

Research report

On understanding creative language: The late positive complex and novel metaphor comprehension



Karolina Rataj^{a,b,*}, Anna Przekoracka-Krawczyk^{c,d}, Rob H.J. van der Lubbe^{b,e}

^a Faculty of English, Adam Mickiewicz University, Poznań, Poland

^b Cognitive Psychology and Ergonomics, University of Twente, Enschede, The Netherlands

^c Laboratory of Vision Science and Optometry, Faculty of Physics, Adam Mickiewicz University, Poznań, Poland

^d Vision and Neuroscience Laboratory of the NanoBioMedical Centre, Adam Mickiewicz University, Poznań, Poland

^e Department of Cognitive Psychology, University of Finance and Management, Warsaw, Poland

ARTICLE INFO

Article history:

Received 8 June 2016

Received in revised form 19 October 2017

Accepted 23 October 2017

Available online 28 October 2017

Keywords:

Novel metaphor

LPC

N400

Creative language

ABSTRACT

Novel metaphoric sentences have repeatedly evoked larger N400 amplitudes than literal sentences, while investigations of the late positive complex (LPC) have brought inconsistent results, with reports of both increased and reduced amplitudes. In two experiments, we examined novel metaphor comprehension in Polish, using the same set of literal, novel metaphoric, and anomalous sentences. The first aim of the study was to test whether novel metaphors would evoke larger or smaller late positivity complex (LPC) amplitudes compared to literal and anomalous sentences. Some earlier studies have shown that whether increased LPC amplitudes are observed or not may be related to the task participants are asked to perform, with explicit acceptability judgments being more likely to evoke the LPC effect. The second aim of the study was, thus, to test whether the observed LPC pattern would be the same across two tasks, the semantic decision task (Experiment 1) and the reading task (Experiment 2). Our results replicated the N400 effect observed in earlier studies on metaphor in both experiments. Most importantly, a reduction in late positivity to novel metaphors relative to anomalous sentences was found in both experiments. Additionally, this difference was broadly distributed over parietal sites in Experiment 1, and clearly left-lateralized in Experiment 2, which might imply differences in the involvement of recollection and semantic processes. Overall, these results seem to indicate that both conventionality and task demands modulate the LPC pattern.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Novel metaphors reflect the human drive to go beyond routine in communication. Although the common association they evoke might be with novels and poetry, they seem to play an important role in everyday communication, by evoking interest, surprise, and thus attracting our attention. Many theories of metaphor and models of metaphor comprehension have pervaded the fields of linguistics, philosophy and psychology, as scholars investigated the linguistic and cognitive structure of metaphors, features that make a metaphoric sentence plausible and apt, as well as processes involved in metaphor comprehension. The latter will be the main focus of the current paper. We will attempt to test the extent to which processes involved in metaphor comprehension and

indexed by two event-related potential (ERP) components, i.e., the late positive complex (LPC) and the N400, are influenced by stimulus related aspects such as novelty of the metaphoric sentence, and task related aspects, such as potential differences between a semantic decision and a reading task. Before specific questions are raised, however, we will shortly outline previous behavioral and ERP findings related to the current study.

One of the first models of metaphor comprehension, the standard pragmatic model (Grice, 1975), stipulated that the literal meaning of a metaphoric utterance needs to be rejected before the figurative meaning is selected. Contrary to these predictions, authors of early behavioral studies on metaphor comprehension argued that the stage of literal meaning rejection is not a necessary step in arriving at non-literal meaning, by showing comparable reaction times to metaphoric and literal sentences (Gibbs et al., 1997; Inhoff et al., 1984; Janus and Bever, 1985; Onishi and Murphy, 1993; Pynte et al., 1996). This was further corroborated by results of studies that used the metaphor interference

* Corresponding author at: Faculty of English, Adam Mickiewicz University, al. Niepodległości 4, 61-874 Poznań, Poland.

E-mail address: krataj@amu.edu.pl (K. Rataj).

technique, in which participants are asked to classify three types of statements (true class-inclusion statements, e.g., *Some birds are robins*, false class-inclusion statements, e.g., *Some birds are apples*, and metaphoric statements, e.g., *Some jobs are jails*) as either literally true or literally false. Several studies have shown that participants take more time to reject a metaphoric statement than a false inclusion statement, and concluded that metaphoric meaning is activated before literal meaning processing has been completed (Glucksberg et al., 1982; Kazmerski et al., 2003; Wolff and Gentner, 2000). It is, however, important to bear in mind that early behavioral studies largely did not distinguish between different levels of conventionality of the experimental sentences, while results of recent research clearly show that conventionality impacts comprehension. This is manifested in longer reaction times and lower accuracy rates to novel metaphors as compared to conventional metaphoric or literal sentences (Arzouan et al., 2007a; Lai et al., 2009). Novel metaphors are frequently perceived as somewhat meaningful, as their meaning has no representation and needs to be constructed based on available semantic information.

This impact of conventionality has also been noted in the pattern of the N400 effect in research on metaphor comprehension. The N400, a negative-going wave observed between 300 and 500 ms after critical stimulus onset, was initially observed in response to semantic anomaly (Bentin et al., 1985; Kutas and Hillyard, 1984; Rugg and Nagy, 1987). One of the widely confirmed results is that novel metaphoric utterances evoke larger N400 amplitudes than literal utterances (Arzouan et al., 2007a,b; Coulson and Van Petten, 2002, 2007; De Grauwe et al., 2010; Goldstein et al., 2012; Lai et al., 2009; Pynte et al., 1996; Tarter et al., 2002). This increase in the N400 amplitude has been interpreted as reflecting difficulty of conceptual retrieval and integration (Arzouan et al., 2007a; Coulson and Van Petten, 2002; Goldstein et al., 2012), effort related to the integration of the unrelated concepts leading to conceptual expansion (Rutter et al., 2012), processing effort in lexical access (Weiland et al., 2014), or the extent of inferences that can be projected from one concept to another (Lai and Curran, 2013). Overall, these interpretations seem in line with the view of the N400 as reflecting the intersection between bottom-up processing of a stimulus and top-down activity in semantic memory, the result of which is an initial, flexible conceptual representation that might, however, need refinement at later stages (Kutas and Federmeier, 2011).

Interestingly, the N400 has shown sensitivity to the degree of conventionality of a metaphoric utterance, with novel metaphors (e.g., *ripe dream*) evoking somewhat larger N400 amplitudes than conventional metaphors (e.g., *transparent intention*), and somewhat smaller than anomalous utterances (e.g., *indirect blanket*) (Arzouan et al., 2007a; Goldstein et al., 2012). This graded N400 effect with similar latency, topography and morphology of the N400 component seems to provide some evidence for the claim that the difference between figurative and literal language is quantitative rather than qualitative, and stems from more complex mapping processes in the case of metaphoric than literal language, which is additionally modulated by the novelty of the mapping.

One model that accounts for both quantitative differences between literal and metaphoric language comprehension and the effects of conventionality of the metaphoric utterances observed in behavioral and electrophysiological studies is the Career of Metaphor Model (Bowdle and Gentner, 2005). This account proposes that metaphor is “a species of analogy” (Bowdle and Gentner, 2005, p. 196), and, thus, comparison is the main mechanism involved in novel metaphor comprehension. In the case of conventional metaphors, the model proposes that comparison is replaced by a faster and more automatic process of categorization in the course of conventionalization.

Two recent ERP studies have lent strong support to the Career of Metaphor Model. Lai and Curran (2013) used sentence-primers and simile-primers preceding metaphoric and literal sentences, and found that comparison induced by simile-primers facilitated mappings in novel metaphors producing reduced N400 amplitudes. This finding suggests that comparison is necessary in understanding novel metaphors. For conventional metaphors, the N400 priming effect was marginal, but it was significant in reaction times at later stages of processing. Thus, the authors concluded that comparison is vital in novel metaphor comprehension, and can be facilitative at later stages in conventional metaphor processing. Goldstein et al. (2012) asked their participants to explain half of the novel and conventional metaphoric word dyads presented in the exposure phase, followed by a test phase, in which participants performed a semantic judgment task for all the word pairs. The results showed that novel metaphoric word dyads that were explained in the exposure phase evoked smaller N400 amplitudes than those that were not explained, while conventional metaphoric word dyads that were explained in the exposure phase evoked larger N400 amplitudes than those that were not explained. This modulation of the N400 amplitudes by recent explanation was interpreted as indicative of reactivation of mappings available for conventional metaphors and sense retrieval for novel metaphors, showing that the semantic representations of both novel and conventional metaphors are dynamic and flexible.

These results clearly indicate that novel metaphor comprehension requires meaning construction rather than retrieval, which points to the creative nature of novel metaphor. One ERP study has recently compared the N400 responses to common and creative uses of objects and found a similar N400 effect to the one reported in studies on metaphor comprehension, with larger N400 amplitudes for creative than common uses (Kröger et al., 2012). This similarity in results could be seen as a link between different instances of creative thinking, and shows that investigating novel metaphor comprehension can bring us closer to understanding some aspects of creative cognition.

Although the N400 pattern observed in studies on metaphor comprehension seems to be quite consistent, it remains debatable that a single component could index all aspects involved in the comprehension of metaphoric sentences, as is evident from research on the LPC and sentence processing (Steinhauer et al., 2010). The LPC, known also as the P600, is a positive-going wave with a peak between 600 and 800 ms after stimulus onset. Although it was initially linked to the processing of syntactic anomalies (Hagoort et al., 1993; Osterhout et al., 1994), as well as to syntactic reanalysis (Friederici and Mecklinger, 1996), or syntactic integration (Kaan et al., 2000), results of later studies revealed its sensitivity to semantic anomalies, which gave rise to research on the semantic P600 (for a review see Bornkessel-Schlesewsky and Schlewsky (2008)).

The P600 following the N400, also dubbed the biphasic N400-P600, has been observed for sentences which involve non-reversible semantic anomalies (van Herten et al., 2005). Brouwer et al. (2012) interpreted this pattern as reflecting retrieval of lexical and semantic information from long term memory (the N400), and the integration of this retrieved information into the mental representation of the sentences (the P600). Within this view, novel metaphors should evoke increased N400 amplitudes compared to literal language, as the retrieval poses greater demands related to meaning construction, and increased late positivity, as the integration of the retrieved information regarding two previously not associated terms is more difficult than in the case of literal language.

Results concerning the LPC in response to metaphoric sentences are, however, far from consistent. First, several studies on metaphor comprehension that reported the N400 effect did not examine

the LPC time window. Furthermore, of those that did, some showed the expected biphasic pattern with increased N400 and LPC amplitudes to metaphoric sentences (Coulson and Van Petten, 2002; De Grauwe et al., 2010; Weiland et al., 2014), while others revealed the opposite pattern, with increased N400 but reduced LPC amplitudes to novel metaphors relative to literal and anomalous utterances (Arzouan et al., 2007a; Goldstein et al., 2012). Aspects such as task type (including semantic judgment tasks, reading tasks, or delayed response procedures) or stimulus selection criteria (with levels of meaningfulness, metaphoricity, familiarity and cloze probability controlled to various degrees) may have contributed to those inconsistencies (De Grauwe et al., 2010; Kuperberg, 2007). Below we discuss specific implications arising from these methodological differences.

