Basic composition and enriched integration in idiom processing: an EEG study

Paolo Canal, Francesca Pesciarelli, Francesco Vespignani, Nicola Molinaro & Cristina Cacciari

Author note: Paolo Canal, Laboratorio di Linguistica Giovanni Nencioni, Scuola Normale Superiore, Pisa, and Center for Neurocognition, Epistemology and theoretical Syntax (NEtS), Institute for Advanced Study (IUSS), Pavia; Paolo Canal, Francesca Pesciarelli and Cristina Cacciari, Department of Biomedical, Metabolical and Neural Sciences, Università degli Studi di Modena e Reggio Emilia, Modena; Francesco Vespignani, Department of Cognition and Formation Sciences, Università degli Studi di Trento, Rovereto; Nicola Molinaro, Basque Center on Cognition, Brain and Language (BCBL), Donostia and Ikerbasque, Basque Foundation for Science, Bilbao, Spain.
This research was carried out at Università di Modena e Reggio Emilia. The experiment was funded by the Italian Ministry of Education, Universities and Research (MIUR) through a grant for "projects of national interest" (PRIN n. 2005119758_003). We are grateful to Chiara Reali and Alice Montanaro for their help in the construction and norming of the experimental materials and in running the ERP experiment. Correspondence concerning this article should be addressed to Paolo Canal, Laboratorio di Linguistica Giovanni Nencioni, Scuola Normale Superiore, p.zza dei Cavalieri 7, 56126 Pisa, Italy. E-mail: paolo.canal@iusspavia.it

Running head: semantic composition in idiom processing

Abstract

We investigated the extent to which the literal meanings of the words forming literally-plausible idioms (e.g., *break the ice*) are semantically composed, and how the idiomatic meaning is integrated in the unfolding sentence representation. Participants read ambiguous idiom strings embedded in highly predictable, literal and idiomatic contexts, while their EEG was recorded. Control sentences only contained the idiom-final word whose cloze values were as high as in literal and idiomatic contexts. Event Related Potentials data showed that differences in the amplitude of a Frontal positivity (PNP) emerged at the beginning and at the end of the idiom strings, with the Idiomatic context condition associated with more positive voltages. The Time-Frequency analysis of the EEG showed an increase in power of the middle gamma frequency band only in the Literal context condition. These findings suggest that sentence revision mechanisms, associated with the frontal PNP, are involved in idiom meaning integration, and that the literal semantic composition of the idiomatic constituents, associated with changes in gamma frequency, is not carried out after idiom recognition.

Keywords

ERPs; Frontal PNP; Middle gamma Frequency band; semantic composition and meaning integration in idioms; Configuration Hypothesis;

1. Introduction

Idioms (e.g., *break the ice, bark up the wrong tree*) are strings of words with a conventional, figurative meaning and a default structure. Idiomatic meanings are stored in semantic memory, but how this affects the cognitive mechanisms responsible for idiom processing is still at issue. On the one hand, the processing of the idiomatic constituents is often speeded up with respect to matched literal constituents in lexical decision tasks (Swinney & Cutler, 1979), in sentence verification tasks (Pesciarelli et al., 2014) or in reading comprehension tasks (Siyanova-Chanturia, Conklin & Schmitt, 2011; Vespignani, Canal, Molinaro, Fonda & Cacciari, 2010), as if idiomatic expressions were simply retrieved from semantic memory without any attempts to compose the meanings of the individual words. On the other hand, the processing time advantage of idioms, compared to literal sentences, has not been systematically replicated with all idiom types and methodologies (e.g., Tabossi Fanari & Wolf, 2009; Cacciari, Padovani & Corradini, 2007). Moreover, fMRI evidence attested a more widespread activation of the language network for idioms, extending to prefrontal areas (Boulenger, Hauk & Pulvermüller, 2009; Romero Lauro, Tettamanti, Cappa & Papagno, 2008; Zempleni, Haverkort, Renken & Sotwe, 2007; for a meta-analysis, see Bohrn, Altmann & Jacobs, 2012), that could be due to several factors including the competition engendered by the literal meaning and the figurative meaning of the expression (Romero Lauro et al., 2008). In fact, the literal meaning of the idiom's constituent words is activated at least until the string is recognized as an idiom (Cacciari, 2014; Cacciari & Corradini, 2015; Vespignani et al., 2010). After recognition, the idiomatic meaning is retrieved from semantic memory and may conflict or interact with the literal meaning of the idiom string. Unfortunately, because of the low temporal resolution of fMRI measures, brain-imaging studies cannot help in disentangling how these two sources of information are accessed and in particular whether or not they are accessed in parallel (but see the MEG study of Boulenger, Shtyrov & Pulvermüller, 2012). In contrast, the investigation of the electrical activity of the brain can provide evidence on whether and how semantic composition and idiom retrieval coexist during comprehension. Providing new evidence on this aspect of idiom processing is the general aim of the present study.

Language comprehension is assumed to be compositional (e.g., Frege, 1892; Pylkkänen, 2008): sentence comprehension proceeds combining the default meaning of individual constituents according to the morpho-syntactic rules of a language. Idioms (e.g., *kick the bucket*) challenge this assumption for several reasons: first, when the idiom has two plausible interpretations, the meanings of the idiom's constituent words can be used to derive a literal interpretation (*kick a pail*) that differs from the idiomatic one (*die*); second, in semantically transparent idioms (e.g., *sleep like a log*) the literal meanings of the constituent words is related to the idiomatic interpretation and can help in determining the global meaning of the string; third, the words forming idiomatic configurations preserve their own lexico-semantic properties while forming at the same time larger units (Molinaro & Carreiras, 2010); lastly, at least part of the literal meaning of the idiom words is activated before the idiom is recognized as such. In a nutshell, idioms seem to behave at the same time compositionally and non-compositionally (Sprenger, Levelt & Kempen, 2006; Cacciari & Corradini, 2015; Boulenger et al., 2012; Holsinger & Kayser, 2013).

How do literal computation and idiomatic meaning retrieval and integration interact during comprehension? The literature has not yet provided an unequivocal answer (for an overview of the behavioral literature, see Cacciari & Corradini, 2015;

Cacciari, 2014). It has been suggested that when processing frequently co-occurring multi-word strings, individual word meanings may not be necessarily computed (Arnon & Snider, 2010; Tremblay & Baayen, 2010). Since multi-word strings are stored as such in semantic memory, in production they "could be easily retrieved and used as wholes without the need to compose them online through word selection and grammatical sequencing" (Tremblay & Baayen, 2010; p.151). However, this may not necessarily apply to idioms and notably to idiom comprehension. In fact, one of the tenets of recent models of idiom processing (e.g., Cacciari & Tabossi, 1988; Sprenger et al., 2006; Vespignani et al., 2010; for an overview see Cacciari 2014) is that at least part of the literal meaning of the idiom's constituent words is activated during idiom comprehension: behavioral evidence has accumulated showing that at least partial literal meaning activation is obligatory and typically precedes idiom retrieval (Cacciari & Corradini, 2015; Holsinger & Kayser, 2013), although it has been shown that a full analysis of the literal meaning of idiomatic constituents may not be carried out (Rommers, Bastiaansen & Dijkstra, 2013; Peterson, Burgess, Dell & Eberhard, 2001). This EEG experiment was devised to further explore the real-time relationship between the composition of the literal meaning of the constituent words and the activation and sentence integration of the figurative meaning.

1.1 Processing idioms: compositionality and other factors

Lexical look-up models were the first to propose that idioms are semantically empty long words directly retrieved from the mental lexicon (Bobrow & Bell, 1973; Gibbs, 1980; Swinney & Cutler, 1979). The most influential hypothesis of this approach, the Lexical Representation Hypothesis (Swinney & Cutler, 1979), posited that idiom retrieval starts at the beginning of the string and runs in parallel with the computation of the literal meaning. Since computing takes longer than retrieving, the idiomatic meaning becomes available before the literal meaning. This model was questioned by several psycholinguistic studies showing that the literal meanings of the idiom's constituent words are to some extent activated (e.g., Cacciari & Tabossi, 1988; Titone & Connine, 1994); and that many idioms can be internally modified (e.g., Gibbs, Nayak & Cutting, 1989) preserving the idiomatic meaning. The starting point of more recent Non-Lexical Hybrid models is that the speaker/reader does not know in advance that what she is currently processing is an idiom until enough information become available to the processing system. In real-time, idiom retrieval occurs only after the reader/listener has recognized that the linguistic input she is processing is (or is highly expected to be) a known idiom (Cacciari & Tabossi 1988; Vespignani et al., 2010). This was initially posited by the Configuration Hypothesis (CH, Cacciari & Tabossi 1988) and then was confirmed by several studies (Cutting & Bock, 1997; Libben & Titone, 2008; Sprenger et al., 2006; Titone & Connine, 1994; Vespignani et al., 2010; Cacciari & Corradini, 2015; Holsinger & Kayser, 2013).

According to the CH, idioms are processed word by word until enough information has accumulated to render the sequence of words identifiable as – or highly expected to be – a memorized idiom. At that point the idiom is recognized. In fact, when the initial fragment of a string triggers high expectancy about a final idiomatic conclusion, recognition of a word providing an unexpected (literal) ending is slowed down (Tabossi, Fanari & Wolf, 2005). The point at which the idiom is identified depends on the idiom characteristics (e.g., length, familiarity, semantic transparency, ambiguity) and on previous context, and is usually defined as the word after which the expression becomes highly predictable based on cloze probability values. When the presence of an idiom is perceived before the string offset and the configuration is activated, also the remaining constituent(s) become highly predictable (Vespignani et al., 2010; Molinaro & Carreiras, 2010; Molinaro, Barraze, & Carreiras, 2013) and the following input is compared with the pre-stored idiom representation in template matching process (Kok, 2001) aimed at monitoring whether or not it is the expected item (Vespignani et al., 2010; Molinaro & Carreiras, 2010). In sum, the recognition of the idiomatic string determines a qualitative change in processing (Vespignani et al., 2010) that makes the composition of the constituents post idiom recognition potentially irrelevant (Peterson et al., 2001; Rommers et al., 2013).

Once the idiom meaning is retrieved from long-term semantic memory, it must be integrated in the sentence representation to form a semantically coherent structure. Current psycholinguistic models of idiom processing (including the CH) are silent on whether idiom meaning integration occurs right after recognition, or whether it requires further time and processing effort. Pragmatic models of idiom comprehension, for instance Vega-Moreno (2001), along the lines of the Relevance Theory approach (e.g., Sperberg & Wilson, 1995) point out that idiom interpretation requires the integration of multiple sources of information: when an utterance contains an idiom the hearer has access "to the concepts encoded by the words in the idiom, to the concept encoded by the idiom as a whole, and often also to a pragmatic routine for bridging the gap between the compositional meaning and the idiomatic meaning (Vega-Moreno, 2001; p.415)". But again how and when the integration of the idiomatic meaning in the sentential representation occurs remains an open question.

