

Representation and processing of multi-word expressions in the brain

Anna Siyanova-Chanturia¹, Kathy Conklin², Sendy Caffarra³, Edith Kaan⁴,
and Walter J. B. van Heuven⁵

¹ School of Linguistics and Applied Language Studies, Victoria University of Wellington; ² School of English, University of Nottingham; ³ Basque Centre on Cognition, Brain and Language; ⁴ Department of Linguistics, University of Florida; ⁵ School of Psychology, University of Nottingham

Corresponding author:

Anna Siyanova-Chanturia

School of Linguistics and Applied Language Studies

Victoria University of Wellington

Kelburn Parade

Wellington

6012

New Zealand

+64 (0)4 463 5922

anna.siyanova@vuw.ac.nz

ABSTRACT

Language comprehension is sensitive to the predictability of the upcoming information. Prediction allows for smooth, expedient and successful communication. While general discourse-based constraints have been investigated in detail, more specific phrase-level prediction has received little attention. We address this gap by exploring the ERPs elicited during the comprehension of English binomials – familiar and predictable multi-word expressions. In Experiment 1a, participants read binomial expressions (*knife and fork*), infrequent strongly associated phrases (*spoon and fork*), and semantic violations (*theme and fork*). In Experiment 1b, participants read the same stimuli without “and”. Experiment 1a revealed that binomials elicited larger P300s and smaller N400s compared to the other conditions, reflecting the activation of a ‘template’ that matches the upcoming information (P300) and pointing to easier semantic integration (N400). In contrast, no differences were observed between binomials and associates in Experiment 1b. We conclude that distinct mechanisms underlie the processing of predictable and novel sequences.

Keywords: multi-word expressions; binomials; frequency; predictability; ERPs; P300; N400.

INTRODUCTION

Although human language can be original and highly creative, we rarely fully exploit its creative potential. Evidence suggests that much of the language we encounter on a daily basis is ‘formulaic’. That is, words tend to co-occur in specific linguistic configurations, known as multi-word expressions (MWEs),¹ and language users tend to draw on a vast, yet limited, repertoire of MWEs. MWEs are familiar sequences of words, such as, collocations (*strong tea*), binomials (*time and money*), multi-word verbs (*rely on*), idioms (*ring a bell*), speech routines (*How’s it going?*), discourse markers (*on the other hand*), lexical bundles (*in the middle of*), and grammatical constructions (*the –er the –er*). A fundamental feature of MWEs is that they are extremely ubiquitous: a native speaker of English is believed to produce about four MWEs in every minute of discourse (Glucksberg, 1989; Pollio et al., 1977). Studies looking at monologues and conversations found that around one quarter of the speech analysed was formulaic in nature (Van Lancker Sidtis & Postman, 2006; Van Lancker & Rallon, 2004). Other estimates suggest that the number of MWEs in American English is comparable to the number of single words (Jackendoff, 1995). Such prevalence clearly makes MWEs an essential component of mature linguistic competence.

¹ In the present paper, we opted for the term *multi-word expressions*, because the focus is on strings of language longer than a single word. Another commonly used term to refer to familiar phrases is *formulaic language*. It is noteworthy, however, that formulaic language encompasses multi-word sequences, as well as single-word items, such as expletives and exclamations (*darn, wow*) and conversational speech formulas (*okay, right, yeah, hello*).

Because MWEs are highly familiar, conventional ways of expressing thoughts and ideas, they render our discourse natural and easily comprehensible. It has long been argued that it is easier and more economic to remember and use language in chunks – which are highly predictable in nature – than having to create novel combinations of words anew (Wray, 2002). In his *Idiom Principle*, Sinclair (1991) proposed that language users have available to them thousands of semi-preconstructed phrases that constitute “single choices”, even though they might be analyzable into individual components (p. 110). Indeed, it has long been acknowledged that chunking is an important strategy in linguistic processing, and that in order to be able to process linguistic input in a smooth and expedient fashion, one has to operate with larger linguistic units, such as chunks (Miller, 1956). It appears then, that the focus of neurolinguistic enquiry should be as much on MWEs, as it has been on single words.

In what follows below, we first consider what is arguably one of the defining characteristics of MWEs – predictability; a feature that has far reaching consequences for how MWEs are processed in the brain. We then turn to a more theoretical debate about the nature of the mental lexicon and the role of multi-word information.

Contextual predictability in language comprehension

Predictive mechanisms play an important role in language comprehension. A multitude of studies employing a range of paradigms and tasks have attributed their findings to the degree of contextual predictability. For example, eye movement studies have shown that predictability of a word, given the preceding sentential or phrasal context, is one of the key factors known to affect fixation durations, number of fixations, as well as the

likelihood of the word being skipped (Balota, Pollatsek, & Rayner, 1985; Rayner & Well, 1996). From a neurophysiological perspective, a negative deflection in the 250-500 ms time window peaking around 400 ms after stimulus onset – known as the N400 – has been linked to a word's predictability given the preceding context (Kutas & Hillyard, 1980, 1984; van Petten & Kutas, 1990). The N400 has also been associated with the reader's or listener's expectations based on their real-world knowledge (Hagoort et al., 2004). In Hagoort et al. (2004), sentences that violated participants' world knowledge (*The Dutch trains are white and very crowded*) resulted in a larger N400 effect than sentences in which this knowledge was not violated (*The Dutch trains are yellow and very crowded*).

Central to the description of the N400 is the concept of cloze probability (Kutas & Hillyard, 1984), which establishes the proportion of respondents that provide the correct completion of a phrase or a sentence. Cloze probability is known to be inversely correlated with N400 amplitudes (Kutas, Van Petten, & Kluender, 2006). The higher the cloze probability of a word, the smaller the N400 amplitudes. In Kutas and Hillyard (1980, 1984), unexpected sentence completions with low cloze probability (the word *hour* in *The bill was due at the end of the hour*) elicited larger N400s than expected ones with high cloze probability (the word *month* in *The bill was due at the end of the month*).

Evidence suggests that contextual cues shape word processing from its earliest stages and pre-activate the features of a likely upcoming word (Kutas & Federmeier, 2011). The predictive mechanisms are so strong that the processing of an unexpected word may, in fact, be facilitated if it shares some features, semantic or orthographic, with the expected stimulus (Laszlo & Federmeier, 2009; for a review, see Kutas &

Federmeier, 2011).

Contextual predictability has long been linked to the N400 component. However, recently, Lau, Namyst, Fogel, and Delgado (2016) pointed out that researchers have often tended to confound congruity and predictability. That is, congruous endings were also predictable given the preceding context (Kutas & Hillyard, 1980). In a series of experiments, Lau et al. (2016) observed reduced N400s for (equally congruous) predictable versus unpredictable adjective-noun pairs where the noun was kept constant (*mashed potato* vs. *shredded potato*), and for (equally unpredictable) congruous versus incongruous adjective-noun pairs where the noun was kept constant (*yellow bag* vs. *innocent bag*). While predictability had a large effect on N400 amplitudes, congruency was found to have only a small effect on the size of the N400. These results suggest that although higher predictability and greater congruency given the preceding word/s may result in the modulations of the same ERP component, N400 effects of predictability and (in)congruency are, in fact, due to different underlying mechanisms (Lau et al., 2016).

Predictive mechanisms and MWEs

While general sentence level constraints exert a powerful influence on how we process language, there are linguistic contexts whose canonical structure and meaning are believed to be stored in memory, and which lend themselves particularly well to the study of predictive mechanisms. If a native speaker of English is asked to complete the phrase “fish and ...”, “you can’t judge a book by its ...”, or “as a matter of ...”, the answers will invariably be the most expected completions *chips*, *cover*, and *fact* (see Van Lancker Sidtis, Cameron, Bridges, & Sidtis, 2015). These and thousands of other MWEs are

familiar, conventional ways of using language. As a result, they are highly, and often uniquely, predictable strings of language. This makes them ideal candidates for the investigation of predictive mechanisms in language comprehension.

In line with probabilistic models of language, information about the co-occurrence of words is represented in a speaker's mind (Gregory et al., 1999; Jurafsky, 1996; McDonald & Shillcock, 2003a, 2003b; Seidenberg & MacDonald, 1999). As McDonald and Shillcock (2003a, 2003b) argue, the vast amounts of language that a speaker encounters are a rich source of statistical knowledge about the way in which the language is used. The brain is capable of storing and subsequently using large amounts of statistical information during language comprehension to estimate the probability of *word n* following *word n-1*. In addition, integrating a word into one's mental lexicon also involves encoding its surrounding (phrasal) context into the mental lexicon (McDonald & Shillcock, 2003b; also see Goldberg, 1995, 2006; Langacker, 1987).

It has been argued that the expectations driven by a highly conventional string of language (the word *cover* following *you can't judge a book by its ...*) should be different from the more general discourse-based constraints, such as those observed in Kutas and Hillyard (1984). Specifically, the activation of the final word within a MWE should be stronger and more categorical than that of a plausible, but not uniquely predictable, word in a sentence that does not contain a MWE (Molinaro & Carreiras, 2010; Vespignani, Canal, Molinaro, Fonda, and Cacciari, 2010). Although electrophysiological research into MWEs is still in its infancy, current evidence tentatively points to two processes associated with the comprehension of such sequences compared to novel language: 1) easier semantic integration of familiar information, and 2) the activation of template

matching mechanisms for uniquely predictable linguistic information.