The semantic decision task and the reading task have been most widely used in ERP studies on metaphor comprehension. In the semantic decision task, participants are asked to decide whether a sentence makes sense or not. This task was used in studies reported by Arzouan et al. (2007a) (Experiment 1 and 3), and Goldstein et al. (2012), which revealed a reduced LPC to novel metaphors, and by De Grauwe et al. (2010), who observed an increased LPC to conventional metaphors. In the reading task, participants are asked to read the sentences presented on the computer screen. This task was used by Arzouan et al. (2007a) (Experiment 2) and Coulson and Van Petten (2002, 2007). Additionally, in some studies (Coulson et al., 2002, 2007), participants were asked to answer follow-up true-false comprehension questions. The use of the reading tasks has also resulted in inconsistent findings, with increased (Coulson and Van Petten, 2002) and reduced (Arzouan et al., 2007a; Coulson et al., 2007) LPC amplitudes to metaphoric sentences. Overall, the semantic decision task and the reading task have revealed both patterns of the LPC effect, with smaller and larger amplitudes to metaphors than literal and anomalous utterances.

Another factor that might be crucial for the LPC findings is the extent to which conventionality was manipulated and controlled in the process of material selection. While Arzouan et al. (2007a) and Goldstein et al. (2012) employed both conventional and novel metaphoric word dyads, De Grauwe et al. (2010) tested exclusively conventional metaphors, Weiland et al. (2014) employed metaphors that were neither familiar nor unfamiliar, and Coulson and Van Petten (2002, 2007) did not directly refer to normative studies on familiarity or conventionality. Overall, results of studies in which novel metaphors were pretested on familiarity rating scales showed decreased LPC to novel metaphors compared to literal or conventional metaphoric utterances.

As the above discussion demonstrates, it remains difficult to account for the inconsistencies in LPC results in terms of task demands or novelty of metaphoric sentences, but novelty of the metaphoric meaning seems to modulate the LPC amplitudes in a more consistent manner. At the same time, the influence of task demands is not reflected in a clear LPC pattern. For these reasons, we designed a study in which we used the same set of novel metaphoric, literal, and anomalous sentences, for which the critical word was the same across the three conditions. For these sentences, we examined the ERP responses in the semantic decision task and the reading task. This design allowed us to tap into the differences in the LPC response between the two tasks, while the differences arising from possible stimulus characteristics were minimized. We expected that the results would provide support for one of the two interpretations of the LPC effect proposed in studies on metaphor comprehension, which we present below.

First, authors who have reported increased late positivity following metaphoric sentences interpreted it as indexing recovery and integration of additional material from semantic memory (Coulson and Van Petten, 2002), enhanced costs due to pragmat-

cally or inferentially driven mapping processes involving two unrelated domains (Weiland et al., 2014), or selection of the metaphoric and suppression of the literal meaning (De Grauwe et al., 2010). In line with the discussion presented by Brouwer et al. (2012), we would argue here for a more parsimonious interpretation that associates an increase in LPC amplitudes with greater effort related to the integration of semantic information retrieved within the N400 time window with the preceding context. In line with research on recognition memory (Paller and Kutas, 1992), it could also index recollection of the entire sentence, which might be more demanding in the case of novel metaphors due low familiarity.

On the other hand, attenuated LPC amplitudes to metaphors were observed in the studies reported by Arzouan et al. (2007a) and Goldstein et al. (2012), but only in response to novel and not conventional metaphoric word pairs. Such reduced LPC amplitudes remain difficult to account for within the interpretation of the LPC as reflecting semantic integration, as novel metaphors would be expected to require more integration than literal sentences and, consequently, should evoke increased LPC amplitudes. One possible explanation postulated by Arzouan et al. (2007a) and Goldstein et al. (2012) is that an overlapping negativity attenuates late positivity amplitudes in the case of novel metaphoric utterances, and might be indicative of access to the non-literal route in the course of novel metaphor comprehension (Arzouan et al., 2007a, Goldstein et al., 2012). Following this interpretation, we would argue that novel metaphors require prolonged retrieval of information from long term memory (possibly related to access to the non-literal route) and mapping between two semantically distant concepts, that is initiated within the N400 time window and later overlaps with the LPC time window. Such overlapping sustained negativity is then reflected in a reduced LPC amplitude to novel metaphors.

This interpretation could be partly supported by a recent study that found sustained negativity between 500 and 900 ms after critical word onset in response to novel metaphors (Rutter et al., 2012). The design of the study differed from other studies on metaphor in several aspects. Neither the semantic decision nor the reading task was used. Instead, participants were asked to decide whether or not a sentence was unusual (question 1) and appropriate (question 2). Also, a delayed response procedure was used. Although sustained negativity was interpreted as reflecting ongoing difficulty in integrating sentence meaning, it could also be related to high demands on working memory (Steinhauer and Drury, 2012; Steinhauer et al., 2010), which in this paradigm could have been triggered by the need to maintain the activated information in memory due to the delayed response procedure, and by task complexity. This interpretation could further be supported by the fact that sustained negativity was associated with all sentence types (literal, metaphoric and anomalous), with amplitude differences consistent with the N400 pattern.

In the current study, we aimed to test whether increased semantic integration (increased LPC amplitudes) or overlapping sustained negativity associated with access to non-literal meaning or continued difficulty in the mapping process (reduced LPC) would be associated with novel metaphor comprehension. Furthermore, we planned to investigate whether the obtained LPC effect would be present regardless of the task the participants were asked to perform. Additionally, we predicted that the N400 effect would be replicated, with novel metaphors evoking larger N400 amplitudes than literal, and smaller than anomalous sentences. Since this effect has been found in several earlier studies that used different task procedures, we expected that it would be present in both experiments.

To these aims, we designed two experiments with the semantic decision task (Experiment 1) and the reading task (Experiment 2). To ensure that the sentences in Experiment 2 were read for com-

prehension, we used yes-or-no comprehension questions on 30% of the trials. The same set of materials including novel metaphoric, literal and anomalous sentences was used in both experiments in order to minimize differences in results that might arise from the different degrees of meaningfulness, familiarity, metaphoricity and cloze probabilities in the experimental sentences. Unlike the materials used in the studies by [Arzouan et al. \(2007a,b\)](#) and [Goldstein et al. \(2012\)](#), who used semantically related, unrelated, conventional and novel metaphoric two-word expressions, our stimuli included novel metaphoric, literal, and anomalous sentences. [Lai et al. \(2009\)](#) also used sentences, but the critical words included in their study were verbs, adjectives, and nouns, while our critical words were nouns. Finally, unlike [De Grauwe et al. \(2010\)](#), we did not use nominal (*A is B*) metaphors. Instead, our critical nouns were preceded by nouns, verbs, or adjectives. Similarly to all of the studies mentioned above (with the exception of Experiment 1 in [De Grauwe et al. \(2010\)](#)), the ERPs were time-locked to the sentence-final word.

2. Results

2.1. Experiment 1

2.1.1. Behavioral results

Two ANOVAs with sentence type as a factor were conducted on the reaction times and the proportion of correct responses recorded in Experiment 1 ([Table 1](#)). The analysis of reaction times revealed a main effect of sentence type [$F(2,58) = 41, p < .001, \eta_p^2 = 0.58$]. Bonferroni corrected pairwise comparisons showed that novel metaphoric sentences evoked the longest reaction times and differed significantly from literal [$p < .01$] and anomalous sentences [$p < .01$], but no difference was observed between the anomalous and literal sentences [$p > .05$]. The analysis of accuracy rates also showed a significant effect of sentence type [$F(2,58) = 143, p < .0001, \eta_p^2 = 0.83$]. Bonferroni corrected pairwise comparisons showed that novel metaphoric sentences differed significantly from literal [$p < .01$] and anomalous sentences [$p < .01$], and that literal sentences differed from anomalous sentences [$p < .05$], which were the easiest to evaluate.

2.1.2. EEG results

2.1.2.1. N400. The results of the anterior-posterior axis \times laterality \times sentence type ANOVA in the time window between 300 and 500 ms revealed a main effect of sentence type [$F(2,58) = 6.10, p < .01, \eta_p^2 = 0.17$]. Literal sentences elicited the smallest ($M = 0.42, SE = 0.14$), anomalous sentences the largest ($M = 0.04, SE = 0.16$), and novel metaphors intermediate ($M = 0.14, SE = 0.15$) N400 amplitudes. One-tailed *t*-tests showed that amplitudes for literal sentences differed significantly from anomalous [$p < .01$], and novel metaphoric sentences [$p < .01$], but no difference was observed between anomalous and novel metaphoric utterances [$p > .05$] (uncorrected, for reasons explicated in [Section 4.4.](#) on p. 33). However, the N400 revealed a linear effect [$F(1,29) = 9.86, p < .01, \eta_p^2 = 0.25$], with amplitudes for novel metaphoric sentences falling in between those for literal and anomalous utterances.

An interaction between the anterior-posterior axis and sentence type was also observed [$F(4,116) = 3.56, p < .05, \eta_p^2 = 0.11$]. To deconstruct this interaction, we performed separate analyses for frontal, central, and parietal sites. As visible in the grand average ERPs ([Fig. 1](#)) and the difference wave maps ([Fig. 2](#), upper panel), the largest difference was observed between the anomalous and literal condition, and a somewhat smaller difference between the anomalous and novel metaphoric condition over central and parietal sites. Statistical tests confirmed this observation, with the effect of sentence type being most pronounced over central sites

Table 1

Means (*M*) and standard errors (*SE*) for reaction times and proportion of correct responses to literal, novel metaphoric, and anomalous sentences (Experiment 1).

	Reaction times (ms) (<i>M</i> and <i>SE</i>)	Proportion of correct responses (<i>M</i> and <i>SE</i>)
Novel metaphoric sentences	1106 (45)	46 (0.022)
Literal sentences	939 (37)	89 (0.025)
Anomalous sentences	892 (34)	94 (0.03)

[$F(2,58) = 9.88, p < .001, \eta_p^2 = 0.25$]. One-tailed *t*-tests revealed that anomalous sentences ($M = -0.73, SE = 0.25$) evoked larger negativity than literal sentences ($M = -0.07, SE = 0.19$) [$p = .001$], and novel metaphoric sentences ($M = -0.46, SE = 0.21$) [$p = .05$] (uncorrected). The difference between literal and novel metaphoric sentences was also significant over the central sites [$p < .01$] (uncorrected). Moreover, a linear effect was observed [$F(1,29) = 17.91, p < .001, \eta_p^2 = 0.38$].

A similar though somewhat smaller effect of sentence type was found over parietal sites [$F(2,58) = 4.81, p = .01, \eta_p^2 = 0.14$]. One-tailed *t*-tests revealed that anomalous sentences ($M = -0.33, SE = 0.26$) evoked larger negativity than literal sentences ($M = 0.29, SE = 0.29$) [$p < .01$] and novel metaphoric sentences ($M = 0.21, SE = 0.21$) [$p < .01$], while no difference was observed between literal and novel metaphoric utterances [$p > .05$] (uncorrected). No effect of sentence type [$p > .05$] was observed over frontal sites.¹

2.1.2.2. LPC. For the LPC time window (500–800 ms), the anterior-posterior axis \times laterality \times sentence type ANOVA revealed a significant anterior-posterior \times laterality \times sentence type interaction [$F(8, 232) = 2.14, p < .05, \eta_p^2 = 0.07$], and a significant anterior-posterior \times sentence type interaction [$F(4, 116) = 7.33, p = .001, \eta_p^2 = 0.20$]. To deconstruct these interactions, we performed separate analyses for frontal, central, and parietal sites.