Compared to literal language processing, idioms comprehension may seem cognitively effortless because their meaning is represented in the mental lexicon and their comprehension hinges on retrieval, rather than on semantic composition mechanisms. However, brain imaging and neuropsychological evidence have shown a different picture. For instance, language-impaired populations (schizophrenic, brain damaged and Alzheimer disease patients) perform poorly in the comprehension of idioms, especially when they also possess a well-formed literal interpretation (for a review see Cacciari & Papagno, 2010). Several factors concur in making idioms cognitively more complex than matched literal language: their meaning is figurative, it is often ambiguous, it may be unrelated to its literal meaning, and it often refers to complex abstract, mental and affective states (Citron et al., 2016).

Compared to other kinds of figurative expressions, idioms may be the least cognitively demanding instances. However, neuroimaging evidence suggests that each kind of figurative language poses specific demands, with different patterns of brain activation associated with the comprehension of idioms, metaphors and irony (Bohrn et al., 2012). For instance, the neural circuitry underlying idiom comprehension at the same time partially overlaps (in Left Inferior Frontal Gyrus) with the network associated with metaphor processing (when both are compared with literal language), and shows idiom-specific activations extending to the dorsolateral prefrontal cortex. The involvement of prefrontal cortices attested by fMRI (e.g., Romero Lauro et al., 2008; and evidence reviewed in Bohrn et al., 2012) and TMS (e.g., Häuser, Titone & Baum, in press; Rizzo, Sandrini & Papagno, 2007) studies points to a crucial role of cognitive control mechanisms in idioms comprehension. As Häuser et al. (in press) argued, the conflict between literal and figurative meanings may require "controlled selection among competing representations".

1.2 The electrophysiology of idiom comprehension

Event-Related Potentials (ERP) and Time Frequency representations (TFR) are complementary aspects of the Electroencephalogram (EEG) recorded from the scalp.

These measures can be used to identify the cognitive mechanisms underlying the processing of an input with millisecond precision. ERPs are microvolt changes of the EEG that are triggered by an external stimulation or a cognitive event. The shape of the ERP brainwaves is determined by the summation of underlying components that reflect the neural activity associated with a particular computational operation (e.g., Luck, 2014). When analyzing the ERPs, important information about induced changes in the underlying EEG *oscillations* (specifically, in the modulation of power amplitude or phase synchrony of different frequency bands) is nearly canceled by the averaging procedure because such changes are not strictly phase-locked to stimulus presentation. These oscillations represent patterns of synchronization of populations of neurons (e.g., Varela, Lachaux, Rodriguez, & Martinerie, 2001) and provide relevant information for studying cognition because they may *change* as a function of stimulus presentation (for an overview see Bastiaansen, Mazaheri & Jensen, 2012) thus providing additional information about the cognitive processes underlying language comprehension.

ERPs have been widely used to investigate the temporal dynamics of language processing in general, and of idioms as well (e.g., Liu, Li, Shu, Zhang & Chen, 2010; Moreno, Federmeier & Kutas, 2002; Vespignani et al., 2010; Rommers et al., 2013; Zhou, Zhou & Chen, 2004). ERP studies on idiom processing have been mostly focused on the mechanisms giving raise to the N400 component, a negative deflection of the brainwaves, distributed in central parietal scalp sites and emerging between 300 and 500 ms (Kutas & Hillyard, 1980; for a review see Kutas & Federmeier, 2011). The N400 is elicited by any meaningful stimulus and is typically larger when words are less predictable in context or are semantically anomalous. Idiom studies in fact found a larger N400 when idiom-inconsistent words replaced idiom-consistent words (e.g., *kick the pail vs. kick the bucket*) (e.g., Liu, Li, Shu, Zhang & Chen, 2010; Moreno, Federmeier & Kutas, 2002; Vespignani et al., 2010; Zhou, Zhou & Chen, 2004). In sum, given the conventionalized, bound nature of idiom strings, when they are familiar idioms "provide a degree of context and cloze probability significantly beyond that of literal statements" (Strandburg et al., 1997, p. 605), which is reflected in the reduction of the N400. Effects on the P300 component have also been reported: the P300 is generally associated with cognitive mechanisms of context updating (Donchin & Coles, 1988) or context closure (Verleger, 1988). Vespignani et al (2010) investigated the electrical correlates of anticipatory mechanisms with unambiguous idiom strings predictable before the string offset. The brain's electrical response to the canonical idiom constituents was different if recorded before or after the idiom recognition point (RP, e.g., *prendere il toro per le*$_{RP}$ *... corna -- take the bull by the*$_{RP}$ *... horns*). Before the idiom recognition point, a reduction of the amplitude of the N400 was observed. In contrast, after idiom recognition, the categorical match with the expected, appropriate idiom word was associated with a P300 (for similar results with other predictable strings see Bornkessel-Schlesewsky et al., 2015; Molinaro & Carreiras, 2010; Roehm, Bornkessel-Schlesewsky, Rösler, & Schlesewsky, 2007). Coherently with the CH, this change was explained assuming that after the recognition of a configuration of words readers can predict the upcoming input categorically matching it with the string retrieved from semantic memory and not simply probabilistically, as typically occurs in sentences composed by freely occurring words.

Recently Rommers, Bastiaansen and Dijkstra (2013) investigated the N400 in the processing of idioms (e.g., *tegen de lamp*, lit. to walk against the lamp, meaning *to get caught*) with a more elaborate paradigm than prior EEG studies. Dutch speaking participants were presented with literally or idiomatically biasing sentence fragments (e.g., *het nieuwe peertje in de ..., the new light bulb into the...* ; *frauder uiteindelijk* *tegen de… , eventually walked against the...*) followed by the expected word (*lamp* –the idiom-final word in idiomatic contexts) or by a semantically related or unrelated word (*kaars*, *candle*, vs. *vis, fish*). Faster lexical decision times were obtained for expected words (*lamp*) than for semantically related or unrelated targets (*kaars* or *vis*), regardless of the literal or idiomatic nature of preceding contexts. Notably, related targets were faster than unrelated targets only in literal contexts. This finding was mirrored by a semantic relatedness effect on the N400 component (Federmeier & Kutas, 1999): in literal and idiomatic contexts, the amplitude of the N400 was larger for semantically related targets (*kaars*) than for expected words (*lamp*). However, the amplitude of the N400 for unrelated targets (*vis*) was larger than for related targets (*kaars*) only in literal contexts. According to Rommers and colleagues, these results, and those emerging from analyses of the frequency domain of the EEG signal, show that the language system may not engage in the process of integrating (unifying) individual word meanings in idiomatic sentences. Since the literal meaning of the idiom words is generally unnecessary to compute the figurative meaning, the language system may not engage, or may only partially engage, in semantically unifying individual word meanings while processing idioms (where semantic unification refers to the integration of word meanings by combining them into larger units, Hagoort, 2005; 2013). However, these conclusions hinge on the comparison between the related and the unrelated conditions, without any comparisons between the waveforms associated to the most expected word in literal and idiomatic contexts. In sum, Rommers et al.'s study did not provide a direct test of the ERP changes associated with the presentation of the expected word in these two contexts. A direct comparison between different types of processing of the same words may instead reveal additional differences related to semantic composition, necessary to literal sentences, and figurative meaning integration, necessary to idiomatic sentences. Therefore in this study we compared the same target word(s) embedded in highly constraining idiomatic and literal contexts.

1.3. Brain electrical correlates of semantic-pragmatic integration: N400 or P600?

The ERP literature on language processing has long debated about the brain correlates of semantic-pragmatic integration processes. The N400 component has been linked to semantic integration/unification processes that combine word meanings and world knowledge with the ongoing mental representation of the sentence (e.g., Brown & Hagoort, 1993; van den Brink, Brown & Hagoort, 2006; Hagoort & Van Berkum, 2007; Hagoort 2005; 2013). However, a substantial part of the N400 component is associated with the access to stored lexical and conceptual representations (e.g., Federmeier & Kutas, 2011) and it is now widely accepted that the N400 cannot *solely* reflect post-access processes (Lau, Phillips & Poeppel, 2008).

Recently, the *Retrieval-Integration hypothesis* (Brouwer, Fitz & Hoeks, 2012; Brouwer & Hoeks, 2013) has proposed that the neural basis of word processing involves a loop between lexical retrieval, reflected in the N400 component, and the integration of retrieved information with prior context, associated with effects on the P600 component. Although a large bulk of evidence has shown that syntactic processing difficulties affect the P600 (e.g., Osterhout & Holcomb, 1992), recently it has been shown that this component is also sensitive to semantic factors (e.g., Hoeks, Stowe & Doedens, 2004; Kim & Osterhout, 2005; Kuperberg, 2007). In sum, the processes engendering the P600 may be more general, operating on both syntactic and semantic information (for a detailed discussion, see Bornkesseln-Schlesewsky & Schlesewsky, 2008; Friederici, 2011). Moreover, the P600 component does not reflect a unitary phenomenon and is better conceived as a family of components with different scalp distributions and functional roles (e.g., a parietal Repair P600, a fronto-central Reanalysis P600, Kaan & Swaab, 2003).

Interestingly, semantic-pragmatic integration processes have also been found to enhance positive shifts that follow the N400. For instance, Late Positive Complex (LPC) effects have been reported in studies on metaphors (e.g., Coulson & Van Petten, 2002; Coulson & Van Petten, 2007; De Grauwe, Swain, Holcomb, Ditman & Kuperberg, 2010; Weiland, Bambini & Schumacher, 2014), jokes (e.g., Coulson & Kutas, 2001) and proverbs (Ferretti, Schwint & Katz, 2007). P600 effects have been described for irony processing (Filik, Leuthold, Wallington & Page, 2014; Regel, Gunter & Friederici, 2011; Regel, Meyer, & Gunter, 2014; Spotorno, Cheylus, Van Der Henst & Noveck, 2013). This evidence suggests that when the composition of the literal meaning is not sufficient to derive the appropriate figurative meaning and an additional effort is required to elaborate the contextually appropriate interpretation, the associated ERPs typically show different types of positive effects. These positivities have been interpreted as reflecting different phenomena, including reanalysis of word meanings (e.g., De Grauwe et al., 2010), pragmatic adjustment (Weiland et al., 2014) and pragmatic interpretation (Regel et al., 2011). These "enriched meaning integration" processes, as we define them, are necessary to achieve a coherent sentential representation whenever the literal composition of the word meanings is not enough to derive the appropriate interpretation.