Of the many types of MWEs, idioms have received by far the most attention in ERP research. In one of the earliest such studies, Strandburg et al. (1993) recorded ERPs on the final word of idiomatic, literal (novel), and nonsensical phrases. Participants had to decide whether or not the phrase was meaningful. Smaller N400s were observed for idiomatic phrases compared to literal and nonsensical ones, with an ordered increase in the N400 amplitudes from idiomatic to literal to nonsensical phrases, suggesting progressive increases in their depth of processing. In another study by Laurent et al. (2006) participants performed a semantic relatedness task on French idioms and metaphors that varied in their degree of salience (strongly and weakly salient). The authors defined salient meanings as those foremost on our mind due to their frequency, conventionality, familiarity, or prototypicality. The authors observed smaller N400 amplitudes on the last word of the strongly salient idioms compared to the last word of the weakly salient novel metaphors. More recently, Vespignani et al. (2010) investigated the processing of Italian idioms before and after the recognition point (the point at which the idiom becomes uniquely recognizable). Three conditions were looked at: one idiomatic and two literal control conditions (substitution and violation) embedded in a sentence context. Similar to the earlier studies, Vespignani and colleagues (2010) found that idiomatic phrases elicited smaller N400s than literal phrases for the word that represented the recognition point of the idiom.

The finding of reduced N400s for conventional idioms in these studies was interpreted as an indicator of easier processing and semantic integration for salient figurative expressions than for their novel counterparts. The N400, however, is not the

only ERP component involved in the processing of highly predictable information. The P300 has also been implicated in phrasal processing, albeit in fewer studies (Molinaro & Carreiras, 2010; Vespignani et al., 2010). The P300 is a positive-going wave peaking around 300 ms after stimulus onset and is largest over parietal sites.² Classically, the P300 component has been associated with a *context-updating theory* (Donchin, 1981; Donchin & Coles, 1988; Donchin & Fabiani, 1991) and a *context-closure account* (Verleger, 1988). Both relate the P300 effects to expectancies that arise during stimulus processing. However, while the context-updating theory predicts larger effects for unexpected events, the context-closure theory accounts for the larger P300 in terms of the closure of certain expectations. In the context-updating theory, the P300 reflects an *expectancy violation*, while in the context-closure theory, the P300 reflects an *expectancy confirmation* (Riess Jones, 1988). Closure or confirmation have been described in terms of post-stimulus activities that lead to the decision that a stimulus belongs, or does not belong, to a particular class (Desmedt, 1980; Verleger, 1988).

The idea of closure or confirmation has important implications for MWEs. Vespignani et al. (2010) hypothesized that the expectations driven by a highly conventional formulaic chunk (idiom) should be different from those of general discourse-based constraints. In order to test this, ERP waveforms were compared for idiomatic and literal phrases before and after the recognition point. Before the recognition

² The P300 includes a number of components, of which the P3a and the P3b are the most common. The more anterior P3a is associated with unexpected events. The more posterior P3b has been shown to be elicited by infrequent task-relevant events (for a review, see Kok, 2001). In the present study, we focus on the P3b (henceforth, the P300).

point, the difference between the conditions was linked to the N400 (discussed above). After the recognition point, however, the idiomatic (but not literal) sentence completions resulted in a larger P300. Given that following the recognition point, only one idiom completion was possible, Vespignani et al. (2010) concluded that the observed P300 effect is the result of categorical template matching. This mechanism is thought to operate “for multi-word expressions ... when the compositional analysis must be integrated with the retrieval of prefabricated meaning from the semantic memory” (p. 1696).

Finally, Molinaro and Carreiras (2010) examined the processing of Spanish figurative and literal collocations embedded in sentence context. ERPs were measured on the final word of collocations and matched novel phrases. All collocations elicited a larger P300 than novel phrases. On the contrary, meaning variation (figurative vs. literal) affected ERPs only in the N400 time window. Molinaro and Carreiras (2010) concluded that two distinct mechanisms were at work: an earlier one determined by high predictability of an upcoming word (in the phrase *en cuerpo y alma* “in body and soul”, the first part *en cuerpo y ...* activates the final word *alma*), linked to the P300, and a later one involved in semantic processing routines, linked to the N400.

Although only two studies to date have linked the P300 to MWE processing, this component has also been observed in other highly constraining linguistic contexts. Roehm et al. (2007), replicating Kutas and Iragui (1998), had participants read sentences ending in highly expected antonym completions (*The opposite of black is ... white*), related but unexpected completions (*yellow*), or unexpected and anomalous ones (*nice*). The anomalous completion (*nice*) resulted in larger N400 amplitudes than the other

conditions. However, when participants were presented with the antonymous word pairs (*black/white*) in an antinomy judgement task, a positive peak was observed for the expected condition. Although the peak was within the N400 time window and appeared akin to a reduced N400, the authors interpreted it as the P300. Roehm and colleagues (2007) argued that the highly expected antonym reflected strong expectations and elicited the P300 component because “the correct identification of the predicted word does not require a lexical search (there is a unique prediction that may either be fulfilled or not)” (p. 1272).

Roehm et al. (2007) concluded that the P300 component signals are functionally distinct from the N400 level of predictive processing. However, as noted by Molinaro and Carreiras (2010), further evidence is needed to support the disassociation between the P300 and the N400 during on-line language comprehension, not least because the N400 can be observed within a relatively large 250-500 ms window and may peak earlier than the often cited 400 ms. As Molinaro and Carreiras (2010) argue, further evidence will allow us to ascertain whether the P300 does indeed represent lexical pre-activation with no or few demands for semantic integration as a result of strong categorical expectation.

A theoretical debate: A distinction between the lexicon and grammar

In contrast to figurative MWEs such as idioms (Bobrow & Bell, 1973; Cacciari & Tabossi, 1988; Cutting & Bock, 1997; Gibbs, 1980; Gibbs & Gonzales, 1985; Gibbs, Nayak, & Cutting, 1989; Swinney & Cutler, 1979), literal and compositional MWEs have received relatively little attention in psycholinguistic and, in particular, neurolinguistic research on language representation and processing. This is largely due to the dominance

of the linguistic theory known as the *words-and-rules approach*, which makes a distinction between the *lexicon*, a collection of memorized and stored forms, and *grammar*, a collection of rules that are applied to these forms (Pinker, 1991; Pinker, 1999; Pinker & Ullman, 2002; Ullman, 2001; Ullman et al., 2005). In line with the words-and-rules approach, the lexicon only contains morphemes, words, and highly idiosyncratic forms, such as idioms. Compositional language is not thought to be part of the lexicon. In this account, frequency effects should only be observed for forms at the word level, and never at the (compositional) phrase level, because the latter are thought to be necessarily computed.

Despite the dominance of this theory, a number of *frequency-based approaches* have argued against the distinction between a word and a phrase, or the lexicon and grammar (Abbot-Smith & Tomasello, 2006; Bod 2006; Bybee, 1998; Christiansen & Chater, 1999; Elman, 1990; Goldberg, 2006; Tomasello, 2003; Pierreumbert, 2001; Rumelhart & McClelland, 1986). The common view is that experiences with a linguistic form ultimately determine our memory representations. In this view, language is not seen as a set of grammar rules, but as a statistical accumulation of experiences that changes every time a linguistic unit, a word or a phrase, is encountered (Bod, 2006; Bybee, 2006). One prominent idea is that linguistic units do not exist in isolation, but form and function in dynamic *networks*, which are regularly updated to reflect the nature of the ever-changing linguistic experiences (Rumelhart & McClelland, 1986). Thus, unlike the words-and-rules approach, frequency-based theories posit that frequency effects are not specific to words and highly idiosyncratic items like idioms, but are, in fact, present for all linguistic exemplars, of any length, complexity, level of abstractness or literality. We will

come back to these theories again in the Discussion.

Frequency effects for words are well established (Balota et al, 2004; Brysbaert et al, 2011), such that higher frequency words are processed more quickly than lower frequency words. Recent behavioural and eye-tracking research has also examined frequency effects at the phrase level. A wealth of studies have demonstrated that MWEs are processed differently from novel language. Comprehension and production studies with idioms, collocations, binomials, lexical bundles, and phrasal verbs have shown that MWEs enjoy quantitatively faster processing times than control phrases (Arnon & Snider, 2010; Bannard & Matthews, 2008; Janssen & Barber, 2012; Janssen & Caramazza, 2011; Matlock & Hereida, 2002; Siyanova-Chanturia, Conklin, and van Heuven, 2011; Siyanova-Chanturia, Conklin, and Schmitt, 2011; Sosa & MacFarlane, 2002; Tremblay, et al., 2011; Tremblay & Tucker, 2011). These studies are important because they provide evidence for phrasal frequency effects in language comprehension and production, and argue (directly or indirectly) against the traditional distinction between grammar and the lexicon. Empirical evidence further suggests that language users store frequency information not just with respect to single words but also larger phrasal units, pointing to the parallels in the processing of words and phrases (Arnon & Snider, 2010; Bannard & Matthews, 2008; Siyanova-Chanturia, Conklin, & van Heuven, 2011) and indicating that all linguistic information may be governed by similar cognitive mechanisms (Arnon & Christiansen, 2017; Arnon, McCauley, Christiansen, 2017;

Christiansen & Chater, 1999; Elman, 2009; also see Christiansen & Arnon, 2017).³

What is still largely missing from the grammar-lexicon debate, however, and from the MWE literature more broadly, is electrophysiological evidence pertinent to the processing of those MWEs that are literal and compositional in nature. The majority of ERP studies to date have focused on figurative language, such as idioms (Laurent et al., 2006; Rommers, Dijkstra, & Bastiaansen, 2013; Strandburg et al., 1993; Vespignani et al., 2010). The aim of the present study was to address this gap in the literature by using ERPs to investigate the cognitive processes involved in the comprehension of English binomial expressions – phrases that are highly familiar and predictable (akin to idioms), frequent (much more so than idioms), and literal (unlike idioms). Our main research question thus centred on the electrophysiological markers involved in the processing of highly conventional, compositional phrases during on-line language comprehension. In

³ It is important not to overestimate the role of frequency for some MWEs. While it is an important factor in the way in which many MWEs are processed (such as collocations, binomials and lexical bundles), frequency may play a lesser role for idioms and proverbs. For example, while the expressions *raining cats and dogs* and *everything but the kitchen sink* are undoubtedly familiar, conventional strings of language, they are, nevertheless, infrequent, appearing just twice in the British National Corpus. For such MWEs, familiarity and predictability, rather than frequency, play a major role (Hallin & Van Lancker Sidtis, 2017; Moon, 1998a, 1998b; Rammell, Pisoni, & Van Lancker Sidtis, in press; Van Lancker Sidtis, 2015). In addition, studies have shown that idioms vary in semantic decomposability, a factor that may also affect how “chunked” such strings may appear (Libben & Titone, 2008; Titone & Libben, 2014).

addition, we wanted to clarify the nature of multi-word information in the mental lexicon.