An ANOVA on amplitudes recorded over frontal sites revealed a laterality \times sentence type interaction [$F(4,116) = 3.27, p < .01, \eta_p^2 = 0.10$]. Follow-up analyses over Fz showed a main effect of sentence type [$F(2,58) = 4.28, p < .05, \eta_p^2 = 0.13$]. Bonferroni corrected pairwise comparisons revealed that novel metaphoric sentences ($M = 2.32, SE = 0.33$) evoked larger LPC amplitudes than anomalous sentences ($M = 1.36, SE = 0.35$) [$p < .01$], but did not differ from literal sentences ($M = 1.97, SE = 0.39$) [$p > .05$]. Amplitudes for literal sentences did not differ from those for anomalous sentences [$p > .05$]. However, a linear effect was observed [$F(1,29) = 11.15, p < .01, \eta_p^2 = 0.28$]. A similar pattern was found over F4, [$F(2,58) = 5.52, p < .01, \eta_p^2 = 0.16$], with novel metaphoric sentences ($M = 1.16, SE = 0.29$) evoking larger positivity than anomalous sentences ($M = 0.17, SE = 0.32$) [$p < .01$]. Here, again, no differences between literal ($M = 0.95, SE = 0.33$) and anomalous [$p > .05$], or literal and novel metaphoric utterances [$p > .05$] were observed, but a linear effect was found [$F(1,29) = 11.88, p < .01, \eta_p^2 = 0.29$]. No main effect of sentence type was found over F3 [$F(1,29) = 2.10, p > .10, \eta_p^2 = 0.07$].

¹ We performed additional analyses on literal sentences evaluated as meaningful, novel metaphoric sentences evaluated as meaningful (nMET_meaningful), novel metaphoric sentences evaluated as meaningless (nMET_meaningless), and anomalous sentences evaluated as meaningless. We found no differences between nMET_meaningful and nMET_meaningless, as well as between nMET_meaningful and literal or anomalous, or between nMET_meaningless and anomalous [$ps > .7$]. The only trend we have found was for the difference between nMET_meaningless and literal sentences [$p = .07$]. Future research is needed to look closer into differences in the ERPs depending on whether novel metaphors are perceived as meaningful or meaningless, with possibly many more trials within all conditions. This is particularly important for novel metaphors, which are less frequently evaluated as meaningful than literal utterances.

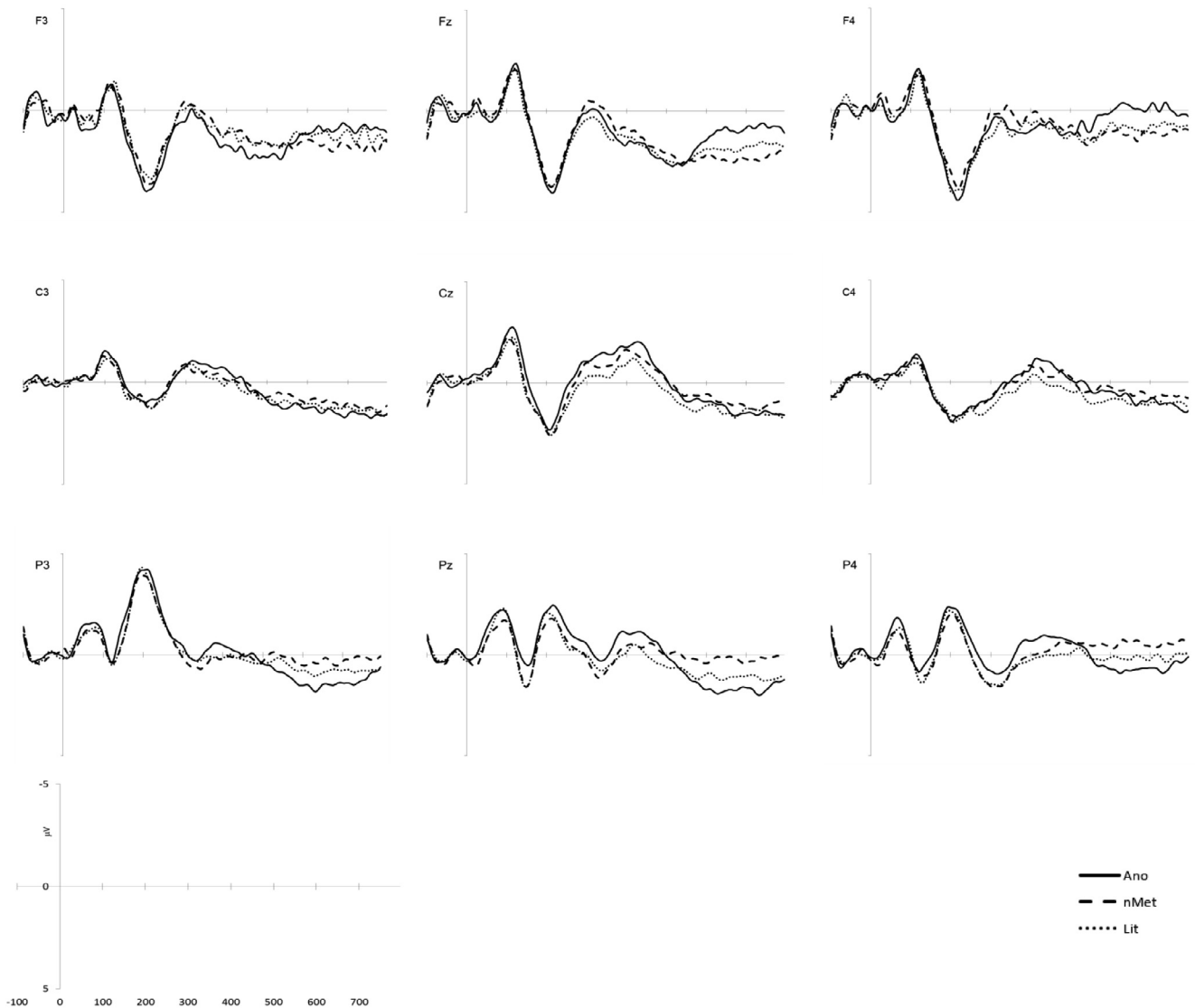


Fig. 1. Grand average ERPs for anomalous (solid lines), literal (dotted lines), and novel metaphoric sentences (dashed lines) in Experiment 1 (the semantic decision task).

A somewhat different pattern emerged from the analyses over parietal sites, where a main effect of sentence type was found [$F(2,58) = 9.34, p < .01, \eta_p^2 = 0.24$], with novel metaphoric sentences evoking reduced LPC amplitudes ($M = -0.05, SE = 0.25$) relative to anomalous sentences ($M = 1.03, SE = 0.25$) [$p = .001$]. Literal sentences evoked intermediate amplitudes ($M = 0.60, SE = 0.27$) and differed from novel metaphoric [$p = .01$] but not anomalous sentences [$p > .05$]. Here again, a linear effect was found [$F(1,29) = 15.79, p < .001, \eta_p^2 = 0.35$]. Over the central sites, only a main effect of utterance type was observed [$F(2,58) = 3.45, p < .05, \eta_p^2 = 0.11$], however pairwise comparisons did not reveal differences between the three sentence types. To further test whether the difference between anomalous and novel metaphoric sentences had a broad parietal distribution, as visible in difference wave maps in Fig. 3 (upper panel), separate analyses on P3, Pz, and P4 were conducted. They revealed the same pattern over all three electrodes, with reduced LPC amplitudes for novel metaphoric relative to anomalous sentences [$ps < .01$]. Literal sentences evoked larger LPC amplitudes than novel metaphoric sentences over Pz [$p < .01$], P4 [$p < .05$], but not P3 [$p > .05$]. Also, amplitudes for anomalous and literal sentences did not differ over any of the electrodes [$p > .05$]. The grand average ERPs for the three sentence types are depicted in Fig. 1.

2.2. Experiment 2

2.2.1. Behavioral results

In Experiment 2, comprehension questions were presented on 30% of the trials to check whether participants were actively reading and comprehending the sentences. The analysis of the proportion of correct responses revealed that participants gave 91% of correct responses ($M = 91, SD = 7$), which means that the sentences were read attentively.

2.2.2. EEG results

2.2.2.1. N400. The results of the anterior-posterior axis x laterality x sentence type ANOVA between 300 and 500 ms revealed only a main effect of sentence type [$F(2,44) = 4.09, p < .05, \eta_p^2 = 0.16$], and no interactions with sentence type [$ps > .10$]. As visible in the grand average ERPs (Fig. 4) and the difference wave maps (Fig. 2, lower panel), literal sentences ($M = 0.38, SE = 0.12$) elicited smaller N400 amplitudes than anomalous sentences ($M = -0.01, SE = 0.16$) [$p < .01$], and novel metaphoric sentences ($M = 0.03, SE = 0.15$) [$p < .05$]. The difference between novel metaphoric and anomalous sentences did not reach significance [$p > .05$]. Like in Experiment 1, a linear effect showed a graded pattern in the recorded amplitudes [$F(1,22) = 6.84, p < .05, \eta_p^2 = 0.24$].