When meaning integration does not proceed smoothly, and/or when predictions on upcoming words are disconfirmed by the arrival of an unexpected yet plausible word, another type of ERP positivity emerges, the post-N400 Positivity (PNP) (for an overview see Van Petten & Luka, 2012). The larger N400 on unexpected words is followed by a PNP on frontal electrodes that has been interpreted as a signal of the cost of correcting a misprediction (e.g., Federmeier, Wlotko, De Ochoa-Dewald & Kutas, 2007; Kutas, 1993; Thornhill & Van Petten, 2012). DeLong, Quante and Kutas (2014) proposed that frontal PNPs may reflect the inhibitory processes needed to override the pre-activation of the most expected sentence completion. However, two recent studies have suggested that plausibility, beyond predictability, plays a role in eliciting the frontal positive effects. Using words with medium-low levels of predictability in context, Molinaro, Carreiras and Duñabeitia (2012) observed frontal PNPs to target words (adjectives in noun-adjective pairs) that provided redundant information (*monstruo horrible, horrible monster*), or contrasting information (*monstruo hermoso, beautiful monster*), compared to neutral word pairs (*monstruo solitario, lonely monster*). The more unnatural the sentences were, the larger was the positivity. Brothers, Swaab and Traxler (2015) directly tested the impact of misprediction on the PNP asking participants to explicitly judge whether the sentence-final words matched or not their prediction. The waveforms associated with expected vs. unexpected completions were then separately analysed. A larger PNP for unexpected than for expected targets was observed; interestingly, the more implausible the sentences were, the larger were the positivities. These PNPs would reflect "meta-linguistic evaluation" processes (Molinaro et al., 2012) and/or "elaborate revisions or inferences" serving semantic integration processes (Brothers et al., 2015).

Finally, the analysis of the Time Frequency (TF) domain of the EEG might provide a promising tool for further clarifying how semantic composition and integration develop in sentence comprehension. In fact, power increase in lower and middle gamma frequency bands (approximately between 35-75 Hz) has been associated with successful semantic unification mechanisms (e.g., Hald, Bastiaansen & Hagoort, 2006; Penolazzi, Angrilli & Job, 2009; Wang, Zhu & Bastiaansen, 2012). Lewis and Bastiaansen (2015) proposed that such activity "reflects a match between strong top-down predictions and bottom-up linguistic input" (p.159). As we mentioned, Rommers et al (2013) directly contrasted the EEG activity in idiomatic and literal sentence processing in the time-frequency domain. They interpreted the power increase in the middle gamma frequency band (50-70Hz) observed after the presentation of the expected target words in literal but not in idiomatic contexts as evidence that semantic unification was not fully carried out during idiom comprehension. Induced changes in the frequency domain of the EEG may be a sensible measure of literal composition since they are not strictly time-locked to the cyclic computation of each single word and may reflect the processing mechanisms that go beyond the single constituent level (e.g., Lewis & Bastiansen, 2015).

We thus expected that literal composition to be associated with changes in the middle gamma frequency band, and in the N400 amplidude. Enriched integration processes may instead be associated with larger positive effects in the ERPs, reflecting a controlled reanalysis of the discurse which could be sustained in time when a substantial revision of the sentence representation is needed (e.g., for irony comprehension see Regel et al., 2011), or short-lived when a minimal revision of the sentence is required (e.g., Molinaro et al., 2012).

2. The Present Study

The general aim of this study was to assess to what extent the literal composition of the words forming an idiom string is indeed performed during idiom processing (namely, when a basic semantic composition is needed) and how readers integrate the idiomatic meaning in the sentence representation once literal composition turns out to be insufficient or defective (namely when an enriched meaning integration is needed). We thus designed an EEG experiment in which short idiomatic strings (e.g., *break the ice*) with both a plausible literal meaning and an idiomatic meaning were embedded in sentences strongly biasing readers towards the literal or the idiomatic meaning. A control condition was also designed formed by literal sentences that only used idiom-final words (e.g., *ice*) with cloze probability values as high as in literal and idiomatic sentences.

Ambiguous idioms provide an ideal test bed for contrasting the literal and the idiomatic processing of the very same sequence of words. In fact, the use of literally-plausible, rather than literally-implausible idioms as in Rommers et al. (2013), may allow us to assess processing differences solely due to idiomaticity and limit the effect of differences in frequency of co-occurrence of words, or familiarity. In fact, as Tabossi, Fanari and Wolf (2009) noted, idioms are familiar to most speakers, whereas matched literal sentences are fairly novel. This represents a potential confound in many idiom studies.

Participants read word by word each of the sentences forming the three conditions (Literal, Idiomatic and Control) while their EEG was recorded. In the Literal and Idiomatic conditions, ERPs were time-locked to the first word of the idiom. In the Control condition, ERPs were time-locked to the idiom-final word to better characterise potential processing differences at the end of the idiom string, where the integration of the idiom meaning in the sentences representation was expected to occur. We minimized the impact of differential sentence constraints, known to elicit N400 effects, and compared the ERPs to semantically well-formed sentences in which expectations towards upcoming words were similarly high and were actually fulfilled (see examples in Table 1). In literal contexts, readers should perform a word-by-word semantic composition of the literal meaning of the idiomatic constituents, namely a basic semantic composition. In contrast, consistent with the Configuration Hypothesis, in idiomatic contexts the literal meanings of the idiomatic constituents should be accessed and semantic composition performed until the idiom recognition point (e.g., Titone & Connine, 1994; Boulenger et al, 2012). The idiom meaning is then retrieved from semantic memory and integrated in the sentential context through an enriched meaning integration where additional cognitive efforts are expected to select the idiomatic meaning and adjust it to the global sentence meaning to obtain a coherent discourse interpretation. The semantic composition of the literal meaning of the words following idiom recognition is often unnecessary (Rommers et al., 2013) if not disruptive, since it would lead to a contextually inappropriate literal interpretation (for related behavioural evidence, Cacciari & Corradini, 2015; Holsinger, 2013).

In sum, we expect that after idiom recognition an effect would be visible on the N400 component (smaller N400 in the Idiomatic than in the Literal condition), since basic semantic composition may decay if not being switched off. Consistent with Rommers et al (2013), this may also lead to changes in the middle gamma frequency band, with a power decrease in the Idiomatic compared to the Literal condition. If idioms behave as metaphors or irony (De Grauwe et al., 2010; Regel et al., 2011), we might find LPC or P600 effects in Parietal electrodes that would reflect the cognitive demands associated with the integration of the figurative meaning. However, because idiom processing should require minimal contextual re-analysis, we might alternatively find that idiomatic meaning integration is associated with Frontal PNP effects, with more positive ERPs in Idiomatic than in Literal and Control conditions, occurring in a time-window following the N400 (but note that PNP effects sometimes show a "substantial temporal overlap" with the N400, see Thornhill & Van Petten, 2012).

-- Table 1 ---

3. Method

3.1 Participants

Forty-two students at the University of Modena-Reggio Emilia participated in the experiment (22 female, mean age 19, range = 18-20) for course credit or a gift reward. Participants were right-handed native speakers of Italian and reported normal or corrected-to normal vision. No participant had a history of neurological disorder. The research was carried out fulfilling ethical requirements in accordance with standard procedures at the University of Modena-Reggio Emilia.

3.2 Materials and design

416 students (273 female) from the University of Modena-Reggio Emilia, not aware of the aim of the experiment, took part in the norming of the experimental materials. Two hundred idiom strings with an idiomatic and a well-formed literal meaning were selected from a collection of Italian idioms and were divided into two lists. Each list was submitted to 20 participants who were asked to rate each idiom for familiarity (from 1: "never heard" to 7: "heard very often") and to paraphrase it. The 90 idioms selected as experimental materials were highly familiar (M=5.4, SD = 0.81, range = 3.78-7) and correctly paraphrased (mean of correct paraphrase = 87%, SD = 13%, range = 61%-100%). 88 idiom strings were formed by Verb, Determiner and

Noun, and two by Verb and Noun. The strings were embedded in sentences that biased towards the figurative meaning in the Idiomatic Condition (henceforth IC), or towards the literal meaning in the Literal Condition (henceforth LC). In the Control Condition (henceforth CC), only the idiom-final word was used embedded in a literal sentence (see Table 1 for examples). The idiom strings were exactly the same in LC and IC, although they were preceded by different contexts. Sentence length did not differ across conditions [IC = 15.74; LC = 15.96; CC = 15.87; $F(2,178) = 1.55$, ns]. To avoid wrap-up effects, on average 3.9 constituents followed the idiom string. The sentence structure was kept as similar as possible across conditions. To test the cloze-probability of the string constituents in the three conditions, different questionnaires containing sentence fragments of increasing length were used. 324 participants were asked to complete the sentence fragments with the first word that came to their mind. Literal filler sentences were intermixed such that the Idiomatic context condition represented only one third of the fragments in each questionnaire. Idiom-final words (henceforth, W3) had a mean cloze probability value of 0.87 (SD=0.11) in IC, of 0.91 (SD=0.09) in LC and 0.89 (SD=0.11) in CC. The first idiom constituent (henceforth, W1) had a cloze probability value of 0.07 (SD=0.17) in IC and 0.11 (SD=0.19) in LC, cloze values greatly increased after the verb (henceforth, W2) [0.60 (SD=0.30) in IC and 0.60 (SD=0.35) in LC]. The cloze probability of W3 did not significantly differ across conditions [$F(2,178) = 2.67$, $p=0.07$]. The Recognition Point was operationalized as the word after which cloze probability increased above 0.65. This led to 47 idioms recognized after the presentation of W1, and 43 after W2.

As Popiel and McRae (1988) pointed out, idiomatic expressions with a well-formed literal interpretation are used more frequently, and perceived as more familiar, when intended figuratively rather than literally. To be sure that readers would not interpret the literal sentences containing the idiom strings figuratively, a different group of 16 students judged whether the sentences forming the Literal conditions had a literal or a figurative meaning using a two-choice task. Participants assigned a literal interpretation to 95% of the literal sentences. The 90 experimental sentences (30 sentences in each of the three conditions) were also rated for naturalness by 36 different participants. Three different lists were constructed with an equal number of sentences of the three experimental conditions. Below each sentence, a scale was reported ranging from 1: "The sentence is totally unnatural" to 7: "The sentence is perfectly natural". The sentences of the Control condition (M=6.1, SD=0.41) were rated as slightly but significantly more natural that those of the Idiomatic [M = 5.71, SD = 0.64, t (89) = 4.83, p < 0.001] and Literal conditions [M = 5.63, SD = 0.72, t (89) = 5.39, p < 0.001] that instead were perceived as equally natural (t < 1). The 90 experimental sentences were divided in three lists according to a Latin square design. Each list contained the same number of Idiomatic, Literal and Control sentences counterbalanced across lists, so that participants were presented with the same idiomatic string or idiom-final constituent only once. Participants were randomly assigned to one of the three lists. Critical sentences were intermixed with 120 literal filler sentences of similar length and structure in a pseudo-randomized order.

3.3 Procedure

Participants were tested individually in a sound-attenuated room. Sentences were visually displayed word by word at the center of a computer screen. Participants were instructed to read the sentences for comprehension and to avoid eye movements and other muscular activity during the sentence presentation. Trials began when participants pressed a keyboard button. A fixation point (a cross) at the center of the screen was substituted with single words presented for 300 ms and separated by a 300 ms blank (SOA = 600 ms). The presentation of each sentence was followed by a 1500 ms blank. Participants pressed the space bar to read the next sentence so that they had additional time to rest between trials. Every 10 sentences on average, participants answered a true–false question about the content of the sentence (always a filler sentence) just read. After each response, feedback was given. The experiment started with a practice session of 5 literal sentences.