THE PRESENT STUDY

The main goal of this study was to explore electrophysiological responses to a highly expected final word within a conventional phrase, such as binomial expressions. The present investigation consists of two ERP experiments. In Experiment 1a, participants' brain activity was recorded as they read three types of phrases: (1) frequent binomial expressions (*knife and fork*), (2) infrequent novel phrases similar in association strength to the binomial condition (*spoon and fork*), and (3) non-associated, unattested semantic violations (*theme and fork*). In binomials and associates, the two content words were matched for the strength of association. This enabled us to directly compare the electrophysiological responses of target words (*fork*) that are part of familiar linguistic patterns and novel phrases.

In Experiment 1b, participants read the same stimuli as in Experiment 1a but without the conjunction “and” (*knife-fork*, *spoon-fork*, *theme-fork*). If frequent and predictable phrases are processed differently from novel phrases at the electrophysiological level, then the waveforms elicited by, for example, *knife and fork* (binomial) and *spoon and fork* (novel phrase) should differ in Experiment 1a. However, in Experiment 1b, when the conjunction “and” is removed, the linguistic form no longer matches the representation (mental template) and should thus lead to comparable waveforms for the two conditions.

Based on the previous findings (Molinaro & Carreiras, 2010; Vespignani et al., 2010), we hypothesized that if a processing advantage for frequent binomial phrases over

novel phrases is due to binomials being uniquely predictable sequences of words (if binomial expressions represent a template), then larger P300 amplitudes should be elicited for binomials relative to novel phrases. Thus, in Experiment 1a, seeing *knife and* should set up an expectation for the word *fork*. In contrast, in Experiment 1b, seeing *knife* (without “and”) should not activate the template, and, thus, no differences between *knife-fork* and the equally strongly associated *spoon-fork* should be observed. Because the N400 component has also been implicated in the processing of frequent and predictable phrasal configurations (Laurent et al., 2006; Strandburg et al., 1993; Paulmann, Ghareeb-Ali & Felser, 2015; Vespignani et al., 2010), we expected the predictability of the second content word in the binomial expression to result in the modulations of the N400 component. Critically, because of the distinct hypotheses with respect to the two experiments in terms of the P300 and N400, we expected to find a significant interaction between the two critical conditions (binomial and associate) and experiment type (Experiment 1a and Experiment 1b) both in the early P300 and the late N400 time windows.

METHODOLOGY

Participants

Forty-eight native speakers of English at a University in the USA were paid \$20 or were given course credit for their participation (23 females; age 18–30; mean age 20.2). The experiment consisted of two parts: Experiment 1a and 1b. The same participants took part in both experiments. The order was fixed: Experiment 1a first and Experiment 1b second. One male participant took only part in Experiment 1a, bringing the total number of

participants in Experiment 1b to 47 (23 females; age 18–30; mean age 20.2). All participants were right-handed as assessed by the Edinburgh handedness inventory (Oldfield, 1971) and had normal or corrected-to-normal vision. The subjects' consent was obtained according to the Helsinki Declaration. An approval to carry out the present study was obtained from the University Institutional Review Board.

The data from 30 (Experiment 1a) and 28 (Experiment 1b) participants were included in the analysis (Experiment 1a: 13 females; age 18–30; mean age 21; Experiment 1b: 12 females; age 18–25; mean age 20). In the retained participants over 50% of the trials were artifact-free. The participant loss was almost entirely due to excessive blinking during reading target sequences.

Stimulus materials

In Experiment 1a, 120 matched triplets were selected. The target stimuli belonged to one of the following conditions: (1) frequent binomial expressions whose two content words were strongly associated (*knife and fork*); (2) grammatically plausible but infrequent (novel) three-word combinations, whose two content words were as strongly associated as the two words in the binomial condition, and which contained the same second content word as did the binomial condition (*spoon and fork*); and (3) non-associated, semantic violation phrases ending in the same content word (*theme and fork*).

Binomial expressions and their frequencies were extracted from the British National Corpus (BNC, <http://www.natcorp.ox.ac.uk/>). Associates were created by replacing the first content word of the binomial condition with a word that was semantically related to the second content word of the binomial condition. The items in

the semantic violation condition were created by replacing the first content word of the binomial condition with a semantically unrelated word. Associate and violation frequencies were extracted from the BNC. The binomials used in the experiment consisted of a conjunction (*and*) and two content words, which were of the same part of speech ('N and N', 'V and V', 'Adj and Adj'). The critical word was always the second content word. The University of South Florida Free Association Norms database (<http://w3.usf.edu/FreeAssociation/>) was used to match the constituents of binomials (*knife and fork*) and associates (*spoon and fork*) in semantic association strength (the forward association was measured). The statistical analysis showed that the two content words in the binomial and associate conditions were equally strongly associated ($t(119) = 1.6, p > .1$). Content words in the semantic violation condition were not associated. The properties of the three experimental conditions are presented in Table 1.

Table 1. Mean phrasal frequency, phrasal length and semantic association strength for the three conditions: binomials, associates and semantic violation phrases (SD in parentheses).

	Binomial	Associate	Violation
Phrasal frequency (BNC)	102.0 (143.0)	0.7 (1.2)	0 (0)
Phrasal length (characters)	13.6 (2.7)	14.6 (2.8)	14 (2.5)
Association strength	0.21 (0.2)	0.25 (0.2)	0 (0)

The materials used in Experiment 1b were identical to those used in Experiment 1a, except for the conjunction “and” which was removed, resulting in: (1) binomials (*knife-fork*); (2) associates (*spoon-fork*); (3) semantic violations (*theme-fork*).

In addition to the three experimental conditions, 40 filler phrases of the type ‘N and

N' were used in Experiment 1a and 1b. They were included for the animal categorization task (see below) and constructed such that half of them contained a word denoting an animal (always a noun) in the first position (Experiment 1a: *lion and prey*; Experiment 1b: *lion-prey*), and the other half had the word denoting an animal in the second position (Experiment 1a: *nest and eagle*; Experiment 1b: *nest-eagle*).

EEG procedure

The 120 target items (40 items of each of the three conditions) were intermixed with 40 fillers. Three counterbalanced lists were created each with 160 items. To encourage participants to read for comprehension, we conducted a secondary go/no-go animal categorization task. All filler items contained exactly one word denoting an animal, while none of the experimental items contained 'animal' words.

The words in each trial were presented using rapid serial visual presentation, that is, one word at a time. Each trial consisted of three (Experiment 1a: *knife, and, fork*) or two (Experiment 1b: *knife, fork*) words, during which participants were asked not to blink. Participants were instructed to read all words for comprehension and to press a designated button as soon as they saw a word denoting an animal. The words were presented in white lower case letters in Courier New font against a black background in the centre of a VGA computer screen. The viewing distance was approximately 100 cm.

Each trial started with a 500 ms fixation cross. The inter-stimulus interval (ISI) was 200 ms and each individual word remained on the screen for 300 ms. There was a longer, 1000 ms, ISI after the last (critical) word to delay participants' eye blinks. After the last word of each trial, there was a 2000 ms blank screen with the word BLINK in the middle

indicating that participants could blink and get ready for the next trial. Participants were urged not to start blinking until they saw the word BLINK and to stop blinking as soon as the word BLINK disappeared. Experiment 1a lasted approximately 20 minutes and consisted of two blocks with a break in between the blocks during which an impedance check was performed. Participants, seated comfortably in a sound proof booth, completed a practice session before the experiment started.

The EEG procedure and stimuli for Experiment 1b were identical to that of Experiment 1a, except that the conjunction “and” was removed from the stimuli. Experiment 1b lasted approximately 15 minutes and also consisted of two blocks with a break during which an impedance check was performed.

All participants took part in Experiment 1a (stimuli with conjunction) and then Experiment 1b (stimuli without conjunction). The experiments happened on the same day, one after the other. The 120 target triplets were counterbalanced across the three lists so that each participant read only one version of each triplet. To avoid repetition as much as possible, if a participant saw List 1 in Experiment 1a (*knife and fork*), they then saw List 2 (*spoon-fork*) or 3 (*theme-fork*) in Experiment 1b. Thus, no participant saw identical trials in the two experiments. Although the repetition of the second content word (*fork*) was unavoidable, it occurred across all conditions and, therefore, it equally affected the three conditions.

EEG recording

EEG was recorded from 39 Ag/AgCl scalp electrodes, using an elastic cap with active shielding (Easy-Cap) combined with an ANT amplifier (ANT software B.V., Enschede,

the Netherlands). Electrode positions were: midline - Fz, FCz, Cz, CPz, Pz; lateral left/right - FP1/2, F7/8, F5/6, F3/4, FT7/8, FC5/6, FC3/4, T7/8, C5/6, C3/4, TP7/8, CP5/6, CP3/4, P7/8, P5/6, P3/4, O1/2. Horizontal and vertical EOG was recorded from electrodes placed on the outer canthi, and below and above the right eye, respectively. Two additional electrodes were placed on the right (A2) and left (A1) mastoids. The signal was acquired using the left mastoid as a common reference and was sampled at a rate of 512 Hz. Electrode impedance was kept below 5 k Ω during the recording.