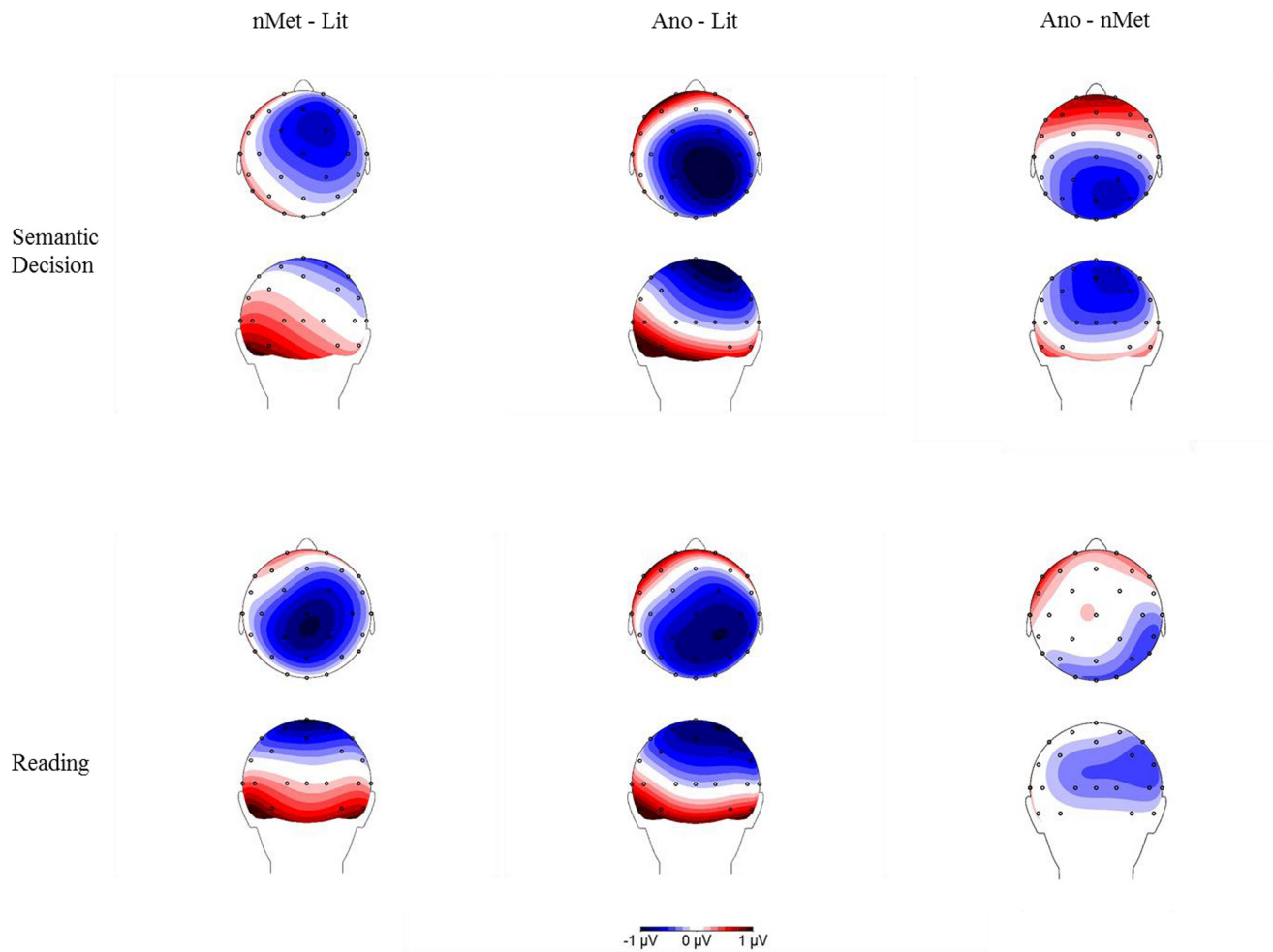


Fig. 2. Difference wave maps showing the N400 effect (300–500 ms) by subtracting the literal sentence condition (LIT) from the novel metaphoric sentence condition (nMET), the literal sentence condition (LIT) from the anomalous sentence condition (ANO), and the novel metaphoric sentence condition (nMET) from the anomalous sentence condition (ANO) in Experiment 1 (the semantic decision task) and Experiment 2 (the reading task).

2.2.2.2. LPC. In the LPC time window, the anterior-posterior axis \times laterality \times sentence type ANOVA revealed a laterality \times sentence type interaction [$F(4,88) = 4.44, p = .01, \eta_p^2 = 0.17$]. Follow-up ANOVAs over left, midline and right sites were performed to deconstruct this interaction. Only the analysis over left sites revealed a main effect of sentence type [$F(2,44) = 7.85, p = .001, \eta_p^2 = 0.26$]. Bonferroni corrected pairwise comparisons revealed that anomalous sentences ($M = 1.19, SE = 0.22$) evoked significantly larger positivity than metaphoric sentences ($M = 0.67, SE = 0.19$) [$p = .001$]. Amplitudes for literal sentences appeared smaller than for anomalous sentences ($M = 0.82, SE = 0.22$), but this difference was only marginally significant [$p = .08$]. No significant difference was observed between the literal and novel metaphoric sentences, but a significant linear effect was found [$F(1,22) = 18.87, p < .001, \eta_p^2 = 0.46$].

To further test whether the difference between anomalous and novel metaphoric sentences had a broad left distribution, separate analyses on F3, C3, and P3 were conducted. As visible in the difference wave maps in Fig. 3 (lower panel), the difference was absent over the frontal sites. Statistical tests confirmed this observation, with no main effect of sentence type over F3 [$F(2,44) = 0.16, p > .70, \eta_p^2 = 0.01$]. Over C3 and P3, both a main effect ([$F(2,44) = 6.60, p < .01, \eta_p^2 = 0.23$], [$F(2,44) = 6.35, p < .01, \eta_p^2 = 0.22$], respectively), and a difference between anomalous and novel metaphoric sentences [$ps < .01$] were observed. LPC amplitudes for literal sen-

tences recorded over C3 and P3 did not differ from those for novel metaphoric and anomalous sentence [$ps > .05$].

No main effect of sentence type was observed over the midline [$F(2,44) = 1.74, p > .10, \eta_p^2 = 0.07$] or right sites [$F(2,44) = 0.78, p > .10, \eta_p^2 = 0.03$]. The grand average ERPs for the three sentence types in the LPC time window are displayed in Fig. 4.

Overall, the analyses revealed a graded N400 effect in both experiments. Although the difference wave maps for Experiment 1 might suggest that the difference between literal and novel metaphoric sentences had a frontal distribution, statistical tests did not confirm this observation. Moreover, LPC amplitudes for novel metaphoric sentences were reduced relative to those for anomalous sentences over parietal sites in both experiments. While this effect was clearly left-lateralized in Experiment 2, it presented a more broad parietal distribution Experiment 1. This difference will be further elucidated in analyses on rescaled data presented below.

2.3. Comparison of Experiment 1 and Experiment 2

To investigate potential topographical differences between the N400 and LPC effects observed in Experiment 1 and Experiment 2, we conducted analyses in which these effects were directly compared. Since topographical differences might stem from overall

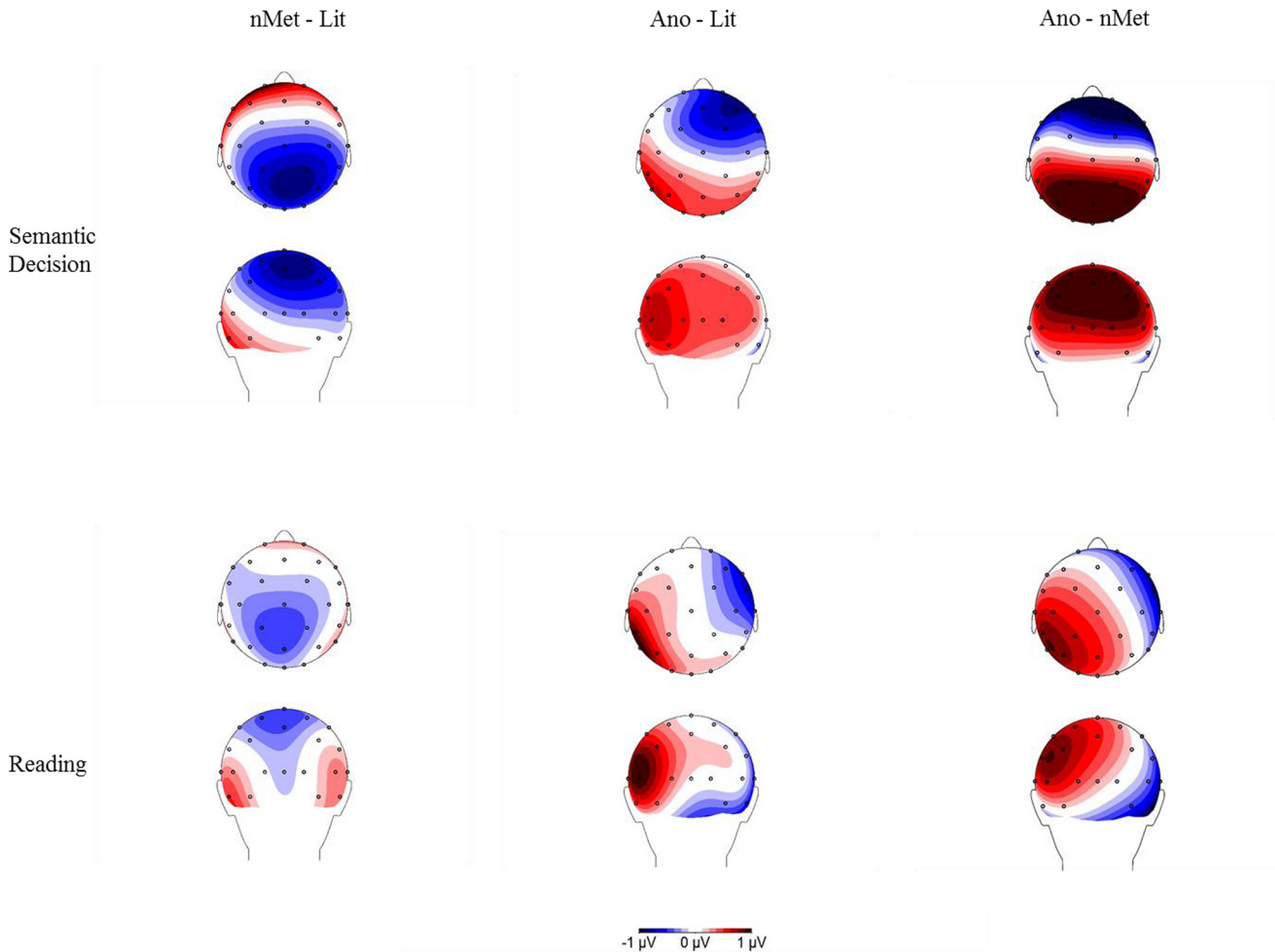


Fig. 3. Difference wave maps showing the LPC effect (500–800 ms) by subtracting the literal sentence condition (LIT) from the novel metaphoric sentence condition (nMET), the literal sentence condition (LIT) from the anomalous sentence condition (ANO), and the novel metaphoric sentence condition (nMET) from the anomalous sentence condition (ANO) in Experiment 1 (the semantic decision task) and Experiment 2 (the reading task).

amplitude differences, we employed a rescaling procedure (see Section 4.4.).

2.3.1. N400

An anterior-posterior \times laterality \times experiment ANOVA on rescaled difference waves revealed no interactions involving the factor experiment [$ps > 0.05$].

2.3.2. LPC

The anterior-posterior \times laterality \times experiment ANOVA on rescaled data for the metaphor-literal difference waves revealed no interactions involving the factor experiment. We only observed a main effect of the anterior-posterior axis [$F(2,102) = 5.40, p < .01, \eta_p^2 = 0.10$], which indicated that the literal-metaphor difference was largest above parietal ($M = 0.11, SE = 0.03$), followed by central ($M = 0.08, SE = 0.03$), and smallest over frontal sites ($M = -0.06, SE = 0.05$).

The anterior-posterior \times laterality \times experiment ANOVA on rescaled data for the anomaly-metaphor difference waves revealed an anterior-posterior \times laterality \times experiment interaction [$F(4,204) = 2.66, p < .05, \eta_p^2 = 0.05$]. A main effect of anterior-posterior axis [$F(2,102) = 8.65, p < .01, \eta_p^2 = 0.15$], and a main effect of laterality [$F(2,102) = 5.10, p < .01, \eta_p^2 = 0.09$] were also observed.