3.4 Data acquisition and analysis

The electroencephalogram (EEG) was amplified and recorded with a BioSemi Active-Two System from 32 active electrodes placed on the scalp (Fp1, Fp2, AF3, AF4, F3, F4, F7, F8, FC1, FC2, FC5, FC6, C3, C4, T7, T8, CP1, CP2, CP5, CP6, P3, P4, P7, P8, O1, O2, PO3, PO4, Fz, Cz, Pz, Oz) plus four electrodes placed around the eyes for eye movement monitoring (2 at the external ocular canthi and 2 below the eyes) and two electrodes placed over the left and right mastoids. EEG and EOG signals were amplified and digitized continuously with a sampling rate of 512 Hz. Two additional electrodes were placed close to Cz, the Common Mode Sense [CMS] and the Driven Right Leg [DRL] electrodes (for a detailed discussion about the Biosemi system, see Van Rijn, Peper & Grimbergen, 1990). EEG signals were online referenced to CMS and re-referenced off-line to the average activity of the two mastoids. The EEG signal was analyzed using EEGLAB (Delorme & Makeig, 2004), and FieldTrip (Oostenveld, Fries, Maris & Schoffelen, 2010) open-source toolboxes for MATLAB (The MathWorks, Inc).

Data were filtered with a high-pass Butterworth filter (0.05 Hz) and long epochs (of 3200 ms) containing the three critical target words in idiomatic and literal contexts (from -1600 ms to 1600 ms from the presentation of the last MWE constituent) were extracted. In this epoch an Independent Component Analysis (ICA) was performed in order to identify and remove ocular artefacts only (e.g., Jung, Makeig, Humphries, Lee, McKeown, Iragui & Sejnowski, 2000). After ocular artefacts removal, two criteria were used to reject EEG epochs containing other types of artefacts: an amplitude criterion of +/- 90 μV and the visual inspection of the remaining epochs. 14 participants were excluded from the analyses because of the high number of rejected epochs (> 40%)[1]. For the 28 participants included in the analysis an average of 25 epochs per condition survived artifact rejection, with no differences between conditions [F(2,54) <1].

3.4.1 Event Related Potentials

For computing ERPs, an additional low pass filter (35 Hz) was applied. The epochs were baseline-corrected to a pre-stimulus interval, from -1500 to -1200 ms preceding the presentation the last word of the idiom, for all three conditions (i.e., the baseline period occurred before the presentation of the first idiomatic constituent). Statistical analyses comparing IC and LC were performed in the epoch corresponding to the presentation of the whole expression (from -1200 to 600 ms). On the last word of the expression (0 to 600 ms), all three conditions were compared. Repeated measures ANOVAs were performed on the single-subject mean voltage activity in fixed time-windows, obtained from each electrode site and aggregated across different levels of the topographic factors of interest. Analyses were performed in R (R Core

1 The relatively high number of participants excluded from the analyses was due to the massive presence of slow drifts in the recordings of some participants, possibly due to sweating (the experiment was carried out in July). After having realized, during pre-processing, that the number of participants to reject was high, we collected data from 12 additional participants in October, to gain the appropriate signal to noise ratio. The ERP analyses were therefore carried out on 28 participants.

Team, 2015). Separate ANOVAs were performed in different time-windows to evaluate for ERP effects affecting the N400 component (300-500 ms time-window from the presentation of each of the three critical words) and the effects on the following positivity (PNP). The selection of the appropriate time-window for the positive effect is not straightforward (Van Petten & Luka, 2012): the effects are observed more often in the 600 to 900 ms time-window, but may also have a "substantial temporal overlap" with the N400, with a distribution focused on Frontal electrodes (e.g., Thornhill & Van Petten, 2012). We thus selected an early (400-600ms) and a late (600-900ms) window. Two types of ANOVAs on different sets of electrodes were conducted. Context (IC, LC, CC) and topographic factors were treated as within-subject factors. 24 lateralized electrodes were included in the lateralized ANOVAs in which the Longitude factor was organized in two levels: Anterior (A: FP1,FP2,AF3,AF4, F7, F8, F3, F4, FC5,FC6,FC1, FC2) vs. Posterior (P: CP5, CP6, CP1, CP2, P7, P8, P3, P4, PO3, PO4, O1, O2). The Hemisphere factor was also organized in two levels (Left vs. Right). In the midline ANOVAs, the Longitude factor had two levels with two electrodes for each level (Fz and Cz in Anterior, and Pz and Oz for Posterior). ANOVAs were Greenhouse-Geisser corrected when appropriate (uncorrected degrees of freedom and corrected p values are reported). Effect size was estimated with partial eta squared (Lakens, 2013).

3.4.2 Time-frequency analysis

TF representations (TFR) of single-trial data were computed using the multi-taper approach described by Mitra and Pesaran (1999). Power changes in TFRs were computed with 400 ms time smoothing and 5 Hz frequency smoothing in 2.5 Hz frequency steps (from 35 to 80Hz) and 10 ms time steps. Power estimates for each trial were then averaged across IC and LC conditions for each participant. Single subject average TFRs are expressed as relative change from a pre-stimulus interval ranging from 400 to 150 ms prior to the presentation of the first word of the expression (Cohen, 2014).

Differences between the Idiomatic and Literal conditions were evaluated using a cluster-based random permutation test (Maris & Oostenveld, 2007). The algorithm identifies clusters of data that differ across conditions and are contiguous in time, frequency and scalp distribution, and controls for multiple comparisons. We looked for power differences in the gamma frequency band (35-80Hz) between the Idiomatic and Literal conditions across all electrodes. We computed separate contrasts for each time-window associated with the presentation of the three words forming the expression.

4. Results

4.1 ERP analysis

The visual inspection of the brainwaves elicited by the three conditions did not show any clear N400 differences (Figure 1). The most visible difference in the ERPs insists on a slightly later time-window in Anterior electrodes rather than Posterior electrodes.

--- Figure 1 ---

--- Table 2 ---

--- Table 3 ---

4.1.1 N400 differences across word positions

For lateralized electrodes, the three ANOVAs (Table 2) performed on the

N400 time-windows (300 to 500 ms after word presentation) yield to significant interactions between Condition and Longitude in W1 [$F(1,27)=11.29$, $p<0.01$], W2 [$F(1,27)=5.27$, $p<0.05$] and W3 [$F(1,27)=6.43$, $p<0.05$]. Considering midline electrodes, a significant interaction between Condition and Longitude emerged in W1 [$F(1,27)=8.24$, $p<0.01$]. To confirm the anterior distribution of the effect, different one-way ANOVAs were performed on each level of Longitude (Table 3) showing that in Anterior electrodes the waveforms were more positive in IC than in LC in W1: differences are reliable when considering all lateral electrodes [+0.78μV, $F(1,27)=5.90$, $p<0.05$] and show a marginally significant trend when considering midline electrodes only [+0.73μV, $F(1,27)=3.51$, $p<0.1$]. No significant differences emerged in Parietal electrodes. The significant interactions between Condition and Longitude in W2 and W3 were explained by a qualitatively similar, but weakened pattern. During W2, effects are not reliable in Frontal nor Parietal electrodes. In W3 the difference between IC and LC was only marginally significant over lateralized [+0.65μV, $F(1,27)=3.40$, $p<0.1$] and not over midline [+0.70μV, $F(1,27)=2.47$, ns] electrodes. The frontal distribution of the difference between IC and LC (see scalp distribution in Figure 1) in W1 is not compatible with the typical N400 topography (e.g., Kutas, Van Petten & Kluender, 2006).

4.1.2 Early and Late PNP on the last word of the expression

--- Figure 2 ---

--- Figure 3 ---

--- Table 4 ---

--- Table 5 ---

We analyzed ERP differences associated with the three different conditions upon the presentation of the idiom-final words (W3). As it can be seen in Figure 2, the differences between IC and the other two conditions are more consistent in Anterior electrodes in the early PNP time-window (sse Figure 2 and 3). Differences in the ERPs overlap with the N400 time-window and seem consistent until 600 ms in Anterior electrodes. We first carried out two omnibus ANOVAs (in lateral and midline electrodes, separately) during the early PNP time-window (see Table 4). Significant Condition X Longitude interactions emerged in both lateralized [$F(2,54)=5.93$, $p<0.01$] and midline electrodes [$F(2,54)=8.60$, $p<0.001$]. In lateralized electrodes, differences among conditions were robust in Anterior [$F(2,54)=4.39$, $p=0.018$, $\varepsilon=0.97$, $\eta^2_p=0.14$] but not significant in Posterior electrodes [$F(2,54)<1$]. Similarly, also in midline electrodes, the effect of Condition was significant in Anterior [$F(2,54)=6.16$, $p=0.008$, $\varepsilon=0.84$, $\eta^2_p=0.19$] but not in Posterior [$F(2,54)<1$] electrodes. During the late PNP time-window (600-900 ms), ERPs show very little differences between conditions (see Table 5). In fact, ANOVAs revealed only one significant (and spurious) interaction between Condition and Hemisphere, which was not due to a larger effect of Condition in one of the two hemispheres [Left: $F(2,54)<1$; Right: $F(2,54)=1.40$, ns], but to amplitude differences in Left and Right scalp sites in CC [$F(1,27)=8.16$, $p=0.008$], rather than in IC [$F(1,27)<1$] or LC [$F(1,27)<1$].

To better analyze the differences between conditions, we analyzed a subset of Frontal electrodes (FP1FP2,AF3,AF4,F3,Fz,F4,FC1,FC2) on which a Frontal PNP has been previously reported (e.g., Brothers et al., 2015; Thornhill & Van Petten, 2012). A significant effect of Condition was confirmed by a one-way ANOVA [$F(2,54)=6.28$, $p=0.004$, $\varepsilon=0.96$, $\eta^2_p=0.19$]: pair-wise comparisons between the three conditions showed that ICs elicited more positive ERPs than LCs [+1.01μV, CI=(+0.20μV : +1.81μV), $t(27)=2.55$, $p=0.017$] and CCs [+1.37μV, CI=(+0.55μV :

+2.19μV), t(27)=3.44, p=0.002], without any significant difference between LCs and CCs [+0.36μV, CI=(-0.31μV : +1.04 μV), t(27)=1.11, ns].

4.1.3 Early PNP on the first two words of the idiom

A visual inspection of Figure 1 revealed that the difference between IC and LC in the N400 time-window is very similar to the difference between these two conditions at W3. Presumably, this can reflect the onset of a similar frontal PNP elicited by the first idiomatic constituent. Indeed, during W1 presentation a significant interaction between Condition and Longitude emerged in both lateral [F(1,27)=8.57, p=0.006, η^2_p=0.24] and midline [F(1,27)=7.32, p=0.012, η^2_p=0.21] electrodes, with IC showing more positive ERPs than LC in Anterior electrodes [midline: +1.09μV , CI=(+0.27μV : +1.91μV), t(27)=2.72, p=0.011; midline: +0.96μV, CI=(+0.26μV : +1.65μV), t(27)=2.83, p=0.008]. In the same time-window but during the presentation of W2, a significant Condition by Longitude interaction emerged only in lateral electrodes [F(1,27)=5.39, p=0.028, η^2_p=0.17], but differences between conditions were not reliable in Anterior (+0.22μV, t<1) or Posterior [-0.61μV, t(27)=-1.34, ns] electrodes.