Analysis and results

Brain Vision Analyzer was used for data analysis. EEG recordings were re-referenced off-line to the average activity of the two mastoids and filtered with a band-pass of 0.01–30Hz. Artifacts due to eye movements were rejected by means of a visual inspection. Artifacts due to muscle activity exceeding ± 100 mV in amplitude were also rejected. On average, 31% of trials in Experiment 1a (binomials: 30.9%, associates: 31.5%, violations: 30.6) and 32.2% in Experiment 1b (binomials: 33.4%, associates: 31.5%, violations: 31.9%) were excluded from the analysis due to ERP artifacts. There were no differences in the number of rejections between the conditions in either of the two experiments (all $P_s > .3$). For each target noun, an epoch of 1100 ms was obtained including a 100 ms pre-stimulus baseline. Average ERP waveforms were computed time-locked to the onset of the second content word.

The following factors were included in the statistical analyses: Type (binomial, associated, violation), Anteriority (anterior: F3, F4, F5, F6, Fz, FC5, FC6, FC3, FC4, FCz; central: C5, C3, Cz, C4, C6; posterior: CP5, CP3, CP4, CP6, CPz, P3, P4, P5, P6,

Pz), and Laterality (left: F5, F3, FC5, FC3, C5, C3, CP5, CP3, P5, P3; medial: Fz, FCz, Cz, CPz, Pz; right: F6, F4, FC6, FC4, C6, C4, CP6, CP4, P6, P4). The Greenhouse-Geisser procedure was applied on every within-subject effect where the sphericity assumption was violated. The time windows for the statistical analyses were chosen based on the visual inspection and earlier research (Molinaro & Carreiras, 2010).

Specifically, the P300 window was defined as 250-350 ms after stimulus onset, while the N400 window was defined as 350-450 ms after stimulus onset.

Experiment 1a (stimuli with conjunction)

250-350 ms. A 3 (Type: binomial, associate, violation) x 3 (Anteriority: anterior, central, posterior) x 3 (Laterality: left, medial, right) repeated-measures ANOVA showed a significant main effect of Type, $F(2,58) = 12.87, p < .0001$. The three conditions differed from each other (Figure 1), with binomials eliciting the largest positivity ($M = 5.20$), violations eliciting the smallest positivity ($M = 2.28$), and the associated condition being in the middle ($M = 4.02$). The planned comparison binomials vs. associates revealed a trend, $t(29) = 1.95, p = .06$, whereas the other planned comparisons were significant: binomials vs. violations, $t(29) = 4.81, p < .0001$; associates vs. violations, $t(29) = 3.32, p = .002$ (Figure 2). Importantly, we found a significant interaction between Type and Laterality, $F(4, 116) = 4.98, p = .002$. Planned comparisons revealed that while the differences between violations and the other two conditions were prominent across left (binomials vs. violations: $t(29) = 4.16, p = .001$; associates vs. violations: $t(29) = 2.93, p = .006$), medial (binomials vs. violations: $t(29) = 4.84, p = .0001$; associates vs. violations: $t(29) = 3.42, p = .001$) and right sites (binomials vs. violations: $t(29) = 5.23, p$

<.0001; associates vs. violations: $t(29) = 3.32, p = .002$), the difference between the two critical conditions – binomials and associates – was right lateralised (right: $t(29) = 2.36, p = .02$; left/medial: $t(29) = 1.69, p = .10$). No other two- or three-way interactions with Type were significant (all $P_s > .4$)

350-450 ms. A 3 (Type: binomial, associate, violation) x 3 (Anteriority: anterior, central, posterior) x 3 (Laterality: left, medial, right) repeated-measures ANOVA revealed a significant main effect of Type, $F(2,58) = 25.91, p < .0001$ (Figure 1). Planned comparisons showed that the three conditions differed from each other, with violations eliciting the largest negativity ($M = -0.10$), binomials eliciting the smallest negativity ($M = 5.12$), and the associate condition being in the middle ($M = 3.18$): binomials vs. associates, $t(29) = 2.88, p = .007$; binomials vs. violations, $t(29) = 6.72, p < .0001$; associates vs. violations, $t(29) = 4.37, p = .0001$ (Figure 3). In addition, we found a significant interaction between Type and Laterality, $F(4, 116) = 8.67, p < .0001$. Planned comparisons revealed that the differences between violations and the other two conditions were highly significant across left (binomials vs. violations: $t(29) = 5.98, p < .0001$; associates vs. violations: $t(29) = 3.86, p = .001$), medial (binomials vs. violations: $t(29) = 6.66, p < .0001$; associates vs. violations: $t(29) = 4.52, p < .0001$) and right sites (binomials vs. violations: $t(29) = 7.33, p < .0001$; associates vs. violations: $t(29) = 4.58, p < .0001$). The differences between the two critical conditions – binomials vs. associates – were also evident across all sites, although the effect was strongest across the right sites (left: $t(29) = 2.66, p = .01$; medial: $t(29) = 2.47, p = .02$; right: $t(29) = 3.41, p = .001$).

Finally, the two factors Type and Anteriority were found to interact significantly, $F(4, 116) = 4.91, p = .008$. Planned comparisons revealed that the differences between

violations and the other two conditions were significant across anterior (binomials vs. violations: $t(29) = 5.18, p < .0001$; associates vs. violations: $t(29) = 3.80, p = .001$), central (binomials vs. violations: $t(29) = 7.03, p < .0001$; associates vs. violations: $t(29) = 4.39, p < .0001$) and posterior sites (binomials vs. violations: $t(29) = 7.64, p < .0001$; associates vs. violations: $t(29) = 4.64, p < .0001$). Although the differences between binomials and associates were also observed across all sites, this effect appeared to be strongest across the central and posterior sites (anterior: $t(29) = 2.19, p = .03$; central: $t(29) = 3.09, p = .004$; posterior: $t(29) = 3.12, p = .004$). The three-way Type x Laterality x Anteriority interaction was not significant, $F(8, 232) = 1.90, p = .09$.

[Figure 1 about here]

[Figure 2 about here]

[Figure 3 about here]

Summary of findings (Experiment 1a). In the early time window (250-350 ms), binomials and associates elicited larger positivity than semantic violations. Importantly, binomials elicited larger positivity relative to the associate condition across the right lateral sites. In the late time window (350-450 ms), binomials elicited smaller negativity than semantic violations and, crucially, associates. Experiment 1a suggests that frequent and predictable phrases elicit a larger P300 and a smaller N400 compared to novel phrases.

Experiment 1b (stimuli without conjunction)

250-350 ms. A 3 (Type: binomial, associate, violation) x 3 (Anteriority: anterior, central, posterior) x 3 (Laterality: left, medial, right) repeated-measures ANOVA revealed no significant main effect of Type, $F(2,54) = 1.83, p = .21$ (Figures 4 and 5). In addition, no significant two- and three-way interactions were observed with the factor Type (all $P_s > .09$).

350-450 ms. A 3 (Type: binomial, associate, violation) x 3 (Anteriority: anterior, central, posterior) x 3 (Laterality: left, medial, right) repeated-measures ANOVA showed a significant main effect of Type, $F(2,54) = 12.87, p = .0001$ (Figure 4). Planned comparisons revealed no significant differences between binomials ($M = 1.65$) and associates ($M=1.30$), $t(27) = .50, p = .61$; while the two conditions were found to differ significantly from violations ($M = -1.20$): binomial vs. violation, $t(27) = 4.23, p = .0002$; associates vs. violation, $t(27) = 5.55, p < .0001$ (Figure 6). Furthermore, a significant interaction was found between Type and Laterality, $F(4, 108) = 9.28, p < .0001$. Planned comparisons showed no significant differences between binomials and associates across left, medial or right sites (all $P_s > .37$); while both conditions differed significantly from violations across the three lateral regions (all $P_s < .001$). No other two- or three-way interactions involving Type were found significant (all $P_s > .09$).

[Figure 4 about here]

[Figure 5 about here]

[Figure 6 about here]

Summary of findings (Experiment 1b). In both the early and late time windows (250-350 ms and 350-450 ms), no differences were observed between binomials and associates. As expected, in the late time window (350-450 ms), semantic violations elicited a larger N400 than binomials and associates.

Comparison between the stimuli with (Experiment 1a) and without the conjunction (Experiment 1b)

As mentioned earlier, we expected to find an interaction between the critical conditions (binomial vs. associate) and whether or not the stimuli had the conjunction (Experiment 1a vs. Experiment 1b). Furthermore, the results of the analyses of Experiments 1a and 1b showed that there was a significant difference between binomials and associates only in Experiment 1a. To confirm whether an interaction could indeed be found in the combined data, a 2 (Type: binomial, associate) x 2 (Conjunction: with [Experiment 1a], without [Experiment 1b]) ANOVA of the mean voltages across the right lateral sites was conducted for the 250-350 ms and 350-450 ms time windows. The following electrodes were included in the analysis: F6, F4, FC6, FC4, C6, C4, CP6, CP4, P6, P4. As expected, we found a significant interaction for both time windows (250-350 ms: $F(1,27) = 5.62, p = .025$; 350-450 ms: $F(1,27) = 5.34, p = .03$).⁴ The two interactions are graphically presented in Figure 7.

⁴ These analyses included 28 participants. However, the data from four participants were included in only one of the two experiments, due to a large number of artefacts associated with the other experiment. Therefore, we also conducted an analysis with the data of 24 participants who were retained in both experiments. A comparable pattern of results was

[Figure 7]

GENERAL DISCUSSION

The aim of the present study was to examine the electrophysiological components involved in the processing of frequent, familiar and highly predictable word sequences – binomial expressions. In addition, we explored the suggestion made in previous studies that the P300-N400 complex is involved in the processing of conventional language (Molinaro & Carreiras, 2010; Vespignani et al., 2010). However, unlike these earlier studies, we focused on literal compositional MWEs which have so far received almost no attention in neurolinguistic research. To this aim, two ERP experiments examined the processing of English binomial expressions versus infrequent but grammatically correct novel phrases and phrases with semantic violations.

