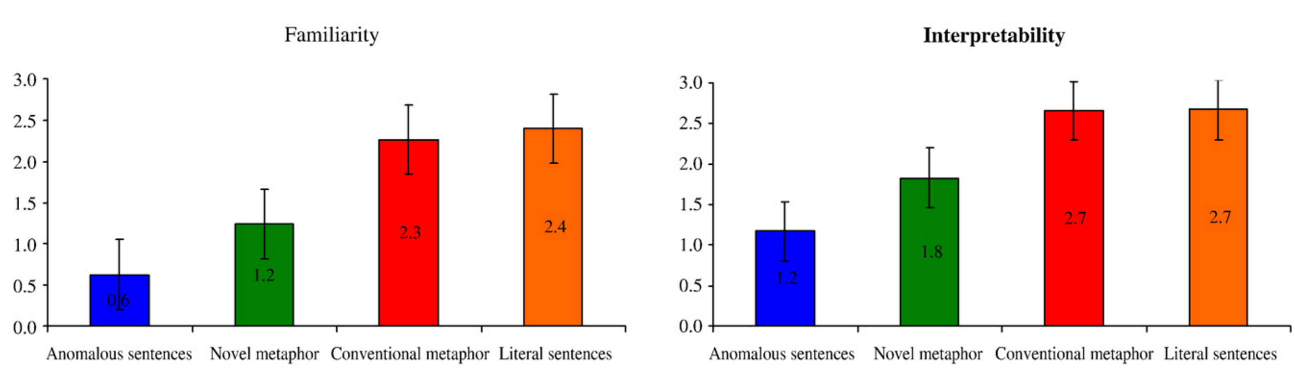

Fig. 7 – Familiarity ratings (left) and interpretability ratings (right) on a scale from 0 to 3 for anomalous sentences, novel metaphors, conventional metaphors, and literal sentences. Error bars indicate the standard errors of the condition difference.

four keys (1, 2, 3, and 4) on the response pad (i.e., *perfect sense*, *some sense*, *little sense*, and *no sense*). Participants responded with the index and middle fingers of both hands. The order of the key assignments (left to right vs. right to left) and the order of the four blocks were counterbalanced across subjects. Once a response was made, the program moved on to the next trial. Instead of having a comprehension task right after each sentence like Coulson and Van Petten (2002), at the end of the EEG session, the sensor net was taken off and participants were asked to do an interpretation generation post-test.

The purpose of the interpretation generation task was to ensure that novel metaphors were all interpretable. In this half-hour task, participants were given all 104 novel metaphorical sentences. Each sentence was presented entirely at once and participants were asked to type in what they thought the sentence meant within 20 s. At 20 s, the program automatically switched to the next trial. The results from the interpretation generation task suggested that all novel metaphors were interpretable, but no formal analysis was conducted due to the qualitative nature of these data.

4.4. *Electrophysiological recording*

Scalp voltages were collected with a 128-channel HydroCel Geodesic Sensor Net™ connected to an AC coupled, 128-channel, high-input impedance amplifier (200 MΩ, Net AmpsTM, Electrical Geodesics Inc., Eugene, OR). Amplified analog voltages (0.1–100 Hz bandpass) were digitized at 250 Hz. Individual sensors were adjusted until impedances were less than 40 kΩ. The EEG was digitally low-pass filtered at 40 Hz. Trials were discarded from analyses if more than 20% of channels were bad (average amplitude over 100 μV or transit amplitude over 50 μV). Eye movements were corrected with an ocular artifact correction algorithm (Gratton, Coles, and Donchin, 1983). Individual bad channels were replaced on a trial-by-trial basis with a spherical spline algorithm (Srinivasan et al., 1996). EEG was measured with respect to a vertex reference (Cz), but transformed to an average mastoids reference for analysis. Event-related potentials (ERP) were obtained by stimulus-locked averaging of the EEG recorded in each condition. ERPs were baseline-corrected with respect to a 200-ms pre-stimulus recording interval.

Acknowledgments

We thank Dr. Albert Kim and Dr. Jos van Berkum for very helpful feedback on a previous version of the manuscript. We also thank Brion Woroch, Casey DeBuse, Brent Young, and Sarah Jirkovsky for help with participant testing.