4.2 TF Analysis

--- Figure 4 ---

Power differences between IC and LC (Figure 3) emerged with an increase in gamma power for LC with respect to IC (summedT= -4192.52, p=0.0629) on W2. No significant clusters were instead identified on W1 and W3. The increase was more pronounced between 55 and 70 Hz, in a time interval ranging from -600 ms (W2 onset) to -100 ms: IC [-4.22%, CI=(-8.45%; -0.23%)] vs LC [+3.91%, CI=(-0.44%;+8.22%)]. A visual inspection revealed that the cluster represented power changes with similar scalp distribution, direction and frequency interval to those reported in Rommers et al (2013). The power change was larger for a cluster of non-peripheral electrodes (see bottom part of Figure 3). Overall, the TF analysis suggested that the differences between contexts started just before the presentation of W2.

5. Discussion

The aim of this study was twofold: first, to assess whether the literal composition of the words forming an idiom string (*basic semantic composition*) was indeed performed during idiom processing; and second to clarify how readers integrated the idiomatic meaning in the sentential representation (*enriched meaning integration*). Using well-formed sentences with highly expected final words, we found no N400 amplitude difference between literal and idiomatic contexts, at variance with our predictions and with the predictions emerging from the Unification view (Hagoort, 2013). Rather, ERP differences emerged in a slightly later time-interval with a positive effect associated with idiom-final words and with the first idiomatic constituent. We hypothesized that enriched integration mechanisms may involve meaning selection and revision mechanisms (we return to this point below). The electrophysiological signature of these processes was a frontal PNP. This result concurs with brain-imaging evidence (Boulenger et al, 2009; Romero Lauro et al., 2008; Zempleni et al., 2007) in showing that idiom processing is more cognitively demanding compared to literal language, especially when idiom strings also have a literal meaning.

5.1 Frontal PNP and idiom meaning integration

While sentence interpretation in literal contexts is carried out combining individual word meanings with the incremental representation of the sentence, idiom processing requires that the conventional meaning of the expression is retrieved and integrated/adapted to the specific sentential meaning. In our study, the difference between basic composition and enriched integration did not affect the N400 but rather the frontal PNP. In fact, a positive shift in frontal electrodes emerged upon presentation of the first and the last words of the idiom strings when embedded in idiomatic but not in literal contexts. The positivity observed here notably differed from the LPC or P600 effects reported in metaphor or irony studies (e.g., Coulson & Van Petten, 2002; Regel et al., 2010; De Grauwe et al, 2010; Weiland et al., 2014). In fact, in our study the differences between conditions were consistent from the early time-window (400 to 600 ms), without any differences in the subsequent time interval (600 to 900 ms) where LPC and P600 usually occur. The different timing of our effect may be explained by due to the difference between idioms and metaphors and irony. Assuming that a frontal distribution of the positive difference reflects sentence reanalysis rather than repair mechanisms (Kaan & Swaab, 2003; Van Petten & Luka, 2012), the frontal PNP may reflect lesser demands for idioms compared to metaphors or irony, usually associated with longer lasting positive differences over parietal electrodes. In fact, the meaning of a metaphor is built on the fly creating a conceptual (and often elaborate) link between topic and vehicle that may involve deeper reanalysis.

The timing and distribution of our effect are very similar to the frontal PNP reported by Thornhill and Van Petten (2012) during the processing of plausible words fitting previous sentence context but “less than highly predictable”. In our case, there was no disconfirmed lexical prediction since the targets’ cloze probability across contexts was carefully balanced, and all target words actually fulfilled contextual expectations. Our PNP effect(s) are better accounted for in terms of sentence revision mechanisms that, in idiomatic contexts, were needed to select the appropriate interpretation of the ambiguous idiom string and adjust it to the sentence representation. Our effect is similar to the “short lasting LPC” reported by Molinaro et al (2012) who found two frontal positive effects varying in onset latency: when contrastive noun-adjective pairs (*beautiful monster*) were compared to neutral pairs (*lonely monster*), the effect was sustained (from 550 to 750 ms), whereas it was short-lived (from 550 to 650 ms) for (more natural) redundant pairs (*horrible monster*). Our results provide additional evidence that frontal PNPs may be linked to sentence re-analysis mechanisms (Brothers et al., 2015), and can be observed also when lexical predictions are actually fulfilled.

Admittedly, the frontal PNP emerging after the presentation of the first idiomatic word was unexpected. One possible explanation is that participants did not wait until the end of the expression to re-analyse the sentence meaning, and that re-analysis is initiated as soon as after the presentation of W1, i.e., immediately after the recognition of the idiom[2] (cloze values show that 47 out of 90 idiomatic expressions were recognized upon presentation of W1). The enriched integration of the idiomatic meaning in the sentence representation might thus be conceived as a distributed process that is initiated as soon as the idiom is recognized and is terminated when the integration of the figurative meaning is completed.

2 We thank an anonymous reviewer for pointing out the importance of this finding.

5.2 Gamma band frequency and semantic composition

As in Rommers et al (2013), the TF analysis showed that the cognitive mechanisms at play during literal and idiomatic processing differed in the temporal development of the power amplitude of the middle gamma frequency band. These differences were sustained in time and emerged prior to the idiom-final word. We observed a larger power increase in the middle gamma frequency band when words were embedded in literal than in idiomatic contexts. This may reflect the fact that, as Rommers et al. (2013) proposed, word-by-word semantic unification processes were more actively engaged in literal than in idiomatic sentence processing.

The interpretation of changes in gamma frequency power is still tentative, given the yet scarce (but growing) evidence on TF effects in language comprehension. In fact, differences have been previously reported for different frequency ranges: power increase in the lower Gamma (35-45Hz) has been associated with the processing of correct sentences compared to sentences containing semantic violations (Hald et al., 2006; Wang et al., 2012; but see the increase in gamma for world knowledge violations observed in Hagoort et al., 2004); also the power increase in middle gamma range (50-70Hz) has been reported for successful sentence processing (Penolazzi et al., 2009; Monsalve, Perez & Molinaro, 2014). Despite the different frequency ranges of lower and middle gamma bands, and the different time development of the effects observed (sustained in time, Hald et al., 2006; Penolazzi et al., 2009; Wang et al., 2012; or short lasting, Monsalve et al., 2014), according to a recent interpretation of these changes the increase in gamma power would reflect reverberatory activity triggered by a successful match "between the pre-activation of the neural representation of the predicted word, and the neural representation of the actually incoming word" (Wang et al., p.11). This explanation (see also Lewis &

Bastiaansen, 2015) would be consistent with our results in literal contexts where a match occurred between the unfolding input and the highly expected word. However, it would fail to explain what happened in idiomatic contexts where the confirmation of the expected constituent should have led to a similar power increase that instead was not observed.

The results obtained from the analysis of the TF domain support the view that basic semantic composition processes were less engaged in idiomatic than in literal contexts for idioms with a plausible, albeit idiom-unrelated, literal interpretation. This extends what has been observed by Rommers et al. (2013) with literally-implausible idioms (e.g., *keep an eye out*) where the composition of literal meanings would have led to a semantically anomalous interpretation of the string. The observed decrease in gamma power may index the suspension of basic combinatorial processes involved in processing highly predictable strings. In idiomatic contexts, once the idiom is recognised, the processing of the remaining idiom constituents may be shallower if not absent and this led to the power differences on W2. In sum, when, as in literal contexts, a literal word-by-word semantic composition was carried out an increase in upper gamma power occurred, triggered by the match between expectations and incoming input. Differently, in idiomatic contexts no increase occurred when the idiomatic expression was recognized suggesting that a compositional analysis was no longer needed because of the availability of the idiom meaning (Peterson et al., 2001).

The study of frequency changes in language processing is a relatively recent field of study and more research is needed to satisfactorily account for the observed effects (Lewis & Bastiaansen, 2015). Still, our ERP results provide further evidence of positive effects in frontal electrodes that are not necessarily associated with differences in predictability (Van Petten & Luka, 2012), and highlight the need for future work providing a coherent theoretical framework unifying still inconsistent pieces of evidence (see the effects on frontal P600, LPC and PNP components).

References

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82. doi:10.1016/j.jml.2009.09.005

Bastiaansen, M. C. M., Mazaheri, A., & Jensen, O. (2012). Beyond ERPs: Oscillatory neuronal dynamics. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 31–50). New York, NY: Oxford University Press.

Bobrow, S. A., & Bell, S. M. (1973). On catching on to idiomatic expressions. *Memory & Cognition*, *1*(3), 343–346. doi:10.3758/BF03198118

Bohrn, I. C., Altmann, U., & Jacobs, A. M. (2012). Looking at the brains behind figurative language—A quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. *Neuropsychologia*, *50*(11), 2669-2683.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on 'semantic P600' effects in language comprehension. *Brain Research Reviews*, *59*(1), 55–73. doi:10.1016/j.brainresrev.2008.05.003

Bornkessel-Schlesewsky, I., Philipp, M., Alday, P. M., Kretzschmar, F., Grewe, T., Gumpert, M., ... & Schlesewsky, M. (2015). Age-Related Changes in Predictive Capacity Versus Internal Model Adaptability: Electrophysiological Evidence that Individual Differences Outweigh Effects of Age. *Frontiers in Aging Neuroscience*, *7*.

Boulenger, V., Hauk, O., & Pulvermüller, F. (2009). Grasping Ideas with the Motor System: Semantic Somatotopy in Idiom Comprehension. *Cerebral Cortex*, *19*(8), 1905–1914. doi:10.1093/cercor/bhn217

Boulenger, V., Shtyrov, Y., & Pulvermüller, F. (2012). When do you grasp the idea?

MEG evidence for instantaneous idiom understanding. *NeuroImage*, *59*(4), 3502–3513. doi:10.1016/j.neuroimage.2011.11.011

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, *136*, 135-149.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–143. doi:10.1016/j.brainres.2012.01.055

Brouwer, H., & Hoeks, J. C. J. (2013). A time and place for language comprehension: mapping the N400 and the P600 to a minimal cortical network. *Frontiers in Human Neuroscience*, *7*. doi:10.3389/fnhum.2013.00758

Brown, C., & Hagoort, P. (1993). The Processing Nature of the N400: Evidence from Masked Priming. *Journal of Cognitive Neuroscience*, *5*(1), 34–44. doi:10.1162/jocn.1993.5.1.34

Cacciari, C. (2014). Processing multiword idiomatic strings: Many words in one? *The Mental Lexicon*, *9*(2), 267–293. doi:10.1075/ml.9.2.05cac

Cacciari, C., & Corradini, P. (2015). Literal analysis and idiom retrieval in ambiguous idiom processing: A reading-time study. *Journal of Cognitive Psychology*, *27*(7), 797-811.