The analyses of Experiment 1a revealed that binomials elicited larger positivity just before 300 ms post stimulus relative to novel but equally strongly associated phrases. The effect was found across the right lateral sites. In line with previous research, we interpret this effect as the P300. Furthermore, the ERPs elicited by the binomial condition and the associate condition around 400 ms post stimulus are compatible in topographic distribution and timing with the N400. Specifically, smaller N400 amplitudes were

found, although due to the reduced number of participants, the interaction for the second window was only a strong trend (350-450 ms: $F(1,23) = 4.22, p = .05$), whereas the interaction for the first window was still significant (250-350 ms: $F(1,23) = 4.64, p = .04$).

observed in the binomial condition relative to the novel condition. The effect was observed across the entire scalp, but was greater over the right sites. When phrases were presented without the conjunction “and” (Experiment 1b), binomial expressions and associates exhibited comparable waveforms in the P300 and N400 time windows. Experiment 1b clearly showed that the second content word in a frequent and strongly associated phrasal configuration and the same word in an infrequent but equally strongly associated phrase elicited very similar waveforms when presented without the conjunction “and”. As expected, the classic N400 effect was observed in the semantic violation condition relative to the other conditions in both experiments.

As pointed out by Molinaro and Carreiras (2010), a traditional account of the N400 might argue that the observed P300 is, in fact, an early N400, given that the latter is known to span 250-500 ms in duration. For example, Lau et al. (2016) observed larger early positivity peaking around 250 ms for the predictable noun in an adjective-noun pair (*mashed potato* vs. *shredded potato*), followed by a reduced N400 relative to the baseline. Both were interpreted as reflecting the same – N400 – component. However, we do not believe that the P300 simply reflects modulations of the N400, or that it is an earlier onset of the N400. Visual examination of Figure 1 suggests a pronounced positive peak for the binomial condition relative to the novel condition just before 300 ms post stimulus (Experiment 1a). In contrast, the two waveforms converged in Experiment 1b. The P300-N400 distinction is further motivated by theoretical accounts specific to the P300 and expectancy confirmation (Kok, 2001; Verleger, 1988). As Verleger (1988) argues, the P300 should be related to the closeness of the match, and inversely related to the difficulty of the task. Critically for the processing of uniquely predictable linguistic

information, a number of researchers have linked the P300 effect to *template matching* (Chao, Nielsen-Bohlman, & Knight, 1995; Ford, 1978), wherein participants develop a neural representation, or a template, of the upcoming stimulus – the closer the match between the incoming information and the template, the larger the amplitude of the P300 (Kok, 2001).

We interpret the early positivity observed in highly constraining contexts as the manifestation of template matching mechanisms, wherein the target sequence, being highly predictable, activates a template in the lexicon that matches the upcoming information. The involvement of the P300 in template matching has been previously reported in the studies on figurative phrasal processing. In Vespignani et al. (2010), idiomatic sentence completions resulted in larger P300s relative to novel phrases after the recognition point. Vespignani and colleagues concluded that the observed P300 effect was the result of categorical template matching that specifically operates for MWEs, albeit in their case for figurative MWEs. Molinaro and Carreiras (2010) found that both literal and figurative Spanish collocations elicited a P300 compared to infrequent novel phrases. The authors argued that the recognition of the conventional string (collocation) led to the pre-activation of the lexical item that completed the collocation. Similar to Molinaro and Carreiras (2010), we argue that the P300 and N400 represent distinct stages of word processing, one associated with the recognition of uniquely predictable prefabricated routine (leading to increased P300s) and the other associated with facilitated processing and semantic integration (eliciting reduced N400s).

It is noteworthy that the elicitation of the P300 component in uniquely predictable contexts has been observed not only in the case of MWEs (Molinaro & Carreiras, 2010;

Vespignani et al., 2010), but also with other types of highly constraining contexts (Fisher et al., 2010; Roehm, et al., 2007). In the study looking at the processing of antonyms, Roehm et al. (2007) attributed the P300 to the correct identification of the uniquely expected word, given the preceding context, which does not require a lexical search. This proposition is consonant with earlier views that an increased P300 reflects speakers' awareness that a given stimulus belongs, or does not belong, to the category of a memorized target event (Kok, 2001).

A comparable early positivity has also been observed with mathematical calculations. Fisher et al. (2010) examined ERPs associated with correct or erroneous calculations, such as additions and divisions. Whereas erroneous calculations led to larger negative amplitudes around 400 ms post stimulus, interpreted as the N400, correct answers to both addition and division resulted in positivity peaking around 300 ms after stimulus onset. Although the positivity is not discussed in the study, the morphological similarity based on the visual inspection between the positivity in Fisher et al. (2010) and the P300 reported in the present investigation for binomials is apparent. It is crucial to note that the correct answer in mathematical calculations ($12+3=15$) is uniquely expected with no other correct answer possible, just as *white* is the only opposite of *black* (Roehm et al, 2007), just as *fork* is the most likely completion of *knife and*. It thus appears that a wide spectrum of uniquely predictable scenarios, both linguistic and non-linguistic, can modulate P300 amplitudes.

In addition to larger P300s amplitudes, binomial processing was characterized by a reduction of the N400 relative to novel phrases and phrases with semantic violation (Experiment 1a), suggesting differences in the processing load for conventional versus

novel language. Similar to earlier research on idioms (Laurent et al., 2006; Strandburg et al., 1993), we interpret smaller N400s elicited by binomials as evidence for reduced processing load and easier semantic integration of familiar phrases.

One of the key findings of the present study pertains to Experiment 1b. When the stimuli were presented without the conjunction “and”, the ERP waveforms on binomials (*knife-fork*) and associates (*spoon-fork*) were found to be similar in both time windows. This finding, in particular with respect to the early positivity, allows us to conclude that what is responsible for the processing differences between binomials and associates in Experiment 1a is nothing but the phrasal, prefabricated, conventional status of binomial expressions, which is why this difference is absent in Experiment 1b, where word sequences are no longer presented in their highly predictable and thus uniquely identifiable form. In other words, in Experiment 1a – but not Experiment 1b – binomial expressions matched the existing template represented in the memory of a speaker.

Our discussion has so far addressed the question of electrophysiological components sensitive to MWE processing. However, the results of the present investigation are also important for the theoretical debate about the nature of the mental lexicon. As was mentioned in the Introduction, two distinct views exist with regard to the mental lexicon. On the one hand, the words-and-rules approach makes a clear distinction between the *lexicon* and *grammar* (Pinker, 1991; Pinker, 1999; Pinker & Ullman, 2002; Ullman, 2001; Ullman et al., 2005). In line with this view, the lexicon contains morphemes, single words, and highly idiosyncratic non-compositional figurative phrases, such as idioms. Literal compositional phrases (such as the binomial expressions used in the present experiment) are not thought to be part of the lexicon and are necessarily

computed during language processing. As a result, frequency effects should only be observed for forms at the word level, and never at the compositional phrase level.

On the other hand, frequency based-approaches, such as usage-based ⁵ (Bybee, 1998; Goldberg, 1995, 2006; Langacker, 1987; Tomasello, 2003) and exemplar-based theories (Bod, 1998, 2006; Pierrehumbert, 2001) have argued against the distinction between the lexicon and grammar. At the core of these theories lies the idea that language learning and processing are affected by the amount of experience that language users have with linguistic exemplars. Importantly, all linguistic material is represented and processed in a comparable way, and frequency effects should be equally observable in smaller, as well as larger units. As Bod (2006) notes, the allocation of representations to linguistic exemplars is accomplished purely on the basis of statistics. According to exemplar models, each instance of a linguistic form a speaker is exposed to is stored in memory in a way that acoustic, articulatory, semantic, contextual and other details are preserved (Bybee & McClelland, 2005). Exemplar models have been drawn on, for example, by Pierrehumbert (2001) to explain frequency effects in the articulation of frequent versus infrequent exemplars (in terms of phonetic reduction). Similarly, usage-based theory adopts a non-minimalist approach to linguistic representation, in which vast and rich experiences with language are stored as exemplars (Bybee & Beckner, 2009). As Bybee (2006) argues, new experiences with linguistic exemplars play a key role in shaping memory representations. As a result, frequently used sequences become more accessible, more entrenched and better integrated (Bybee & McClelland, 2005).

Consequently, usage-based and exemplar views predict more expedient processing for all

⁵ The term ‘usage-based’ was originally coined by Langacker (1987).

frequent events over less frequent ones, irrespective of their internal structure or meaning. Much of recent work within the realm of usage-based linguistics has focused specifically on multi-word processing in a first and second, child and adult language (Arnon & Snider, 2010; Arnon, McCauley, & Christiansen, 2017; Hernandez, Costa & Arnon, 2016; Siyanova-Chanturia, Conklin & van Heuven, 2011; also see Christiansen & Arnon, 2017).

Another approach to language acquisition and processing is the connectionist approach (Christiansen & Chater, 1999; Elman, 1990; Rumelhart & McClelland, 1986). According to connectionist theory, linguistic units do not exist in isolation; rather, they form and exist in relationships (networks) with each other. Similar to usage-based and exemplar models, connectionism puts emphasis on statistical properties of the input in language learning and processing (Saffran, Aslin, & Newport, 1996) and argues that the same mechanisms operate for regular and irregular forms. As Harris (1996) suggests, a lexicon containing variable-sized units fits well into a connectionist framework, according to which the units of representation are not part of a ‘fixed architecture’, but appear via experience and extracting regularities. The more strongly associated the structures are, the more likely they are to facilitate and prime each other.