To elucidate the interaction, separate analyses for each experiment were conducted on the rescaled data. The anterior-poster-

ior \times laterality ANOVA for Experiment 1 revealed a main effect of the anterior-posterior axis [$F(2,58) = 13.13, p < .001, \eta_p^2 = 0.31$]. The anomaly-metaphor difference was largest over parietal sites ($M = 0.20, SE = 0.05$), followed by frontal sites ($M = -0.13, SE = 0.05$), and smallest over central sites ($M = 0.09, SE = 0.04$). The difference over frontal sites was inverted relative to the posterior sites, which confirms the findings observed in the analysis on unscaled average amplitudes in Experiment 1. Moreover, an interaction between the anterior-posterior axis and laterality was found [$F(4,116) = 3.10, p < .05, \eta_p^2 = 0.10$], with the largest anomaly-metaphor difference observed over the Pz ($M = 0.24, SE = 0.06$), followed by P4 ($M = 0.18, SE = 0.05$), and P3 electrode ($M = 0.17, SE = 0.06$). The anterior-posterior \times laterality ANOVA for Experiment 2 revealed a main effect of laterality [$F(2,44) = 6.08, p < .01, \eta_p^2 = 0.22$]. The anomaly-metaphor difference was largest over left sites ($M = 0.15, SE = 0.04$), followed by midline sites ($M = 0.07, SE = 0.05$), and smallest over right sites ($M = -0.05, SE = 0.04$).

The anterior-posterior \times laterality \times experiment ANOVA on rescaled data for the anomaly-literal difference waves revealed no interactions with experiment as a factor. A main effect of anterior-posterior axis [$F(2,102) = 4.37, p < .05, \eta_p^2 = 0.08$], and a main effect of laterality were observed [$F(2,102) = 7.00, p < .01, \eta_p^2 = 0.12$]. The main effect of anterior-posterior axis indicated that the anomaly-literal difference was largest above parietal ($M = 0.09, SE = 0.04$), followed by frontal ($M = -0.06, SE = 0.04$), and smallest

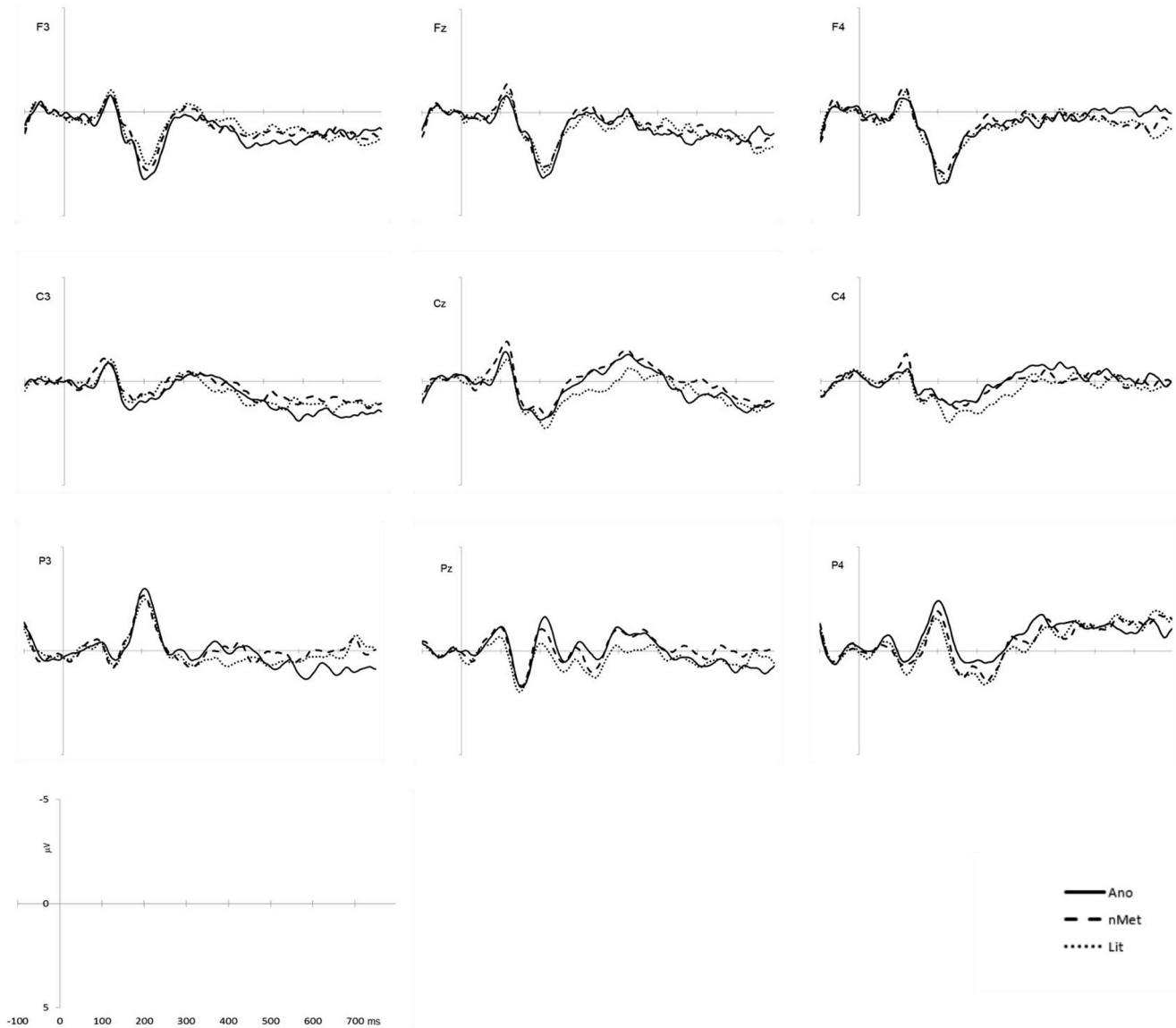


Fig. 4. Grand average ERPs for anomalous (solid lines), literal (dotted lines), and novel metaphoric sentences (dashed lines) in Experiment 2 (the reading task).

over central sites ($M = -0.01$, $SE = 0.03$). The main effect of laterality indicated that the anomaly-literal difference was largest above left ($M = 0.08$, $SE = 0.03$), followed by right ($M = -0.05$, $SE = 0.03$), and smallest over midline sites ($M = -0.0004$, $SE = 0.03$).

3. Discussion

In the current study, novel metaphoric, literal, and anomalous sentences were used to investigate the processes involved in novel metaphor comprehension. The main aim was to test whether novel metaphors would elicit increased or reduced LPC amplitudes compared to the remaining sentence types, and whether this pattern would be observed in both the semantic decision and the reading task. The secondary aim was to examine whether the N400 effect could be replicated in both tasks. The same set of sentences was used in both experiments to ensure that the observed differences would not stem from differences in the degrees of metaphoricity, meaningfulness, familiarity, or cloze probabilities of the experimental sentences.

The analysis of the N400 time window corroborated previous findings regarding the N400 effect in both experiments, with ampli-

tudes for novel metaphoric sentences falling in between those for literal and anomalous sentences. Although the effect was somewhat less pronounced in Experiment 2 (where amplitudes for novel metaphoric and anomalous sentences did not differ), the direction of the effect was the same in both experiments, as confirmed by significant linear effects. Although inspection of difference wave maps suggested a topographical difference between the two experiments, with novel metaphors evoking larger negativity than literal sentences over frontal sites in the semantic decision task and over centro-parietal sites in the reading task, analyses on rescaled data did not corroborate these observations.

Overall, the N400 effect observed in the current study remains in line with previous findings, and can be interpreted as an index of increased activation in semantic memory related to the difficulty of the mappings required for the comprehension of novel metaphors as compared to literal sentences, and of anomalous as compared to novel metaphoric utterances. The comparable latency and topography observed for all conditions is also in line with previous studies that employed novel metaphors, and might indicate that the difference between literal and novel metaphoric language is quantitative rather than qualitative. This may lend some support

for the Career of Metaphor Model (Bowdle and Gentner, 2005), which predicts that mappings underlying novel metaphor comprehension require aligning information from semantically more distant concepts than in the case of literal language, which implies more activity in semantic memory. It is crucial to note, however, that any definite conclusions in favor of this or other models of metaphor comprehension fall outside the scope of the current study.

The analysis of the LPC time window revealed a more complex pattern of results. Contrary to some earlier studies (Coulson and Van Petten, 2002; De Grauwe et al., 2010; Weiland et al., 2014), a reduction in positivity for novel metaphoric sentences as compared to anomalous and literal sentences was observed in Experiment 1. These findings are in accordance with two previous reports that made a clear distinction between novel and conventional metaphoric sentences in the experimental materials (Arzouan et al., 2007a; Goldstein et al., 2012). Studies that reported increased late positivity investigated familiar metaphors (De Grauwe et al., 2010), metaphors that were evaluated as neither familiar nor unfamiliar in the normative studies preceding the EEG experiments (Weiland et al., 2014), or metaphors that were not explicitly pretested on familiarity or conventionality scales (Coulson and Van Petten, 2002). Thus, a major conclusion stemming from the current findings is that conventionality modulates not only the N400, but also the LPC effect.

The reduction in the LPC amplitudes in response to novel metaphoric sentences remains difficult to interpret in light of current theories of the LPC, which postulate that an increase in its amplitude indicates increased secondary semantic integration demands. In line with this interpretation, novel metaphors should evoke larger LPC amplitudes than literal sentences. In line with what Arzouan et al. (2007a) proposed, we interpret this result as reflecting late negativity overlapping with the LPC, thus reducing its amplitude. Further evidence in favor of this interpretation can be found in the study reported by Rutter et al. (2012), who observed sustained negativity between 500 and 900 ms after critical word onset for novel metaphoric sentences as compared to literal sentences. This negativity followed the N400 effect, in which literal sentences evoked the smallest, anomalous sentences the largest and novel metaphors intermediate N400 amplitudes. Interestingly, in the later time window, the sustained negativity effect mimicked the N400 effect (anomalous sentences evoked the most negative amplitudes). This pattern, though, might have been linked to increased working memory demands due to a delayed response procedure, or to task complexity (participants were asked to decide whether or not a sentence was unusual (question 1) and appropriate (question 2)).

Task complexity might potentially be related to our findings in that the semantic decision task employed in Experiment 1 might have posed more difficulty in the case of novel metaphors (somewhat meaningful) than literal and anomalous sentences (clearly meaningful/meaningless). The meaningfulness ratings obtained in the normative studies seem to reflect this pattern. They showed that novel metaphoric sentences fell in between literal and anomalous sentences on a 5 point rating scale. Since the binary decision task employed in the EEG experiment did not include the middle option (somewhat meaningful), the decision might have been particularly taxing for novel metaphoric sentences. This argument can be supported by significantly longer reaction times and lower accuracy rates for novel metaphoric sentences compared to literal and anomalous sentences in the behavioral results. However, task demands cannot account for the reduction of the LPC amplitudes observed for novel metaphors in the reading task (Experiment 2).

Negativity overlapping with the LPC might be associated with higher working memory demands (Steinhauer et al., 2010) and

might index sustained activity in memory needed to perform the mapping process when comprehending a somewhat meaningful sentence, that is a novel metaphor. In the case of anomalous sentences, increased LPC amplitudes might reflect processes engaged in memory retrieval (Paller and Kutas, 1992), in that the recollection of the entire sentence includes the retrieval of the anomaly, which might constitute additional information when compared to the recollection of a literal sentence. Literal sentences, at last, evoked intermediate LPC amplitudes. This effect might be due to the fact that recollection is less demanding in the case of literal relative to anomalous sentences, and at the same time does not require the novel mapping, which might underlie reduced amplitudes for novel metaphoric utterances.