Cacciari, C., Padovani, R., & Corradini, P. (2007). Exploring the relationship between individuals' speed of processing and their comprehension of spoken idioms. *European Journal of Cognitive Psychology*, *19*(3), 417–445. doi:10.1080/09541440600763705

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, *27*(6), 668–683. doi:10.1016/0749-596X(88)90014-9

Citron, F. M., Cacciari, C., Kucharski, M., Beck, L., Conrad, M., & Jacobs, A. M. (2016). When emotions are expressed figuratively: Psycholinguistic and Affective Norms of 619 Idioms for German (PANIG). *Behavior research methods, 48(1)*, 91-111.

Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. MIT Press.

Coulson, S., & Kutas, M. (2001). Getting it: human event-related brain response to jokes in good and poor comprehenders. *Neuroscience letters*, *316*(2), 71-74.

Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory & Cognition*, *30*(6), 958–968. doi:10.3758/BF03195780

Coulson, S., & Van Petten, C. (2007). A special role for the right hemisphere in metaphor comprehension?: ERP evidence from hemifield presentation. *Brain Research*, *1146*, 128-145.

Cutting, J. C., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory & Cognition*, *25*(1), 57–71. doi:10.3758/BF03197285

Davenport, T., & Coulson, S. (2011). Predictability and novelty in literal language comprehension: An ERP study. *Brain Research*, *1418*, 70–82. doi:10.1016/j.brainres.2011.07.039

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.doi:10.1016/j.neuropsychologia.2010.03.017

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, *61*, 150–162. doi:10.1016/j.neuropsychologia.2014.06.016

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. doi:10.1016/j.jneumeth.2003.10.009

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*(03), 357–374. doi:10.1017/S0140525X00058027

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. doi:10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*, *41*(4), 469–495. doi:10.1006/jmla.1999.2660

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84. doi:10.1016/j.brainres.2006.06.101

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

Filik, R., Leuthold, H., Wallington, K., & Page, J. (2014). Testing theories of irony processing using eye-tracking and ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 811.

Frege, G. (1892). Über Sinn und Bedeutung. Zeitschrift für Philosophie und philosophische Kritik C, 25–50.

Friederici, A. D. (2011). The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews*, *91*(4), 1357–1392. doi:10.1152/physrev.00006.2011

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

Gibbs Jr., R. W., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language*, *28*(5), 576–593. doi:10.1016/0749-596X(89)90014-4

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, *9*(9), 416–423. doi:10.1016/j.tics.2005.07.004

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology*, *4*. doi:10.3389/fpsyg.2013.00416

Hagoort, P., & Berkum, J. van. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 801–811. doi:10.1098/rstb.2007.2089

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of Word Meaning and World Knowledge in Language Comprehension. *Science*, *304*(5669), 438–441. doi:10.1126/science.1095455

Hald, L. A., Bastiaansen, M. C., & Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and language*, *96*(1), 90-105.

Häuser, K. I., Titone, D. A., & Baum, S. R. (in press). The Role of the Ventro-Lateral Prefrontal Cortex in Idiom Comprehension: An rTMS Study. *Neuropsychologia*.

Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, *19*(1), 59–73. doi:10.1016/j.cogbrainres.2003.10.022

Holsinger, E., & Kaiser, E. (2013). Processing (non)compositional expressions: Mistakes and recovery. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 866–878. doi:10.1037/a0030410

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(02), 163–178.

Kaan, E., & Swaab, T. (2003). Repair, Revision, and Complexity in Syntactic Analysis: An Electrophysiological Differentiation. *Journal of Cognitive Neuroscience*, *15*(1), 98–110. doi:10.1162/089892903321107855

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205–225. doi:10.1016/j.jml.2004.10.002

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49. doi:10.1016/j.brainres.2006.12.063

Kutas, M., 1993. In the company of other words: electrophysiological evidence for single- word and sentence context effects. Language and Cognitive Processes 8, 533–572.

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. doi:10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. doi:10.1126/science.7350657

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology*, *4*, 863.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933. doi:10.1038/nrn2532

Lewis, A. G., & Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex*, *68*, 155-168.

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

Liu, Y., Li, P., Shu, H., Zhang, Q., & Chen, L. (2010). Structure and meaning in Chinese: An ERP study of idioms. *Journal of Neurolinguistics*, *23*(6), 615–630. doi:10.1016/j.jneuroling.2010.06.001

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. MIT Press.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. doi:10.1016/j.jneumeth.2007.03.024

Mitra, P. P., & Pesaran, B. (1999). Analysis of Dynamic Brain Imaging Data. *Biophysical Journal*, *76*(2), 691–708. doi:10.1016/S0006-3495(99)77236-X

Monsalve, I. F., Pérez, A., & Molinaro, N. (2014). Item parameters dissociate between expectation formats: a regression analysis of time-frequency decomposed EEG data. *Frontiers in Psychology*, *5*. doi:10.3389/fpsyg.2014.00847

Molinaro, N., Barraza, P., & Carreiras, M. (2013). Long-range neural synchronization supports fast and efficient reading: EEG correlates of processing expected words in sentences. *NeuroImage*, *72*, 120-132.

Molinaro, N., & Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biological Psychology*, *83*(3), 176–190. doi:10.1016/j.biopsycho.2009.12.006

Molinaro, N., Carreiras, M., & Duñabeitia, J. A. (2012). Semantic combinatorial processing of non-anomalous expressions. *NeuroImage*, *59*(4), 3488–3501. doi:10.1016/j.neuroimage.2011.11.009

Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching Languages, Switching Palabras (Words): An Electrophysiological Study of Code Switching. *Brain and Language*, *80*(2), 188–207. doi:10.1006/brln.2001.2588

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2010). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, e156869. doi:10.1155/2011/156869

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806. doi:10.1016/0749-596X(92)90039-Z

Papagno, C., & Cacciari, C. (2010). The role of ambiguity in idiom comprehension: The case of a patient with a reversed concreteness effect. *Journal of* *neurolinguistics*, *23*(6), 631-643.

Penolazzi, B., Angrilli, A., & Job, R. (2009). Gamma EEG activity induced by semantic violation during sentence reading. *Neuroscience Letters*, *465*(1), 74–78. doi:10.1016/j.neulet.2009.08.065

Peterson, R. R., Burgess, C., Dell, G. S., & Eberhard, K. M. (2001). Dissociation between syntactic and semantic processing during idiom comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1223.

Popiel, S. J., & McRae, K. (1988). The figurative and literal senses of idioms, or all idioms are not used equally. *Journal of Psycholinguistic Research*, *17*(6), 475–487. doi:10.1007/BF01067912

Proverbio, A. M., Crotti, N., Zani, A., & Adorni, R. (2009). The role of left and right hemispheres in the comprehension of idiomatic language: an electrical neuroimaging study. *BMC Neuroscience*, *10*, 116. doi:10.1186/1471-2202-10-116

Pylkkänen, L. (2008). Mismatching meanings in brain and behavior. *Language and Linguistics Compass*, *2*(4), 712-738.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. >Vienna: R Foundation for Statistical Computing.

Regel, S., Gunter, T. C., & Friederici, A. D. (2011). Isn’t It Ironic? An Electrophysiological Exploration of Figurative Language Processing. *Journal of Cognitive Neuroscience*, *23*(2), 277–293. doi:10.1162/jocn.2010.21411

Regel, S., Meyer, L., & Gunter, T. C. (2014). Distinguishing neurocognitive processes reflected by P600 effects: Evidence from ERPs and neural oscillations. *PloS one*, *9*(5), e96840.

Rijn, A. C. M. van, Peper, A., & Grimbergen, C. A. (1990). High-quality recording of bioelectric events. *Medical and Biological Engineering and Computing*, *28*(5), 389–397. doi:10.1007/BF02441961

Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlesewsky, M. (2007). To Predict or Not to Predict: Influences of Task and Strategy on the Processing of Semantic Relations. *Journal of Cognitive Neuroscience*, *19*(8), 1259–1274. doi:10.1162/jocn.2007.19.8.1259

Rizzo, S., Sandrini, M., & Papagno, C. (2007). The dorsolateral prefrontal cortex in idiom interpretation: An rTMS study. *Brain research bulletin*, *71*(5), 523-528.

Romero Lauro, L. J. R., Tettamanti, M., Cappa, S. F., & Papagno, C. (2008). Idiom Comprehension: A Prefrontal Task? *Cerebral Cortex*, *18*(1), 162–170. doi:10.1093/cercor/bhm042

Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent Semantic Processing in the Human Brain: Evidence from Idiom Comprehension. *Journal of Cognitive Neuroscience*, *25*(5), 762–776. doi:10.1162/jocn_a_00337

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, *27*(2), 251–272. doi:10.1177/0267658310382068

Spotorno, N., Cheylus, A., Van Der Henst, J. B., & Noveck, I. A. (2013). What's behind a P600? Integration operations during irony processing. *PLoS One*, *8*(6), e66839.

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

Strandburg, R. J., Marsh, J. T., Brown, W. S., Asarnow, R. F., Guthrie, D., Harper, R., ... & Nuechterlein, K. H. (1997). Event-related potential correlates of linguistic information processing in schizophrenics. *Biological Psychiatry*, *42*(7), 596-608.

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

Tabossi, P., Fanari, R., & Wolf, K. (2005). Spoken idiom recognition: Meaning retrieval and word expectancy. *Journal of Psycholinguistic Research*, *34*(5), 465-495.

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

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*(3), 382–392. doi:10.1016/j.ijpsycho.2011.12.007

Titone, D. A., & Connine, C. M. (1994). Comprehension of idiomatic expressions: Effects of predictability and literality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1126–1138. doi:10.1037/0278-7393.20.5.1126

Tremblay, A. & Baayen, R. H. (2010) Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood, *Perspectives on formulaic language: Acquisition and communication* (pp. 151-173). London:

The Continuum International Publishing Group.

Van den Brink, D., Brown, C. M., & Hagoort, P. (2006). The cascaded nature of lexical selection and integration in auditory sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 364–372. doi:10.1037/0278-7393.32.3.364

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. doi:10.1016/j.ijpsycho.2011.09.015

Van Rijn, A. M., Peper, A., & Grimbergen, C. A. (1990). High-quality recording of bioelectric events. *Medical and Biological Engineering and Computing*, *28*(5), 389-397.

Varela, F., Lachaux, J.-P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, *2*(4), 229–239. doi:10.1038/35067550

Vega-Moreno, R. E. (2001). Representing and processing idioms. *UCL Wording Papers in Linguistics*, *13*, 73-107.

Verleger, R. (1988).Event-related potentials and cognition: A critique of the context-updating hypothesis and an alternative interpretation of the P300.Behavioral and Brain Sciences,11, 343–427.