Usage-based, exemplar-based, and connectionist models, often conceived as emergentist approaches to language acquisition, processing and use,⁶ differ in a number of ways (for a discussion, see Bybee & McClelland, 2006). For example, while the focus

⁶ The term ‘emergentist’ is used in the sense that language structure emerges, rather than is given a priori, from language use (Bybee & Beckner, 2009; Bybee & McClelland, 2005).

of exemplar and usage-based models is on exemplar storage and entrenchment, connectionist approaches posit that language knowledge is not stored in the form of exemplars (or rules) but in the form of connections between processing units, and the changes to the strengths between these connections (Bybee & McClelland, 2005; McClelland & Rumelhart, 1985; Rumelhart & McClelland, 1986; also see Elman, 1990). However, what is clear is that the above approaches have in common a key tenet that there is no obvious distinction between the lexicon and grammar, or words and rules, or between stored and computed linguistic units, and thus all linguistic information, irrespective of its (ir)regularity, idiosyncrasy, form or meaning is processed in a similar fashion and should be similarly affected by frequency. What shapes our knowledge and sensitivity to general and specific information is the many and varied experiences with language forms and their contexts of use (Bybee & McClelland, 2005). Experience thus has an enduring effect on our mental representations (Bybee & Beckner, 2009).

The finding of the electrophysiological differences in the processing of frequent versus novel phrases sheds further light on the nature of the lexicon, suggesting that together with memorized arbitrary forms, such as morphemes, words, and idioms, the lexicon also contains regular compositional phrases. The present study provides important electrophysiological evidence in support of the view that the human processor is sensitive to phrasal frequency, just as it is tuned to lexical (single word) frequency.

CONCLUSION

In this study, we employed English binomial expressions in order to examine the electrophysiological responses to frequent, familiar and highly predictable compositional

multi-word expressions. Distinct mechanisms were found to underlie the processing of predictable and novel language, as evidenced by an increased P300 and a reduced N400 for the former. This finding supports the view that frequent multi-word expressions are characterized by a reduced processing load and easier semantic integration, as well as pre-activation of the mental template that uniquely matches the unfolding configuration. In addition, the findings reported in the present study add to our understanding of the nature of the lexicon, contesting the lexicon-grammar dichotomy and offering support to frequency-based accounts of language acquisition, processing and use.

ACKNOWLEDGMENTS

XX was supported by an ESRC grant PTA-XXX-XXXX-XXXXX and the University of X prize. XX was supported by the Spanish Ministry of Economy and Competitiveness (PSI2014-54500-P) and the Basque Government (Grant PI_2015_1_25).

REFERENCES

- Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review*, 23, 275–290.
- Arnon, I. & Christiansen, M. H. (2017, early view). The role of multiword building blocks in explaining L1-L2 differences. *Topics in Cognitive Science*.
- Arnon, I., McCauley, S., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-Acquisition effects for multiword phrases, *Journal of Memory and Language*, 92, 265-280.
- Arnon, I., & Snider, N. (2010). More than words: frequency effects for multi-word phrases. *Journal of Memory and Language*, 62, 67-82.
- Arnon, I., McCauley, S.M. & Christiansen, M.H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language*, 92, 12(1).
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, 133, 283–316.
- Balota, D., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, 17, 364-390.
- Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: the effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, 19, 241-248.
- Bobrow, S. & Bell, S. (1973). On catching on to idiomatic expressions. *Memory and Cognition*, 1(3), 343-46.

- Bod, R. (1998). *Beyond grammar: An experience-based theory of language*. Center for the Study of Language and Information, CA.
- Bod, R. (2006). Exemplar-based syntax: How to get productivity from exemplars. *The Linguistic Review*, 23, 291-320.
- Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bólte, J., & Böhl, A. (2011). The word frequency effect: a review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, 58(5), 412-424.
- Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82, 711-733.
- Bybee, J., & Beckner, C. (2009). Usage-based theory. In Heine, B., & Narrog, H. (Eds.) *The Oxford handbook of linguistic analysis, Oxford handbooks online*, 1-30.
- Bybee, J. (1998). The emergent lexicon. In *Chicago Linguistic Society*, 34, 421-435.
- Bybee, J., & McClelland, J. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*, 22, 381-410.
- Cacciari, C. & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, 27, 668-83.
- Chao, L., Nielsen-Bohlman, L., & Knight, R. (1995). Auditory event-related potentials dissociate early and late memory processes. *Electroencephalography and Clinical Neurophysiology*, 96, 157-68.
- Christiansen, M.H. & Arnon, I. (2017, early view). More than words: the role of multiword sequences in language learning and use. *Topics in Cognitive Science*.

- Christiansen, M., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23, 157-205.
- Cutting, J., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory and Cognition*, 25(1), 57-71.
- Desmedt, J. (1980). P300 in serial tasks: An essential post-decision closure mechanism. In H. Kornhuber & L. Deecke (Eds.), *Motivation, Motor and Sensory Processes of the Brain: Electrical Potentials, Behavior and Clinical Use*. Amsterdam: Elsevier, 682– 686.
- Donchin, E. (1981). Surprise . . . surprise? *Psychophysiology*, 18, 493-513.
- Donchin, E., & Coles, M. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, 11, 357-74.
- Donchin, E., & Fabiani, M. (1991). The use of event-related brain potentials in the study of memory: Is P300 a measure of event distinctiveness? In J. R. Jennings and M. Coles (Eds.) *Handbook of Cognitive Psychophysiology: Central and Autonomic Nervous System Approaches* (pp. 471-98). New York: John Wiley.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.
- Fisher, K., Bassok, M., & Osterhout, L. (2010). When two plus two does not equal four: event-related potential responses to semantically incongruous arithmetic word problems. *Proceedings of the Cognitive Science Society*.
- Ford, J. (1978). Does P300 reflect template match or mismatch? In D. A. Otto (Ed.) *Multidisciplinary Perspectives in Event-related Brain Potential Research*, 181-83.

- Gibbs, R. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory and Cognition*, 8, 449-56.
- Gibbs, R. & Gonzales, G. (1985). Syntactic frozenness in processing and remembering idioms. *Cognition*, 20, 243-59.
- Gibbs, R., Nayak, N. & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language*, 28, 576-93.
- Glucksberg, S. (1989). Metaphors in conversation: How are they understood? Why are they used? *Metaphor and Symbolic Activity*, 4, 125-43.
- Goldberg, A. (1995). *Constructions*. Chicago: University of Chicago Press.
- Goldberg, A. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: OUP.
- Gregory, M., Raymond, W., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society*, 35, 151-166.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438-441.
- Hallin, A. E. & Van Lancker Sidtis, D. (2015). A closer look at formulaic language: Prosodic characteristics of Swedish proverbs. *Applied Linguistics*, doi: 10.1093/applin/amu078.
- Harris, C. (1996). Recognition of common word combinations: Towards a lexicon of variable sized units. *Research Report*, Psychology Department, Boston University, MA.
- Hernández, M., Costa A., & Inbal A. (2016). More than words: Multiword frequency

- effects in non-native speakers. *Language, Cognition and Neuroscience*. 31(6), 785-800.
- Jackendoff, R. (1995). The boundaries of the lexicon. In M. Everaert, E. van der Linden, A. Schenk, & R. Schreuder (Eds.), *Idioms: Structural and Psychological Perspectives*. Hillsdale, NJ: Erlbaum, 133-165.
- Janssen, N. & Barber, H. (2012). Phrase frequency effects in language production. *PLoS ONE*, 7(3): e33202. doi:10.1371/journal.pone.0033202.
- Janssen, N., & Caramazza, A. (2011). Lexical selection in multi-word production. *Frontiers in Language Sciences*, 2, 81.
- Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, 20, 137-194.
- Kok, A. (2001). On the utility of P300 amplitude as a measure of processing capacity. *Psychophysiology*, 38, 557-577.
- Kutas, M. & Iragui, V. (1998). The N400 in a semantic categorization task across six decades, *Electroencephalography and Clinical Neurophysiology*, 108, 456-471.
- 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, 621-647
- Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207, 203-05.
- Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161-163.
- Kutas, M., Van Petten, C., & Kluender, R. (2006). Psycholinguistics electrified II: 1994-

2005. In: M. Traxler and M.A. Gernsbacher (Eds.), *Handbook of Psycholinguistics, 2nd Edition*, New York: Elsevier. 659-724.
- Langacker, R. (1987). *Foundations of cognitive grammar*. Vol. 1. Stanford: Stanford University Press.
- Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, 61, (3), 326-338.
- Lau, E., Namyst, A., Fogel, A., & Delgado, T. (2016). A direct comparison of N400 effects of predictability and incongruity in adjective-noun combination. *Collabra*, 2(1), 13, 1-19.
- Laurent, J., Denhières, G., Passerieux, C., Iakimovac, G., & Hardy-Baylé, M. (2006). On understanding idiomatic language. *Brain Research*, 1068, 151-160.
- Libben, M., & Titone, D. (2008). The multidetermined nature of idiomatic expressions. *Memory & Cognition*, 36, 1103-1131.
- Matlock, T. & Heredia, R. (2002). Understanding phrasal verbs in monolinguals and bilinguals. In R. Heredia, and J. Altarriba (Eds.), *Bilingual Sentence Processing* (pp. 251-74). Elsevier Press.
- McClelland, J. & Rumelhart, D. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology. General*, 114, 159-188.
- McDonald, S., & Shillcock, R. (2003a). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research*, 43, 1735-1751.