Another important finding of the current study is that while the difference between the anomalous and novel metaphoric sentences in the LPC amplitudes in the semantic decision task presented a broad parietal distribution, it was left-lateralized in the reading task. Interestingly, this left-lateralized effect in the reading task resembles the old/new effect observed in recognition memory studies (Stróžak et al., 2016a,b; Vilberg and Rugg, 2009). In the task most frequently used in recognition memory studies, participants were presented with stimuli (words or pictures) either in the study and test phase (old information), or only in the test phase (new information). Larger left-lateralized positivity over parietal regions was observed on trials correctly recognized as old than on trials correctly recognized as new. This effect has been interpreted as reflecting recollection memory, with more information available for retrieval evoking larger positivity.

This left-lateralized LPC effect was further dissociated from semantic priming effects in a study by Voss and Federmeier (2011), who showed that while recollection memory processes were indexed by late left posterior positivity, semantic processes were associated with late right posterior positivity. This result might provide a possible explanation of the difference between the two experiments found in the current study. The analysis of the anomaly-metaphor difference waves (see Section 2.3.2.) revealed an interaction between the anterior-posterior axis, laterality, and experiment. The results showed that while in Experiment 1 this difference was largest over the midline parietal site (Pz), in Experiment 2 it was clearly left-lateralized. This effect might be related to task demands. In line with the findings reported by Voss and Federmeier (2011), the broad distribution of the LPC effect in the semantic decision task might suggest the involvement of both recollection and semantic processes. In Experiment 2 (the reading task), recollection processes might be more relevant due to the requirement of responding to comprehension questions following the presentation of the sentences, which might provide a possible interpretation of the lateralization of the effect.

It is important to treat this interpretation with caution due to large differences between the recognition memory studies and the current study, both in terms of materials (individual words and pictures vs literal, novel metaphoric, and anomalous sentences) and procedures (old/new classification vs the semantic decision and reading tasks). It remains beyond the scope of this paper to draw conclusions about a possible dissociation between memory processes and semantic processes as indexed by LPC effects in the context of anomaly and novel metaphor processing, and further research is needed to elucidate this effect. Furthermore, the LPC pattern might be confounded by other factors that have been shown to influence its amplitude and topography, and have rarely been considered in studies on metaphor, such as valence and arousal (Xu et al., 2015), or animacy (Kuperberg et al., 2007). Future research is needed to investigate these factors in a systematic manner.

Finally, the analysis of the rescaled anomaly-metaphor difference waves (Section 2.3.2) confirmed yet another finding observed in the analysis on unscaled LPC amplitudes in Experiment 1 (Section 2.1.2.2). Namely, while novel metaphoric sentences evoked reduced LPC amplitudes over parietal sites, an enhancement of these amplitudes was found over frontal sites. At the same time, LPC amplitudes to anomalous sentences were larger over posterior and reduced over frontal sites. Late frontal positivity has previously been reported in response to low cloze probability sentences with a semantically somewhat plausible continuation, while low probability implausible (anomalous) sentences evoked late posterior positivity (De Long et al., 2014). Novel metaphoric sentences employed in the current study might be potentially similar to sentences evoking frontal positivities in the study by De Long et al. (2014), in that they might appear somewhat plausible. Thus the topographical difference in late positivities observed here might reflect the influence of plausibility and probability levels on the components. One possible explanation of such frontal positivity is that it indexes the inhibition of the pre-activated expected meaning, so that a less plausible but probable continuation is processed (De Long et al., 2014). This interpretation remains, however, very tentative here, as in the current study plausibility was not directly manipulated and probability levels were quite homogeneous. Moreover, the frontal effect observed by De Long et al. (2014) had a broad distribution, while in the current study the difference between anomalous and novel metaphoric sentences had a slight right hemisphere bias. Future research is therefore needed to test the parietal and frontal positivities in novel metaphor comprehension, with possible modulations of the components dependent on the plausibility and probability levels.

There are a few potential caveats in the current study. First, 5-point scales were used for the normative studies on meaningfulness and metaphoricity, while a 7-point scale was employed for familiarity. A more consistent approach would be to use the same number of points for all scales. Second, the presentation of the whole sentence context instead of the word-by-word presentation differs from other studies on metaphor comprehension. We chose this mode of stimulus presentation as a word-by-word presentation exerts increased demands on working memory and induces a graded, rather than one stage, build-up of contextual information preceding the critical word. The effects we observed are, however, in line with previous research on metaphor comprehension, and the differences between Experiment 1 and 2 cannot be accounted for by the differences in the presentation mode.

For the current study, we recruited students of the Physics department, which is a departure from previous studies that have mostly recruited students of psychology. The similarity in the N400 findings might point to little differences that would stem from participants' educational profile. It is important to note, though, that for any firm conclusions, future research is needed that would directly compare participant groups of different educational backgrounds.

In sum, the present study replicated the N400 effect observed in previous studies in Experiment 1 and 2, with literal sentences evoking the smallest, anomalous the largest, and novel metaphoric sentences intermediate amplitudes. This effect indicates that comprehension of novel metaphoric sentences requires increased activation of information in semantic memory relative to comprehension of literal sentences, regardless of task demands. Although, in light of the ongoing debate on the functional role of the LPC, any firm conclusions based on the current study would be preliminary, two crucial findings emerge here. First, novel metaphors displayed a reduced LPC amplitude relative to anomalous sentences in Experiment 1 and Experiment 2. Although at first this effect seems at odds with the current theories of the LPC, together with results of previous studies that employed novel metaphors, it

suggests that conventionality matters not only in the early stages (the N400), but also in the later stages (LPC) of metaphoric language processing. Second, some support has been gathered for the influence of task demands on LPC amplitudes. That is, while in the semantic judgment task the difference in amplitudes between the anomalous and novel metaphoric sentences was broadly distributed over parietal sites, in the reading task, it presented a left-lateralized effect. This finding might suggest a possible dissociation between semantic and recollection processes and could serve as inspiration for future research. After Kuperberg (2007), we argue that task demands should not be easily discarded as irrelevant for language processing, as they modulate the observed ERP patterns.

4. Experimental procedures

4.1. Experiment 1

4.1.1. Participants

Thirty-five students of the Faculty of Physics at Adam Mickiewicz University in Poznań, Poland participated in Experiment 1 in exchange for course credits. Five participants were excluded from the final analysis due to a low number of correct responses. Thus, data from 30 participants (7 male and 23 female, $M_{age} = 23$, range 20–26) were introduced in the final analysis. All participants were right-handed native speakers of Polish, had normal or corrected to normal vision, and no history of neurological disorders.

4.1.2. Materials

All critical words were nouns with the mean number of characters equal to 7.25 ($SD = 1.9$, range = 4–13) and mean frequency of occurrence per million equal to 28 ($SD = 15$, range: 10–63). Mean frequency was calculated based on the IPI PAN Corpus of Polish (Przepiórkowski, 2004). Each critical word appeared in the literal, novel metaphoric and anomalous condition (see Table 2). Two hundred novel metaphoric, literal, and anomalous utterances were created by the experimenters. In order to ensure that the sentences fulfilled the target criteria, three normative studies were performed. Means and standard deviations for all the normative tests are presented in Table 3. Independent judges who completed the normative tests were recruited from the population of students of Adam Mickiewicz University. The judges did not participate in the EEG experiments.

First, the probability cloze test was conducted, in which a group of 34 participants (19 females, $M_{age} = 21$, range = 18–24) saw the target sentences without the critical words and completed each sentence with the word that first came to their minds. Since previous research has shown that cloze probabilities modulate the N400 amplitudes, we wanted to ensure that the critical words were equally unexpected in all three conditions. Those sentences for which the critical word was repeated twice or more in the responses were excluded from the final set of stimuli. Cloze probability for the final list of sentences in all conditions was kept below 2%. Although there was a significant effect of sentence type [$p < .01$], it stemmed from the differences between the anomalous and literal [$p < .01$], and anomalous and novel metaphoric sentences [$p = .001$]. There was no difference between the literal and novel metaphoric sentences [$p > .05$].

Second, another group of 34 independent judges (18 females, $M_{age} = 21$, range = 18–23) was asked to evaluate the sentences on a 5 point meaningfulness scale, ranging from very meaningful (1) to very meaningless (5). This normative test was performed to ensure that the novel metaphoric and literal sentences were perceived as meaningful while the anomalous sentences as clearly meaningless. Based on previous research, we expected that novel

Table 2

Examples of novel metaphoric, literal, anomalous, and filler sentences together with nonliteral translations (in italics). The same sentences were used in Experiment 1 and Experiment 2.

Condition	Sentence context	Critical word
Novel metaphoric	Po chwili jej oczom ukazała się procesja	butelek.
	<i>After a while, she saw a procession of</i>	<i>bottles.</i>
	Zachwyciła go kolejna zgniła	teoria.
	<i>He was amazed by another moldy</i>	<i>theory</i>
	Mają same wykrochmalone	szużące.
	<i>They hire only starched</i>	<i>maids.</i>
Literal	Przelknął kolejną łyżkę	cierpienia.
	<i>He swallowed another spoon of</i>	<i>suffering.</i>
	Jego zwinne palce dźgały	powietrze
	<i>His agile fingers jabbed</i>	<i>the air.</i>
	Co noc słycać głośne tłuczenie	butelek.
	<i>Every night you can hear a loud sound of breaking</i>	<i>bottles.</i>
Anomalous	Istnieje ponoć jakaś nowa	teoria.
	<i>Supposedly, there is a new</i>	<i>theory.</i>
	Wtedy zwolnili wszystkie	szużące.
	<i>At that time, they fired all</i>	<i>maids.</i>
	Długo rozmawiali na temat	cierpienia.
	<i>They talked for a long time about</i>	<i>suffering.</i>
Filler sentences	Tam jest bardzo rozrzedzone	powietrze.
	<i>In there, there is very thin</i>	<i>air.</i>
	Kupiła na rynku pęczek dojrzałych	butelek.
	<i>In the market, she bought a bunch of ripe</i>	<i>bottles.</i>
	W pokoju stała stara, wygodna	teoria.
	<i>In the room, there stood an old, comfortable</i>	<i>theory.</i>
Filler sentences	Wyrwała z zeszytu dwie zapisane	szużące.
	<i>She tore out of the notebook two written-up</i>	<i>maids.</i>
	Kropłe wody kapwały z przemoczonego	cierpienia.
	<i>Water was dripping from soaked</i>	<i>suffering.</i>
	Na końcu nitki należy zrobić	powietrze.
	<i>At the end of the thread you should make</i>	<i>the air.</i>
Filler sentences	Po roku poparcie dla niego się	skurczyło.
	<i>After a year, the support for him</i>	<i>has shrunk.</i>
	Po południu wszystkie dzieci śpiewają i	rysują.
	<i>In the afternoon, all children sing and</i>	<i>draw.</i>
	Wiemy dokładnie co oni	robią.
	<i>We know exactly what they are</i>	<i>doing.</i>
Filler sentences	Te truskawki są bardzo	sprzeczne.
	<i>These strawberries are very</i>	<i>contradictory.</i>
	Szedł z zakupami pod	podeszwą.
	<i>He carried the shopping under his</i>	<i>sole.</i>

metaphors would be evaluated as more meaningful than the anomalous but less meaningful than the literal sentences. Sentences which were evaluated contrary to our expectations (with scores below 3.6 for anomalous sentences, above 2 for literal sentences, and above 3.5 for novel metaphoric sentences) were removed from the final set of stimuli. The analysis of the data for the remaining sentences revealed the main effect of sentence type [$F(2,66) = 585, p < .001, \eta_p^2 = 0.95$]. Bonferroni corrected pairwise comparisons showed significant differences [$p < .001$] between all three sentence types.