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

Wang, L., Zhu, Z., & Bastiaansen, M. (2012). Integration or Predictability? A Further Specification of the Functional Role of Gamma Oscillations in Language Comprehension. *Frontiers in Psychology*, *3*. doi:10.3389/fpsyg.2012.00187

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

Zempleni, M.-Z., Haverkort, M., Renken, R., & A. Stowe, L. (2007). Evidence for bilateral involvement in idiom comprehension: An fMRI study. *NeuroImage*, *34*(3), 1280–1291. doi:10.1016/j.neuroimage.2006.09.049

Zhou, S., Zhou, W., & Chen, X. (2004). Spatiotemporal Analysis of ERP During Chinese Idiom Comprehension. *Brain Topography*, *17*(1), 27–37. doi:10.1023/B:BRAT.0000047334.48256.9f

Figure Captions

Figure 1. *ERP correlates associated with the presentation of the whole expression*. Above. ERPs (further 25Hz low pass filtered) to the onset of the three words expressions (W1, W2, W3), in IC (black line) and LC (red line) from eleven representative electrodes (Negative voltage is plotted upwards). Below. Scalp maps of the mean average difference between IC and LC in the N400 typical time interval (300 to 500 ms), across the three word positions.

Figure 2. *ERPs associated with the last constituent of the expression*.
Above. ERPs (further 25Hz low pass filtered) from nine representative electrodes (Negative voltage is plotted upwards). IC (black line), LC (red line) and CC (green line) are compared at the onset of the last word of the expression (W3). Vertical dashed lines mark the two-windows of interest (from 400 to 600 ms and from 600 to 900 ms).

Figure 3. *Distribution of ERP differences*. Scalp maps of the mean average differences between IC and LC on the left and to IC and CC on the right are displayed for both early PNP (first row) and late PNP (second row) time-windows.

Figure 4. *Time Frequency Analysis*. The first three rows depict the TF representation of EEG data from the presentation of W1 to W3 from electrode FC1: relative changes in Power with respect to the baseline period are plotted as function of time (x axis) and frequency bin (y axis), for IC, LC and IC minus LC. Dark blue corresponds to -20% power change; dark red corresponds to +20% power change. The fourth row

depicts the temporal development of gamma power differences (in the relevant upper gamma frequency range from 60 to 75 Hz) between IC and LC associated with with standard error bars (Morey, 2008). The topographic distribution of the significant negative cluster is displayed on the left of the fourth row. The fifth row shows the results of paired t tests (IC vs LC) carried out on a series of consecutive 10 ms time-window. Showing that differences between conditions onset upon presentation of W2 and are consistent for a time interval of roughly 300 ms.

Table 1. Examples of the experimental materials organized in the three experimental conditions[a].

Idiomatic Context Condition	*1a) La maestra aveva notato che Nicola disturbava i compagni, ma la prima volta chiuse un occhio e continuò la lezione.* (The teacher saw Nick was bothering his desk mate but for the first time she closed an eye and kept on teaching.) *2a) Gli appassionati di montagna talvolta nelle escursioni camminano per giorni e spesso devono stringere i denti fino al traguardo.* (Those who love mountains hike for several days and often have to grit their teeth until they reach the destination.) *3a) Il ragazzo era sempre così arrogante con gli altri che bisognava proprio fargli abbassare la cresta e renderlo più educato.* (The youngster was always so arrogant with everyone that it was necessary to make his crest come off to make him more polite.) *4a) La ragazza pensava già che il colloquio fosse andato male e non la smetteva di fasciarsi la testa mentre aspettava l'esito.* (She was worried because she felt the interview was gone bad and she could not keep from bandage her head up while waiting for the response.) *5a) Per riuscire a preparare l'esame e nel frattempo portare avanti il lavoro Aldo si è spaccato la schiena ma ora è soddisfatto.* (To be prepared for the exam, and keep his job meanwhile, Aldo broke his back but now is satisfied.)
Literal Context Condition	*1b) Alla visita oculistica Enrico, prima di leggere le lettere indicate sulla lavagna luminosa, chiuse un occhio per valutare la miopia.* (At the Ophthalmological visit, before starting to read the letters on the panel aloud Henry closed an eye in order to evaluate his nearsightedness.) *2b) Il dentista sistemò la nuova dentiera in bocca al paziente e poi gli chiese di stringere i denti per completare l'intervento.* (The dentist fixed the new dentures in the patient's mouth and then asked her to grit her teeth to complete the operation.) *3b) Il punk non riusciva a mettersi il casco e perciò andò dal parrucchiere per abbassare la cresta di qualche centimetro.* (The punkster could not wear his helmet and had to go to the hairdresser to cut the Mohawk down a few inches.) *4b) Mario aveva sbattuto contro la roccia senza il casco, e cercava una garza per fasciarsi la testa perché aveva un taglio.* (Mario beat his head on the rock with no helmet and was looking for a gauze to bandage his head up for the wound.) *5b) Il masso franato dalla montagna ha colpito sul dorso il povero mulo, che si è spaccato la schiena fra lo spavento di tutti.* (The rock felt from the mountain hit the donkey on the spine, which broke the back while everyone was scared.)
Control Condition	*1c) Giovanni ha rotto gli occhiali durante la rissa perché ha preso un pugno in un occhio e gli sono caduti a terra.* (Jack broke his glasses during the fight because got a punch in his eye and fell on the ground.) *2c) Matteo a pranzo ha mangiato cosi tanto torrone alle mandorle che alla sera aveva male ai denti e dovette andare dal dentista.* (At lunch, Matteo had so many almond and caramel candy that a few hours later had pain on the teeth, and went to the dentist.) *3c) Il nonno insegnava che il gallo si distingue dalle galline anche perché in testa porta una grande cresta di un rosso acceso.* (Granpa used to say that the rooster differs from the hen also because it has a large vivid red crest on the top of the head.) *4c) Per quanto riguarda i treni eurostar, la prima classe spesso è in coda, altre volte è in testa o in mezzo al treno.*

	(In Eurostar trains, the first class is often on the back of it, sometimes it is on the <u>head</u> of it, sometimes in the middle of it.) *5c) Mia nonna che è molto anziana e ha le ossa fragili porta un busto per tenere dritta la schiena che tende a curvarsi*. (My very old Granma has fragile bones and uses a bust to keep her <u>back</u> straight as it tends to bend.)

[a] The figurative meaning of the expression *"chiudere un occhio"* roughly corresponds to *"turn a blind eye"*; *"stringere i denti"* corresponds to *"gritting teeth"*; *"abbassare la cresta"* corresponds to *"fly down", "fasciarsi la testa"* corresponds to *"crossing the bridge when coming to the river", "rompersi la schiena"* corresponds to *"staying on the grind"*.

Table 2. Analysis of Variance of the N400 time window (from 300 to 500 ms) across word positions.

	W1 (*break*)			W2 (*the*)			W3 (*ice*)		
Lateral electrodes									
	F(1,27)	p	η^2_p	F(1,27)	p	η^2_p	F(1,27)	p	η^2_p
Condition	0.26	0.61	0.01	0.32	0.57	0.01	0.77	0.39	0.03
Condition X Longitude	11.29**	0.002	0.29	5.62*	0.025	0.17	6.43*	0.017	0.19
Condition X Hemisphere	0.88	0.35	0.03	0.37	0.55	0.01	1.95	0.17	0.07
Condition X Longitude X Hemisphere	0.53	0.47	0.02	0.41	0.53	0.01	0.66	0.42	0.02
Midline electrodes									
	F(1,27)	p	η^2_p	F(1,27)	p	η^2_p	F(1,27)	p	η^2_p
Condition	0.13	0.72	0.00	0.90	0.35	0.03	1.03	0.32	0.04
Condition X Longitude	8.24**	0.007	0.23	0.73	0.40	0.03	3.04 .	0.09	0.10

Table 3. Planned comparison between the average amplitude of the N400 in Idiomatic and Literal conditions across the three word positions

	W1 (*break*)		W2 (*the*)		W3 (*ice*)	
Lateral electrodes						
	Voltage difference	F value (1,27)	Voltage difference	F value (1,27)	Voltage difference	F value (1,27)
IC vs LC (Anterior):	+0.78μV	5.90*	+0.61μV	2.46	+0.65μV	3.40 .
IC vs LC (Posterior):	-0.46μV	1.43	-0.16μV	0.13	-0.05μV	0.02
Midline electrodes						
IC vs LC (Anterior):	+0.73μV	3.51 .	+0.55μV	1.52	+0.70μV	2.47
IC vs LC (Posterior):	-0.45μV	0.94	+0.25μV	0.28	+0.06μV	0.02

Table 4. Analysis of Variance of at the last word of the expression considering all three conditions during the early PNP time-window (400-600ms).

	DF	F	p	η^2_p	ε
Lateral electrodes					
Condition	(2,54)	1.21	0.31	0.04	ε=0.94
Condition X Longitude	(2,54)	5.93**	0.006	0.18	ε=0.91
Condition X Hemisphere	(2,54)	0.63	0.53	0.02	ε=0.69
Condition X Longitude X Hemisphere	(2,54)	0.20	0.82	0.01	ε=0.91
Midline electrodes					
Condition	(2,54)	2.11	0.13	0.07	ε=0.94
Condition X Longitude	(2,54)	8.60***	<0.001	0.24	ε=0.96

Table 5. Analysis of Variance of at the last word of the expression considering all three conditions during the late PNP time-window (600-900ms).