- McDonald, S., & Shillcock, R. (2003b). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, 14(6), 648-652.
- Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63, 81-97.
- 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.
- Moon, R. E. (1998a). *Fixed Expressions and Text: A Study of the Distribution and Textual Behaviour of Fixed Expressions in English* (Oxford Studies in Lexicography and Lexicology). Oxford: Clarendon Press.
- Moon, R. E. (1998b). Frequencies and forms of phrasal lexemes in English. In A. P. Cowie (ed.), *Phraseology: Theory, Analysis, and Applications* (Oxford Studies in Lexicography and Lexicology), Oxford: Oxford University Press, pp. 79–100.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, 9, 97-113.
- Paulmann, S., Z. Ghareeb-Ali & C. Felser (2015). Neurophysiological markers of phrasal verb processing: Evidence from L1 and L2 speakers. In Roberto R. Heredia and Anna B. Cieslicka (eds.), *Bilingual figurative language processing*. Cambridge: Cambridge University Press, 245-267.
- Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee and P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: Benjamins, 137-157.
- Pinker, S. (1991). Rules of Language. *Science*, 253, 530-535.

- Pinker, S. (1999). *Words and Rules: The Ingredients of Language*. New York: Harper Collins.
- Pinker, S., & Ullman, M. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, 6, 456-463.
- Pollio, H., Barlow, J., Fine, H. & Pollio, M. (1977). *Psychology and the Poetics of Growth: Figurative Language in Psychology, Psychotherapy, and Education*. Lawrence Erlbaum.
- Rammell, C. S., Pisoni, S., & Van Lancker Sidtis, D. (in press). Perception of formulaic and novel expressions under acoustic degradation: Evidence for a unitary memory trace. *The Mental Lexicon*.
- Rayner, K., & Well, A. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin and Review*, 3, 504-509.
- Riess Jones, M. (1988). What does expectancy mean? *Behavioral and Brain Sciences*, 11, 387-88.
- Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlewsky, M. (2007). To predict or not to predict: Influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience*, 19, 1259-1274.
- Rommers, J., Dijkstra, T., & Bastiaansen, M. C. M. (2013). Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, 25(5), 762-776.
- Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs. In: D. Rumelhart & J. McClelland, J. (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. MIT Press, 216-271.

- Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Seidenberg, M., & MacDonald, M. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, 23(4), 569-588.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase ‘time and again’ matters: The role of phrasal frequency in the processing of multi-word sequences. *Journal of Experimental Psychology: Language, Memory, and Cognition*, 37(3), 776-784.
- Siyanova-Chanturia, A., Conklin, K., and Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and nonnative speakers. *Second Language Research*, 27, 251-72.
- Sosa, A. & MacFarlane, J. (2002). Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word *of*. *Brain and Language*, 83, 227-36.
- Strandburg, R., Marsh, J., Brown, W., Asarnow, R., Guthrie, D., & Higa, J. (1993). Event-related potentials in high-functioning adult autistics. *Neuropsychologia*, 31, 413-434.
- Swiney, D. & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behaviour*, 18, 523-34.
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, London: Harvard University Press.
- Titone, D., & Libben, M. (2014). Time-dependent effects of decomposability,

- familiarity, and literal plausibility on idiom meaning activation: A cross-modal priming investigation. *The Mental Lexicon*, 9, 3, 473-496.
- Tremblay, A., & Tucker, B.V. (2011). The effects of N-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon*, 6(2), 302-24.
- Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61(2), 569-613.
- Ullman, M. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience*, 2, 717-726.
- Ullman, M., Pancheva, R., Love, T., Yee, E., Swinney, D., & Hickok, G. (2005). Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language*, 93, 185-238.
- Van Lancker Sidtis, D. (2015). Formulaic language in an emergentist framework. In M. MacWhinney and W. O'Grady (Eds.), *Handbook of Language Emergence*, Wiley-Blackwell, pp. 578-599.
- Van Lancker Sidtis, D., Cameron, K., Bridges, K. & Sidtis, J. J. (2015). The formulaic schema in the minds of two generations of native speakers. *Ampersand*, 2, 39-48.
- Van Lancker Sidtis, D. & Postman, W.A. (2006). Formulaic expressions in spontaneous speech of left- and right-hemisphere damaged subjects. *Aphasiology*, 20(5), 411-426.
- Van Lancker, D. & Rallon, G. (2004). Tracking the incidence of formulaic expressions in everyday speech: Methods for classification and verification. *Language and*

Communication, 24, 207-240.

- Van Petten C., & Luka, B. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83, 176-90
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*, 18, 380-93.
- 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), 1682-700.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Figure 1. Nine representative electrodes (Experiment 1a – stimuli with conjunction).

Violations in blue, associates in red, binomials in black. Negativity is plotted upwards.

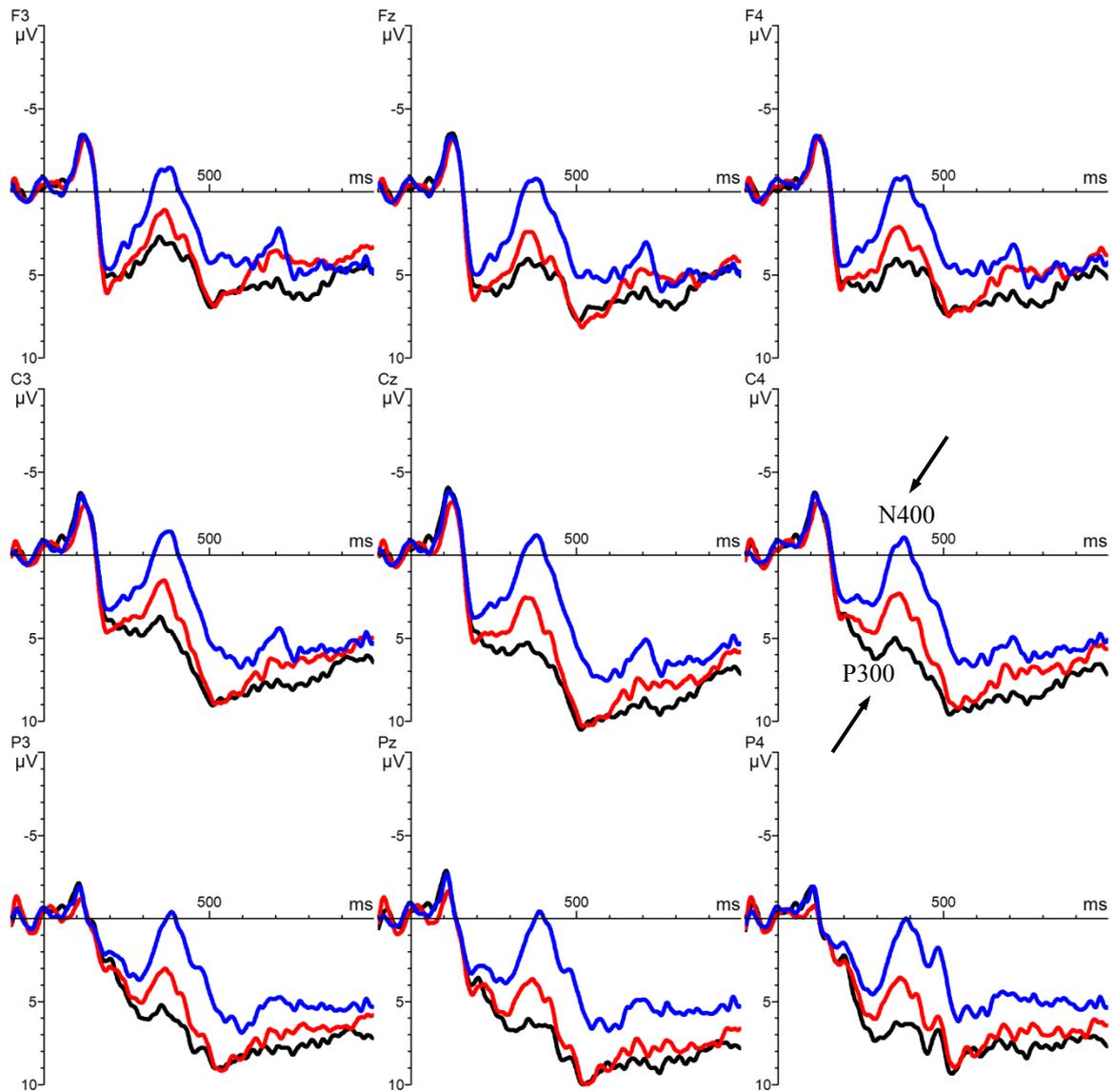


Figure 2. Isovoltage maps of the difference waves for the 250-350 ms time window
(Experiment 1a – stimuli with conjunction).

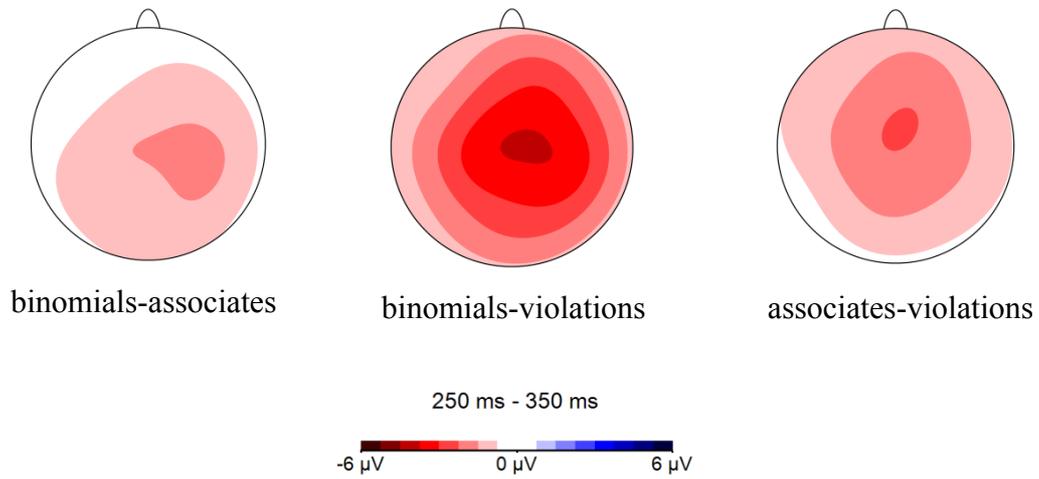


Figure 3. Isovoltage maps of the difference waves for the 350-450 ms time window
(Experiment 1a – stimuli with conjunction).