REFERENCES

Arzouan, Y., Goldstein, A., Faust, M., 2007. Brainwaves are stethoscopes: ERP correlates of novel metaphor comprehension. Brain Res. 36, 222–231.

Barrett, S.E., Rugg, M.D., 1990. Event-related potentials and the phonological matching of pictures. Brain Lang. 38, 424–437.

Bentin, S., McCarthy, G., Wood, C.C., 1985. Event-related potentials, lexical decision, and semantic priming. Electroencephalogr. Clin. Neurophysiol. 60 (4), 343–355.

Besson, M., Macar, F., 1987. An event-related potential analysis of incongruity in music and other non-linguistic contexts. Psychophysiol. 24, 14–25.

Blasko, D.G., Connine, C.M., 1993. Effects of familiarity and aptness on metaphor processing. J. Exper. Psychol., Learn., Mem., Cogn. 19, 295–308.

Bowdle, B., Gentner, D., 2005. The career of metaphor. Psychol. Rev. 112, 193–216.

Brown, C., Hagoort, P., 1993. The processing nature of the N400: evidence from masked priming. J. Cogn. Nerosci. 5, 34–44.

Budiu, R., Anderson, J., 2002. Comprehending anaphoric metaphors. Mem. Cogn. 30, 158–165.

Chwilla, D.J., Brown, C.M., Hagoort, P., 1995. The N400 as a function of the level of processing. J. Psychophysiol. 32, 274–285.

Coulson, S., Matlock, T., 2001. Metaphor and the space structuring model. Metaphor. Symbol 16, 295–316.

Coulson, S., Federmeier, K. D., in press. Words in context: ERPs and the lexical/postlexical distinction. J. Psycholinguist Res.

Coulson, S., Van Petten, C., 2002. Conceptual integration and metaphor: an event-related potential study. Mem. Cogn. 30, 958–968.

Frisson, S., Pickering, M.J., 2001. Obtaining a figurative interpretation of a word: support for underspecification. Metaphor Symbol 16, 149–171.

Gentner, D., Wolff, P., 1997. Alignment in the processing of metaphor. J. Mem. Lang. 37, 331–355.

Gentner, D., Bowdle, B.F., 2001. Convention, form, and figurative language processing. Metaphor Symbol 16, 223–247.

Gentner, D., Bowdle, B.F., Wolff, P., Boronat, C., 2001. Metaphor is like analogy. In: Gentner, D., Holyoak, K.J., Kokinov, B.N. (Eds.), The analogical mind: Perspectives from cognitive science. MIT Press, Cambridge, MA, pp. 199–253.

Gerrig, R., Healy, A., 1983. Dual processes in metaphor understanding: comprehension and appreciation. Mem. Cogn. 9, 667–675.

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

Gibbs, R.W., 2001. Evaluating contemporary models of figurative language understanding. Metaphor Symbol 16, 317–333.

Giora, R., 1997. Understanding figurative and literal language: the graded salience hypothesis. Cogn. Linguist. 8, 183–206.

Giora, R., 1999. On the priority of salient meanings: studies of literal and figurative language. J. Pragmatics. 31, 919–929.

Giora, R., Fein, O., 1999. On understanding familiar and less-familiar figurative language. J. Pragmatics. 31, 1601–1618.

Giora, R., 2003. On our Mind: Salience, Context, and Figurative Language. Oxford University Press, New York.

Glucksberg, S., 2003. The psycholinguistics of metaphor. Trends Cogn. Sci. 7, 92–96.

Glucksberg, S., McGlone, M., Manfredi, D., 1997. Property attribution in metaphor comprehension. J. Mem. Lang. 36, 50–67.

Glucksberg, S., Gildea, P., Bookin, H.B., 1982. On understanding non-literal speech: can people ignore metaphors. J. Verbal Learn Verbal BE 21, 85–98.

Gratton, G., Coles, M.G., Donchin, E., 1983. A new method for off-line removal of ocular artifact. Electroen. Clin. Neuro. 55, 468–484.

Greenhouse, S.W., Geisser, S., 1959. On methods in the analysis of profile data. Psychometrika 24, 95–112.