Third, the metaphoricity scale was used to ensure that novel metaphoric and literal sentences were perceived as intended by the experimenters. The third group of 34 independent judges (21 females, $M_{age} = 20$, range = 18–22) was asked to evaluate the sen-

tences on a 5 point metaphoricity scale ranging between very metaphoric (1) and very literal (5). Sentences which were evaluated contrary to our expectations (with scores below 3.6 for literal sentences and above 2.4 for novel metaphoric sentences) were removed from the final set of stimuli. As expected, participants judged novel metaphoric sentences as more metaphoric than literal sentences [$t(33) = 18.69, p < .001, r = 0.91$].

Finally, in order to check whether novel metaphors were perceived as less familiar than literal sentences, a 7-point familiarity scale was administered. The fourth group of 40 independent judges (25 females, $M_{age} = 21$, range = 19–24) was asked to decide how often they had come across a given sentence, with 1 representing very often and 7 very rarely. As expected, participants evaluated novel metaphoric sentences as less familiar than literal sentences [$t(39) = 15.22, p < .0001, r = 0.86$]. However, literal sentences were not perceived as highly familiar, as the mean score was close to the middle of the scale.

Overall, after the normative studies, the final set of stimuli (see Table 3) consisted of 75 novel metaphoric, 75 literal and 75 anomalous sentences (for a complete set of experimental sentences, see Supplementary materials). Moreover, 225 filler sentences were used as distractors from the experimental sentences, and were excluded from final analyses. Filler sentences differed in the syntactic structure from experimental sentences, but they were constructed to reflect the categories represented in the experimental sentences. Mimicking the experimental sentences, two third of the fillers were meaningful sentences, and one third of the filler sentences were meaningless. To avoid a literal bias, half of the meaningful sentences were conventional figurative utterances and half were literal.

4.1.3. Procedure

Procedures for Experiment 1 and Experiment 2 were approved by the Ethics Committee for Research with Human Participants at Adam Mickiewicz University in Poznan, Poland. Experiment 1 was performed in the Vision and Neuroscience Laboratory of the NanoBioMedical Centre at Adam Mickiewicz University in Poznań, Poland. All stimuli were presented on a 22-in LCD monitor with refresh rate 70 Hz, driven by Presentation Software (v. 16.4, Neurobehavioral Systems Inc.). Participants sat in a dark quiet room separate from the experimenter and the computers in a comfortable chair at a distance of 90 cm from the computer screen. Words were displayed in black font (Times New Roman, font size: 23) on a gray background. First, the fixation point (black square) was displayed on the computer screen for 500 ms, then the context of the sentence remained on the screen for 1800 ms, which was followed by a blank screen (500 ms) and another fixation point (300 ms). Finally, the critical word was displayed in the centre of the computer screen until the response was provided. If the participant did not respond within 1000 ms, a blank screen was displayed which lasted until the response was given and was followed by a 1000 ms inter-trial interval (a blank screen). Participants were asked to decide whether a given sentence makes or does not make sense. Responses to novel metaphoric and literal sentences were considered correct if participants indicated that the sentence made sense, while responses to anomalous sentences

Table 3

Mean meaningfulness ratings (where 1 = meaningful and 5 = meaningless), metaphoricity ratings (where 1 = metaphoric and 5 = literal), familiarity ratings (where 1 = very often and 7 = very rarely), and cloze probabilities are presented for literal, novel metaphoric, and anomalous sentences.

	Meaningfulness M (SD)	Metaphoricity M (SD)	Familiarity M (SD)	Cloze probabilities M (SD)
Literal sentences	1.71 (0.45)	4.31 (0.31)	3.68 (0.21)	0.01 (0.01)
Novel metaphoric sentences	3.14 (0.7)	1.71 (0.36)	5.61 (0.15)	0.01 (0.01)
Anomalous sentences	4.47 (0.27)	NA	NA	0.00 (0.00)

were considered correct if participants indicated that the sentence did not make sense.

Since each critical word was used in all three conditions, the experimental sentences were counterbalanced across 3 blocks so that each participant saw each critical word once and evaluated 150 sentences. The sentences were presented in random order. Each experimental session started with 2 practice blocks. Practice block 1 was designed to allow participants to practice manual responses. Letters M and W appeared on the computer screen. Participants were asked to press the down arrow if they saw M and left arrow if they saw W. Practice block 2 was designed to make participants familiar with the task used in the experiment. Participants were presented with sentence contexts (target sentences without the critical word) for 1800 ms, and when the critical word appeared on the screen, they were asked to decide whether the sentence made or did not make sense. The practice block consisted of 10 sentences, which represented all three conditions used in the experiments. None of these sentences was presented in the experimental block, which followed the two practice blocks. In order to avoid the movement related lateralization effects in the EEG data, the use of the right and left hand was counterbalanced across participants, so that half of the group executed all responses with their right, and half with their left hands.

4.2. Experiment 2

4.2.1. Participants

24 students of the Faculty of Physics at Adam Mickiewicz University in Poznań, Poland participated in Experiment 2 in exchange for course credits. Data from one participant were excluded from further analysis due to a low number of correct responses on the comprehension questions. Among the 23 participants, 11 were female and 12 were male ($M_{age} = 24$, range 21–26). All participants were right-handed native speakers of Polish, had normal or corrected to normal vision, and no history of neurological disorders.

4.2.2. Materials

The set of experimental and filler sentences was identical to the one used in Experiment 1.

4.2.3. Procedure

Experiment 2 was performed in the Vision and Neuroscience Laboratory of the NanoBioMedical Centre at Adam Mickiewicz University in Poznań, Poland. The procedure in Experiment 2 was the same as in Experiment 1, with the exception that Experiment 2 did not involve the semantic decision task. Here, participants were asked to read the sentences. In order to ensure that participants read and comprehended the sentences, comprehension questions were displayed on the computer screen following 30% of the sentences (24 comprehension questions). For example, after a target sentence *An innocent foreigner was arrested yesterday*, a blank screen was displayed for 2000 ms and a comprehension question *Was a drunk foreigner arrested yesterday?* appeared on the screen, to which the participant's expected answer was "No". The comprehension question remained on the screen until the participant gave a response. If no comprehension question followed a given sentence, the inter-trial interval (a blank screen) was displayed for 1000 ms, after which the next trial began. The order of the sentences was randomized. Responses were given by indicating "yes" or "no" using the down and left arrow keys on the keyboard. In order to avoid the lateralization effects in the EEG data related to movement preparation, the use of the right and left hand was counterbalanced across participants, so that half of the participants gave all responses with their right hands and half with their left hands. The EEG recording was the same as in Experiment 1.

4.2.4. Electrophysiological recording

EEG signals were recorded from 32 active Ag/AgCl electrodes (Brain Products): FP1, FP2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, TP9, CP5, CP1, CP2, CP6, TP10, P7, P3, Pz, P4, P8, PO9, O1, Oz, O2, PO10 at the standard extended 10/20 positions with the ground placed at AFz. For blinks artifacts, the electrooculography was recorded bipolarly from above and below the right eye (vEOG). For horizontal eye movements bipolar electrooculography was used, situated horizontally from positions next to the outer rims of the eyes. EEG data were amplified by a QuickAmp (BrainProducts GmbH) with average reference and filtered with lowpass filters (cutoff frequency = 80 Hz) and with highpass filters (cutoff frequency = 0.015 Hz). The sampling rate was 250 Hz. Electrode impedance were maintained below 5 k Ω for all electrodes. All signals were stored on the computer for offline analyses.

Offline EEG analyses were conducted using Brain Vision Analyzer 2.0 software. Data were bandpass filtered from 0.5 Hz (24 dB/oct.) to 30 Hz (24 dB/oct.) using the Butterworth zero phase filter. They were segmented from 300 ms before stimulus onset to 1000 ms afterward, referred to baseline – 100–0 ms before stimulus onset, and edited for artifacts (rejecting trials with zero lines, voltage differences higher than 150 μ V or voltage steps higher than 50 μ V). Ocular artifacts were corrected by the Gratton and Coles method as implemented in Brain Vision Analyzer 2.0.

4.4. Analysis of the behavioral and electrophysiological data

The analysis of reaction times was performed on correctly classified trials. No premature (RTs below 300 ms) or too slow (RTs above 3500 ms) responses were observed. Five participants were excluded from the final analysis in Experiment 1 due to a low number of correct responses on the semantic decision task, and one participant was excluded from the final analysis in Experiment 2 due to a low number of correct responses to the comprehension questions.

Nine electrodes were used in the analysis. For the anterior-posterior factor we selected F3, Fz, F4 (frontal); C3, Cz, C4 (central); and P3, Pz, P4 (parietal). The laterality factor included F3, C3, P3 (left), Fz, Cz, Pz (midline), F4, C4, P4 (right). The inspection of the EEG data revealed the N100 and P200, the exogenous perceptual components, over frontal and central sites. Next, negativity was observed around 200 ms after critical word onset over posterior sites and around 400 ms over central and posterior sites. Finally, a positive component peaking around 600 ms (with the onset around 500 ms and offset around 800 ms) was observed over parietal sites. Based on previous studies on metaphor comprehension and the visual inspection of the data gathered in the current study, we selected the N400 (between 300 and 500 ms) and the LPC component (500–800 ms) for the statistical analysis. To avoid possible bias discussed by Van Rullen (2011), and caused by brain activity associated with correct responses rather than stimulus category, all responses were entered in the analysis. It also increased the signal to noise ratio in Experiment 1, which made the comparison between the two experiments more meaningful. Since, based on previous studies (Arzouan et al., 2007a; Goldstein et al., 2012), we predicted a graded N400 effect, one-tailed *t*-tests were employed to test the differences between the three sentence types. For the LPC amplitudes we did not have a directional hypothesis, for which reason Bonferonni corrected pairwise comparisons were conducted. For all analyses, the Greenhouse–Geisser correction was applied, where appropriate. Partial eta squared (η_p^2) is reported as a measure of effect size.

In Section 2.3. (Comparison of Experiment 1 and Experiment 2), results of analyses are reported on rescaled data of difference waves to draw conclusions regarding topographical differences between conditions and experiments. We decided to employ the

McCarthy and Wood (1985) vector rescaling procedure as this method can be used for both between and within-subjects comparisons, although other methods may be superior when dealing solely with within-subjects comparisons (Jing et al., 2006). Urbach and Kutas (2002) indicated that the rescaling method of McCarthy and Wood may encounter problems due to baseline differences between conditions and the influence of noise. Wilding (2006) indicated that these concerns are reduced considerably when vector scaling is computed over difference scores, which is the approach taken here.

Acknowledgements

The work of Karolina Rataj was funded by the project ConCreTe. The project ConCreTe acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under H2020 Future and Emerging Technologies grant number 611733.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.brainres.2017.10.030>.