	DF	F	p	η^2_p	ε
Lateral electrodes					
Condition	(2,54)	0.68	0.51	0.02	ε=0.99
Condition X Longitude	(2,54)	1.51	0.23	0.05	ε=0.99
Condition X Hemisphere	(2,54)	4.04*	0.033	0.13	ε=0.77
Condition X Longitude X Hemisphere	(2,54)	0.23	0.78	0.01	ε=0.96
Midline electrodes					
Condition	(2,54)	0.19	0.13	0.01	ε=0.99
Condition X Longitude	(2,54)	3.19 .	0.056	0.10	ε=0.88

Figure1 Click here to download Figure Figure_1.jpg

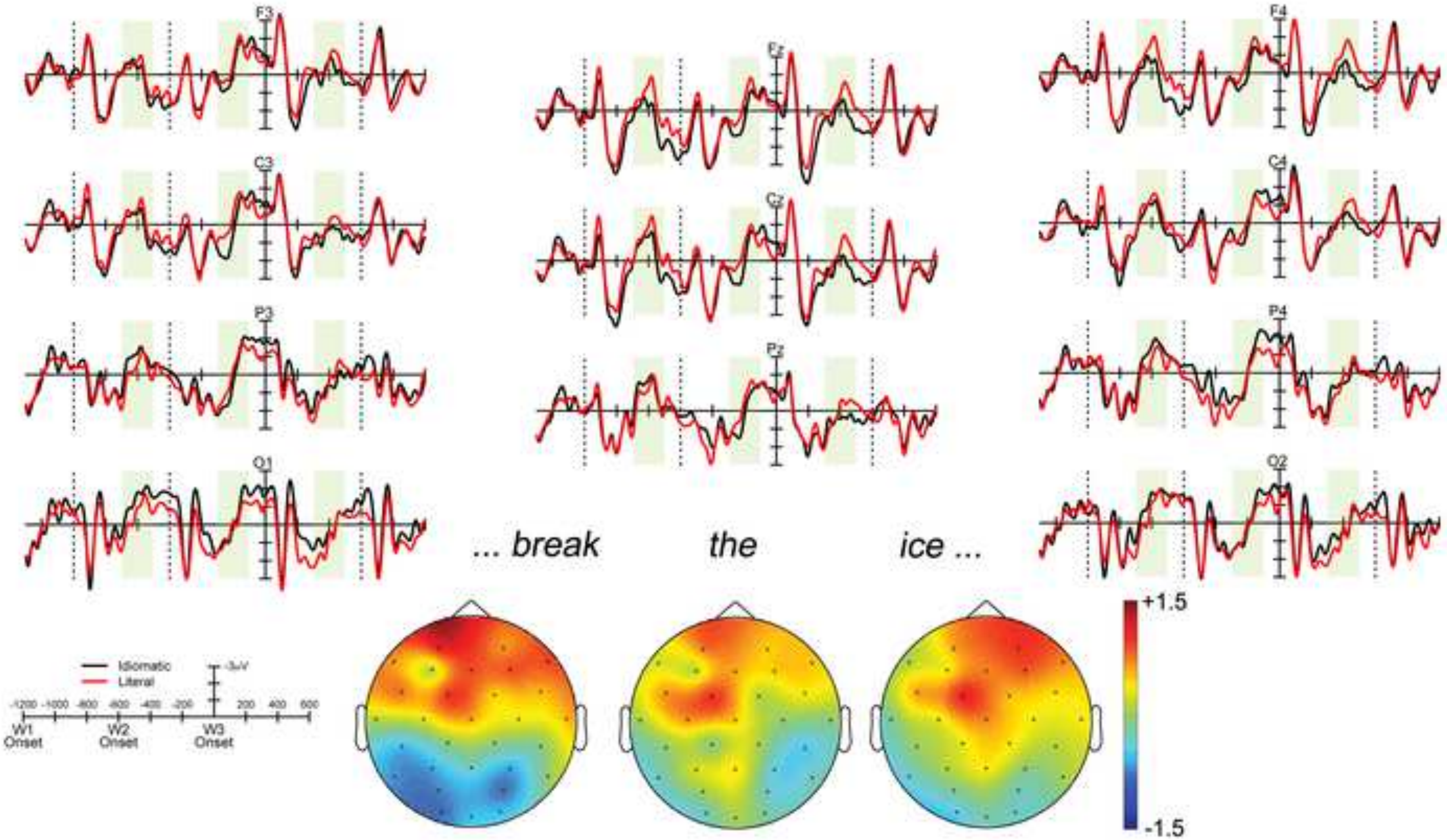
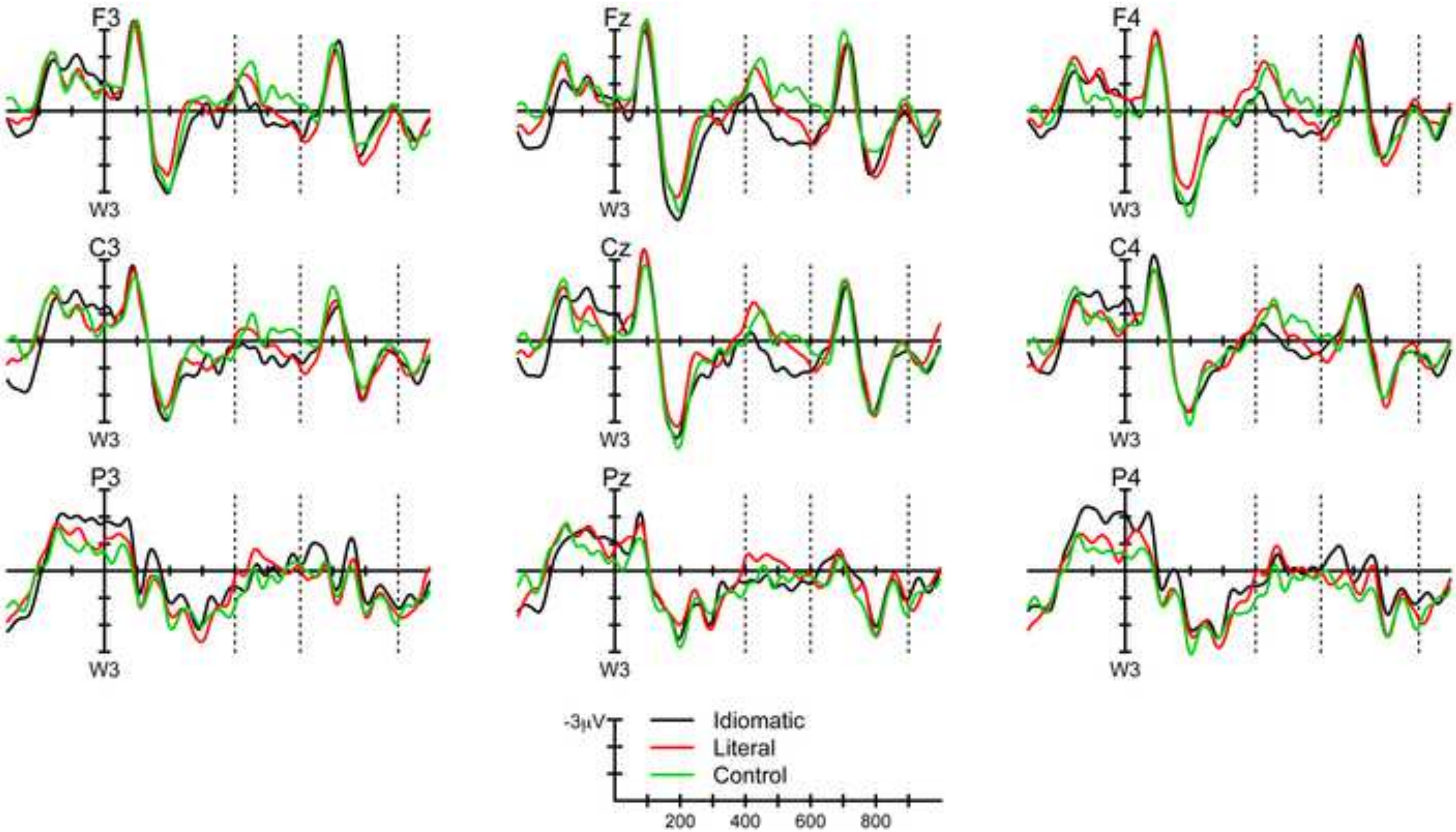


Figure2


F3
Fz
F4
C3
Cz
C4
P3
Pz
P4
W3
-3μV
Idiomatic
Literal
Control
200
400
600
800

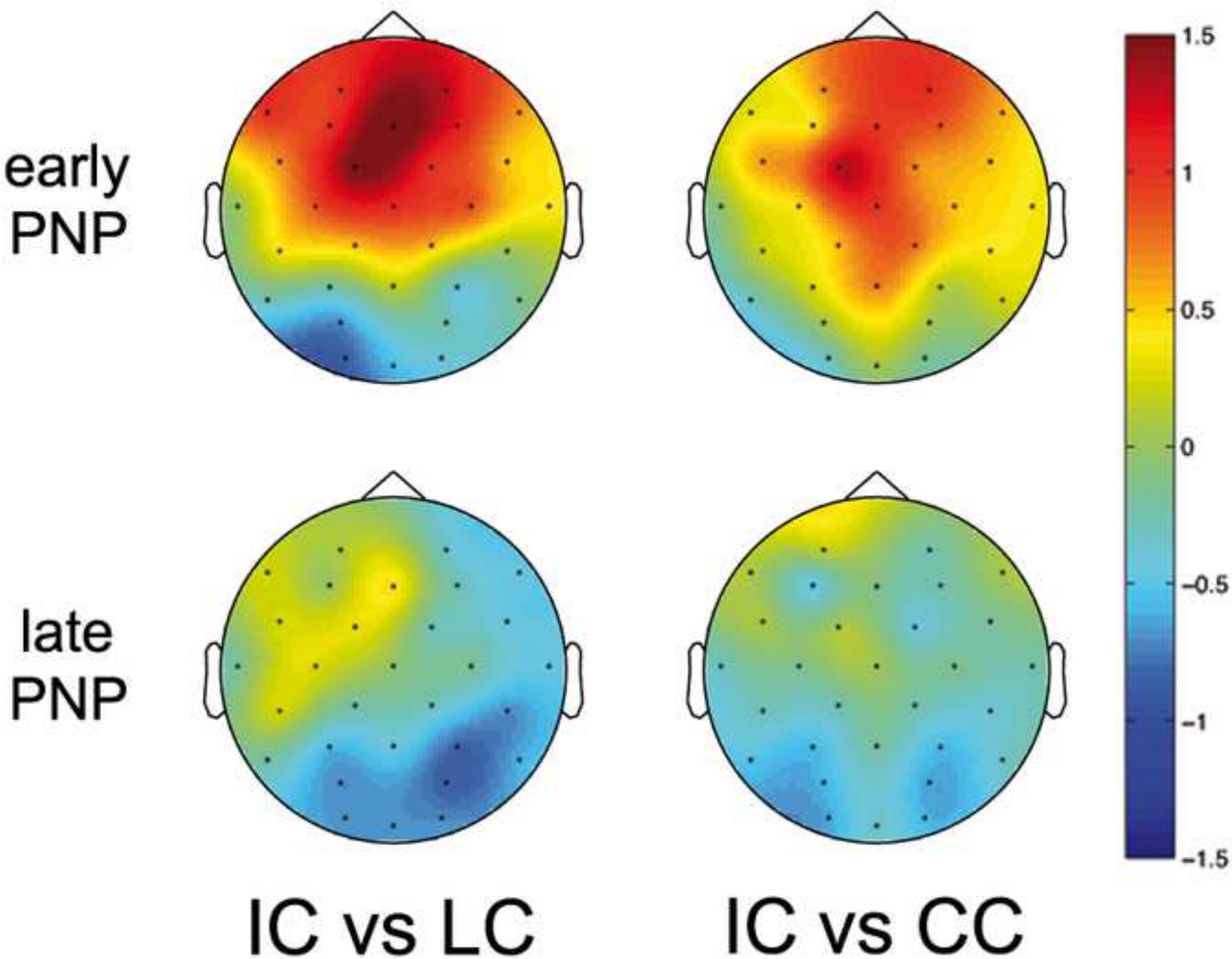

early
PNP
late
PNP
IC vs LC
IC vs CC
1.5
1
0.5
0
−0.5
−1
−1.5

Figure4

Click here to download Figure Figure_4.jpg