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

Hagoort, P., Hald, L., Bastiaansen, M., Petersson, K-M., 2004. Integration of word meaning and world knowledge in language comprehension. Science 304.

Iakimova, G., Passerieux, C., Laurent, J.P., Hardy-Bayle, M.C., 2005. ERPs of metaphoric, literal, and incongruous semantic processing in schizophrenia. Psychophysiology 42, 380–390.

Janus, R.A., Bever, T.G., 1985. Processing of metaphoric language: an investigation of the three-stage model of metaphor comprehension. J. Psycholinguist. Res. 14, 473–487.
Keysar, B., 1989. On the functional equivalence of literal and metaphorical interpretations in discourse. J. Mem. Lang. 28, 375–385.
Kutas, M., Hillyard, S.A., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207, 203–205.
Kutas, M., Federmeier, K.D., 2000. Electrophysiology reveals semantic memory use in language comprehension. Trends Cogn. Sci. 4, 463–470.
Kutas, M., Van Petten, C., Kluender, R., 2006. Psycholinguistics electrified II: 1994–2005, In: Traxler, M.J., Gernsbacher, M.A. (Eds.), Handbook of Psycholinguistics, Second Edition. Elsevier Press, New York, pp. 659–724.
Lakoff, G., Johnson, M., 1980. Metaphors we Live by. University of Chicago, Chicago.
Lakoff, G., Turner, M., 1987. More than Cool Reason: A Field Guide to Poetic Metaphor. University of Chicago Press, Chicago.
Lakoff, G., 1993. The contemporary theory of metaphor, In: Ortony, Andrew (Ed.), Metaphor and Thought, 2nd ed. Cambridge, Cambridge University Press.
Lau, E., Phillips, C., Poeppel, D., 2008. A cortical network for semantics: [de]constructing the N400. Nature. 9, 920–931.
Onifer, W., Swinney, D., 1981. Accessing lexical ambiguities during sentence comprehension: effects of frequency of meaning and contextual bias. Mem. Cogn. 9, 225–236.
Ortony, A., Schallert, D.L., Reynolds, R.E., Antos, S., 1978. Interpreting metaphors and idioms: some effects of context on comprehension. Verbal Learn Verbal BE. 17, 465–477.
Pynte, J., Besson, M., Robichon, F.H., Poli, J., 1996. The time-course of metaphor comprehension: an event-related potential study. Brain Lang. 55, 293–316.
Rugg, M.D, Nagy, M.E., 1987. No lexical contribution to nonword-repetition effects: evidence from event-related potentials. Mem. Cog. 15 (6), 473–481.
Searle, J., 1979. Metaphor. In: Ortony, A. (Ed.), Metaphor and Thought. Cambridge University Press, New York, pp. 92–123.
Srinivasan, R., Nunez, P.L., Tucker, D.M., Silberstein, R.B., Cadusch, P.J., 1996. Spatial sampling and filtering of EEG with spline Laplacians to estimate cortical potentials. Brain Topogr. 8, 1–12.
Seidenberg, M.S., Tanenhaus, M.K., Leiman, J.M., Bienkowski, M., 1982. Automatic access of the meanings of ambiguous words in context: some limitations of knowledge-based processing. Cogn. Psyc. 14, 489–537.
Tartter, V.C., Gomes, H., Dubrovsky, B., Molholm, S., Stewart, R.V., 2002. Novel metaphors appear anomalous at least momentarily: evidence from N400. Brain Lang. 80, 488–509.
Titone, D.A., Connine, C.M., 1994. The comprehension of idiomatic expressions: effects of predictability and literality. Mem Cogn. 20, 1126–1138.
Van Berkum, J.J.A., 2008. Understanding sentence in context: what brain waves can tell us. Curr. Dir. Psychol. Sci. 17 (6), 376–380.
Van Berkum, J.J.A., in press. The neuropragmatics of "simple" utterance comprehension: an ERP review. In: Sauerland, U., Yatsushiro, K. (Eds.), Semantics and pragmatics: From experiment to theory. Basingstoke: Palgrave.
Van Petten, C, Coulson, S, Rubin, S, Plante, E, Parks, M., 1999. Time course of word identification and semantic integration in spoken language. Mem. Cogn. 25, 394–417.