References

- Arzouan, Y., Goldstein, A., Faust, M., 2007a. Brainwaves are stethoscopes: ERP correlates of novel metaphor comprehension. *Brain Res.* 1160, 69–81. <https://doi.org/10.1016/j.brainres.2007.05.034>.
- Arzouan, Y., Goldstein, A., Faust, M., 2007b. Dynamics of hemispheric activity during metaphor comprehension: electrophysiological measures. *Neuroimage* 36 (1), 222–231. <https://doi.org/10.1016/j.neuroimage.2007.02.015>.
- Bentin, S., McCarthy, G., Wood, C.C., 1985. Event-related potentials, lexical decision and semantic priming. *Electroencephalogr. Clin. Neurophysiol.* 60 (4), 343–355.
- Bornkessel-Schlesewsky, I., Schlesewsky, M., 2008. An alternative perspective on “semantic P600” effects in language comprehension. *Brain Res. Rev.* 59 (1), 55–73. <https://doi.org/10.1016/j.brainresrev.2008.05.003>.
- Bowlde, B.F., Gentner, D., 2005. The career of metaphor. *Psychol. Rev.* 112 (1), 193–216. <https://doi.org/10.1037/0033-295X.112.1.193>.
- Brouwer, H., Fitz, H., Hoeks, J., 2012. Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain Res.* 1446, 127–143. <https://doi.org/10.1016/j.brainres.2012.01.055>.
- Coulson, S., Van Petten, C., 2002. Conceptual integration and metaphor: an event-related potential study. *Mem. Cognit.* 30 (6), 958–968.
- Coulson, S., Van Petten, C., 2007. A special role for the right hemisphere in metaphor comprehension? ERP evidence from hemifield presentation. *Brain Res.* 1146, 128–145. <https://doi.org/10.1016/j.brainres.2007.03.008>.
- De Grauwe, S., Swain, A., Holcomb, P.J., Ditman, T., Kuperberg, G.R., 2010. Electrophysiological insights into the processing of nominal metaphors. *Neuropsychologia* 48 (7), 1965–1984. <https://doi.org/10.1016/j.neuropsychologia.2010.03.017>.
- De Long, K.A., Quante, L., Kutas, M., 2014. Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia* 61, 150–162. <https://doi.org/10.1016/j.neuropsychologia.2014.06.016>.
- Friederici, A.D., Mecklinger, A., 1996. Syntactic parsing as revealed by brain responses: first-pass and second-pass parsing processes. *J. Psycholinguist. Res.* 25 (1), 157–176. <https://doi.org/10.1007/Bf01708424>.
- Gibbs, R.W., Bogdanovich, J.M., Sykes, J.R., Barr, D.J., 1997. Metaphor in idiom comprehension. *J. Mem. Lang.* 37 (2), 141–154. <https://doi.org/10.1006/jmla.1996.2506>.
- Glucksberg, S., Gildea, P., Bookin, H.B., 1982. On understanding non-literal speech – can people ignore metaphors. *J. Verb. Learn. Verb. Behav.* 21 (1), 85–98. [https://doi.org/10.1016/S0022-5371\(82\)90467-4](https://doi.org/10.1016/S0022-5371(82)90467-4).
- Goldstein, A., Arzouan, Y., Faust, M., 2012. Killing a novel metaphor and reviving a dead one: ERP correlates of metaphor conventionalization. *Brain Lang.* 123 (2), 137–142. <https://doi.org/10.1016/j.bandl.2012.09.008>.
- Grice, H.P., 1975. *Logic and conversation*. In: Cole, P., Morgan, J. (Eds.), *Syntax Semantics*. Academic Press, New York, pp. 41–58.
- Hagoort, P., Brown, C., Groothusen, J., 1993. The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Lang. Cognit. Proc.* 8 (4), 439–483. <https://doi.org/10.1080/01690969308407585>.
- Inhoff, A.W., Lima, S.D., Carroll, P.J., 1984. Contextual effects on metaphor comprehension in reading. *Mem. Cognit.* 12 (6), 558–567.
- Janus, R.A., Bever, T.G., 1985. Processing of metaphoric language – an investigation of the 3-stage model of metaphor comprehension. *J. Psycholinguist. Res.* 14 (5), 473–487. <https://doi.org/10.1007/Bf01666722>.
- Jing, H., Pivik, R.T., Dykman, R.A., 2006. A new scaling method for topographical comparisons of event-related potentials. *J. Neurosci. Methods* 151 (2), 239–249. <https://doi.org/10.1016/j.jneumeth.2005.08.002>.
- Kaan, E., Harris, A., Gibson, E., Holcomb, P., 2000. The P600 as an index of syntactic integration difficulty. *Lang. Cognit. Proc.* 15 (2), 159–201.
- Kazmerski, V.A., Blasko, D.G., Dessalegn, B.G., 2003. ERP and behavioral evidence of individual differences in metaphor comprehension. *Mem. Cognit.* 31 (5), 673–689.
- Kröger, S., Rutter, B., Stark, R., Windmann, S., Hermann, C., Abraham, A., 2012. Using a shoe as a plant pot: neural correlates of passive conceptual expansion. *Brain Res.* 1430, 52–61. <https://doi.org/10.1016/j.brainres.2011.10.031>.
- Kuperberg, G.R., 2007. Neural mechanisms of language comprehension: challenges to syntax. *Brain Res.* 1146, 23–49. <https://doi.org/10.1016/j.brainres.2006.12.063>.
- Kuperberg, G.R., Kreher, D.A., Sitnikova, T., Caplan, D.N., Holcomb, P.J., 2007. The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain Lang.* 100 (3), 223–237.
- Kutas, M., Hillyard, S.A., 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307 (5947), 161–163.
- Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>.
- Lai, V.T., Curran, T., 2013. ERP evidence for conceptual mappings and comparison processes during the comprehension of conventional and novel metaphors. *Brain Lang.* 127 (3), 484–496. <https://doi.org/10.1016/j.bandl.2013.09.010>.
- Lai, V.T., Curran, T., Menn, L., 2009. Comprehending conventional and novel metaphors: an ERP study. *Brain Res.* 1284, 145–155. <https://doi.org/10.1016/j.brainres.2009.05.088>.
- McCarthy, G., Wood, C.C., 1985. Scalp distribution of event-related potentials: an ambiguity associated with analysis of variance models. *Electroencephalogr. Clin. Neurophysiol.* 62, 203–208.
- Onishi, K.H., Murphy, G.L., 1993. Metaphoric reference: when metaphors are not understood as easily as literal expressions. *Mem. Cognit.* 21 (6), 763–772.
- Osterhout, L., Holcomb, P.J., Swinney, D.A., 1994. Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *J. Exp. Psychol. Learn. Mem. Cogn.* 20 (4), 786–803.
- Paller, K.A., Kutas, M., 1992. Brain potentials during memory retrieval provide neurophysiological support for the distinction between conscious recollection and priming. *J. Cogn. Neurosci.* 4 (4), 375–392. <https://doi.org/10.1162/jocn.1992.4.4.375>.
- Przepiórkowski, A., 2004. *The IPI PAN Corpus: Preliminary Version*. Institute of Computer Science, Polish Academy of Sciences, Warsaw.
- Pynte, J., Besson, M., Robichon, F.H., Poli, J., 1996. The time-course of metaphor comprehension: an event-related potential study. *Brain Lang.* 55 (3), 293–316. <https://doi.org/10.1006/brln.1996.0107>.
- Rugg, M.D., Nagy, M.E., 1987. Lexical contribution to nonword-repetition effects: evidence from event-related potentials. *Mem. Cognit.* 15 (6), 473–481.
- Rutter, B., Kröger, S., Hill, H., Windmann, S., Hermann, C., Abraham, A., 2012. Can clouds dance? Part 2: an ERP investigation of passive conceptual expansion. *Brain Cogn.* 80 (3), 301–310. <https://doi.org/10.1016/j.bandc.2012.08.003>.
- Steinhauer, K., Drury, J.E., 2012. On the early left-anterior negativity (ELAN) in syntax studies. *Brain Lang.* 120 (2), 135–162. <https://doi.org/10.1016/j.bandl.2011.07.001>.
- Steinhauer, K., Drury, J.E., Portner, P., Walenski, M., Ullman, M.T., 2010. Syntax, concepts, and logic in the temporal dynamics of language comprehension: evidence from event-related potentials. *Neuropsychologia* 48 (6), 1525–1542. <https://doi.org/10.1016/j.neuropsychologia.2010.01.013>.
- Stróżak, P., Abedzadeh, D., Curran, T., 2016a. Separating the FN400 and N400 potentials across recognition memory experiments. *Brain Res.* 1635, 41–60. <https://doi.org/10.1016/j.brainres.2016.01.015>.
- Stróżak, P., Bird, C.W., Corby, K., Frishkoff, G., Curran, T., 2016b. FN400 and LPC memory effects for concrete and abstract words. *Psychophysiology* 53 (11), 1669–1678. <https://doi.org/10.1111/psyp.12730>.
- 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 (3), 488–509. <https://doi.org/10.1006/brln.2001.2610>.
- Urbach, T.P., Kutas, M., 2002. The intractability of scaling scalp distributions to infer neuroelectric sources. *Psychophysiology* 39, 791–808.
- van Herten, M., Kolk, H.H., Chwilla, D.J., 2005. An ERP study of P600 effects elicited by semantic anomalies. *Cogn. Brain Res.* 22 (2), 241–255. <https://doi.org/10.1016/j.cogbrainres.2004.09.002>.
- Van Rullen, R., 2011. Four common conceptual fallacies in mapping the time course of recognition. *Front Psychol.* 2, 365. <https://doi.org/10.3389/fpsyg.2011.00365>.
- Vilberg, K.L., Rugg, M.D., 2009. Functional significance of retrieval-related activity in lateral parietal cortex: Evidence from fMRI and ERPs. *Hum. Brain Mapp.* 30 (5), 1490–1501. <https://doi.org/10.1002/hbm.20618>.
- Voss, J.L., Federmeier, K.D., 2011. FN400 potentials are functionally identical to N400 potentials and reflect semantic processing during recognition testing. *Psychophysiology* 48 (4), 532–546. <https://doi.org/10.1111/j.1469-8986.2010.01085.x>.
- Weiland, H., Bambini, V., Schumacher, P.B., 2014. The role of literal meaning in figurative language comprehension: evidence from masked priming ERP. *Front. Hum. Neurosci.* 8, 583. <https://doi.org/10.3389/fnhum.2014.00583>.

- Wilding, E.L., 2006. The practice of rescaling scalp-recorded event-related potentials. *Biol. Psychol.* 72, 325–332. <https://doi.org/10.1016/j.biopsycho.2005.12.002>.
- Wolff, P., Gentner, D., 2000. Evidence for role-neutral initial processing of metaphors. *J. Exp. Psychol. Learn. Mem. Cogn.* 26 (2), 529–541.
- Xu, H., Zhang, Q., Li, B., Guo, C., 2015. Dissociable effects of valence and arousal on different subtypes of old/new effect: evidence from event-related potentials. *Front. Hum. Neurosci.* 9, 650. <https://doi.org/10.3389/fnhum.2015.00650>.