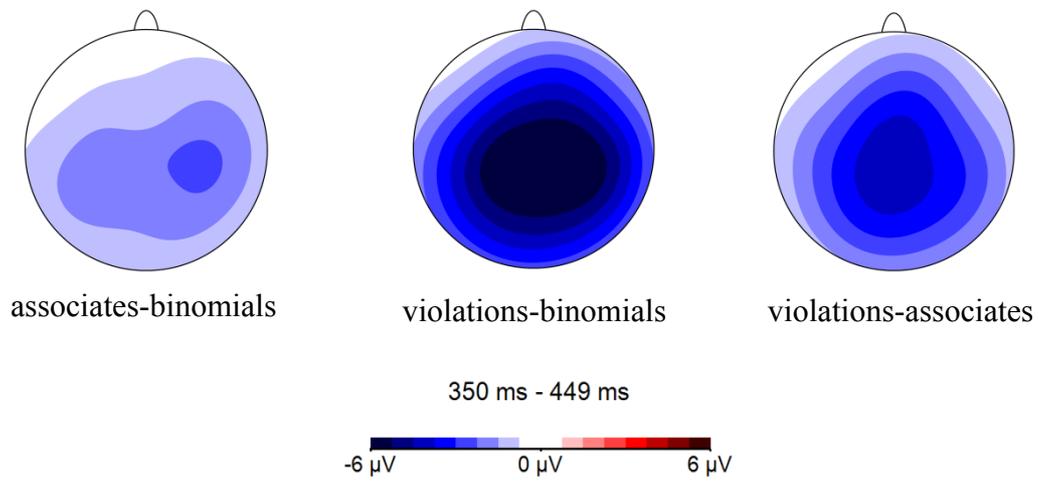


Figure 4. Nine representative electrodes (Experiment 1b – stimuli without conjunction). Violations in blue, associates in red, binomials in black. Negativity is plotted upwards.

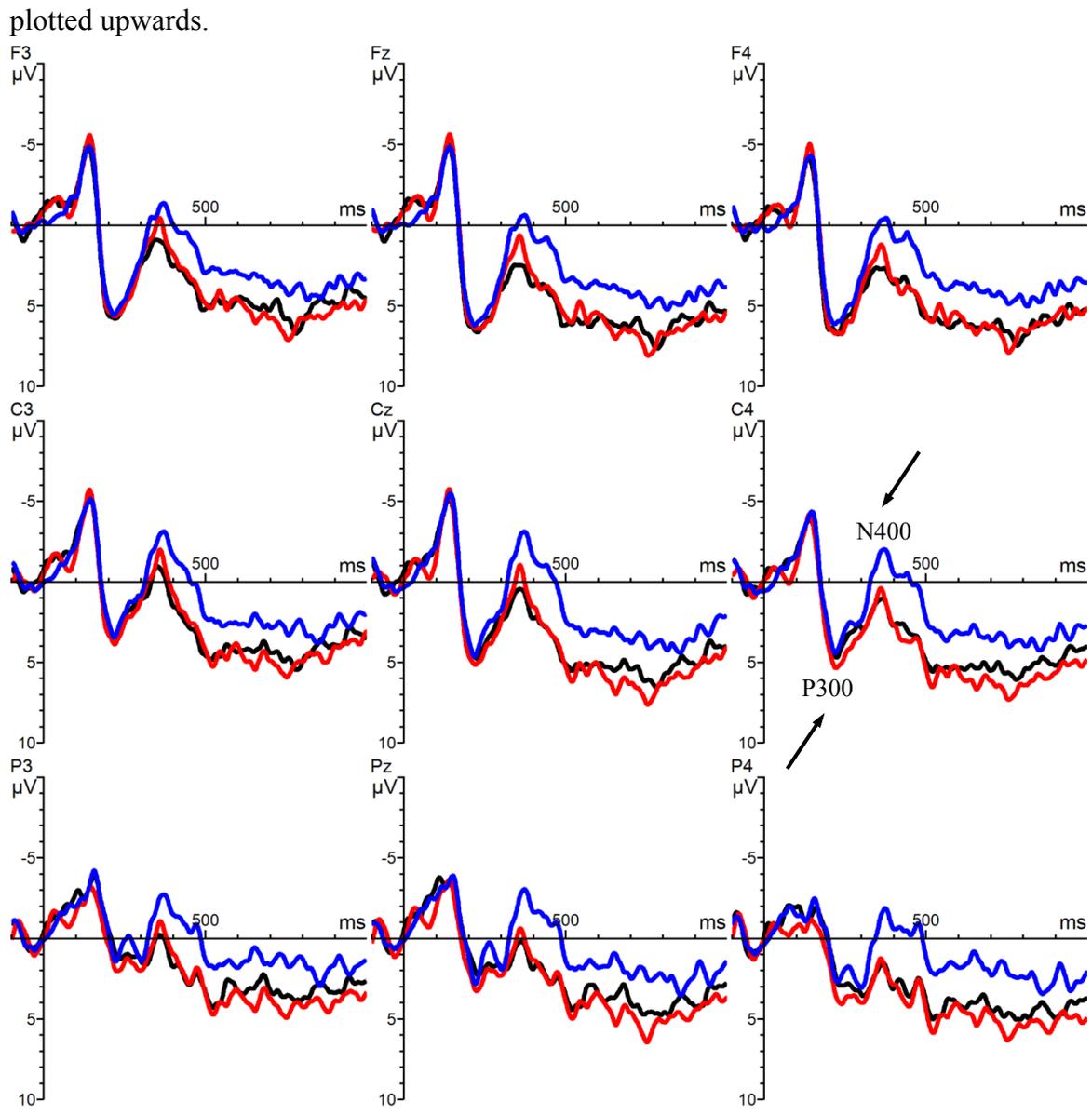


Figure 5. Isovoltage maps of the difference waves for the 250-350 ms time window
(Experiment 1b – stimuli without conjunction).

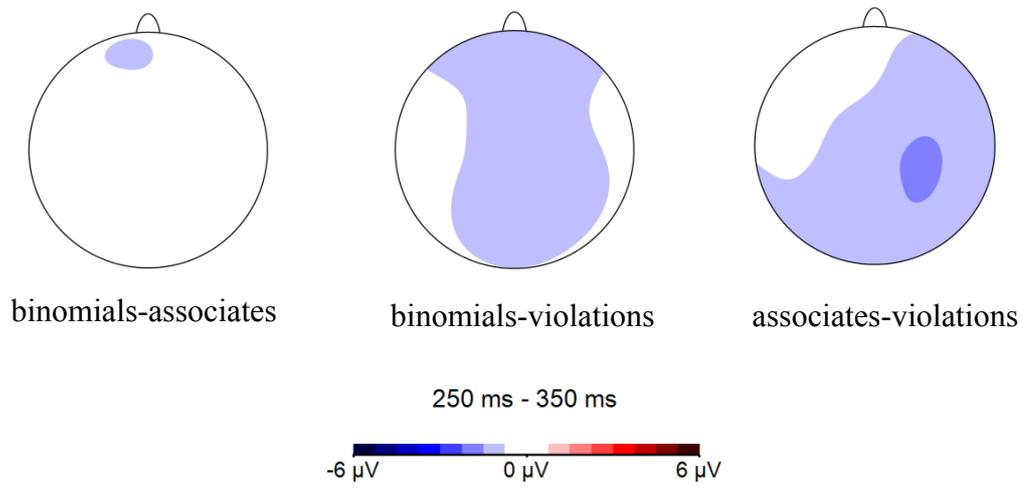


Figure 6. Isovoltage maps of the difference waves for the 350-450 ms time window
(Experiment 1b – stimuli without conjunction).

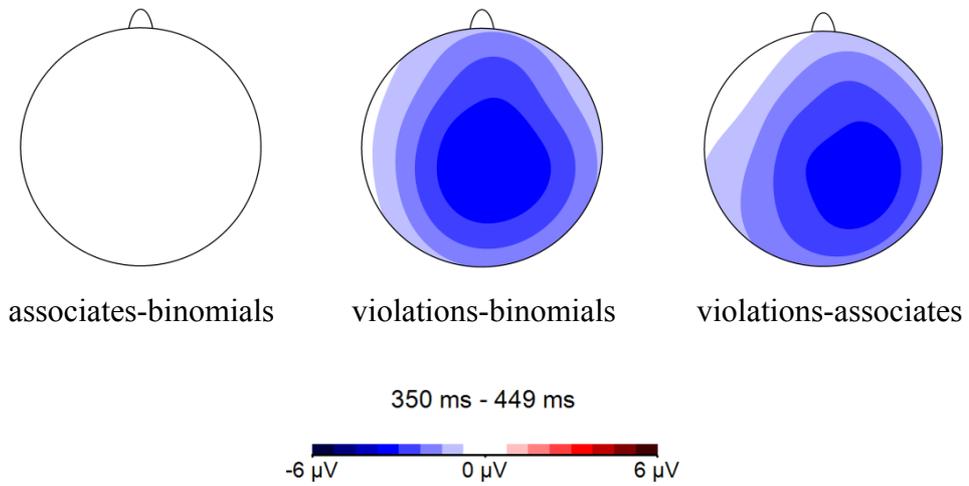
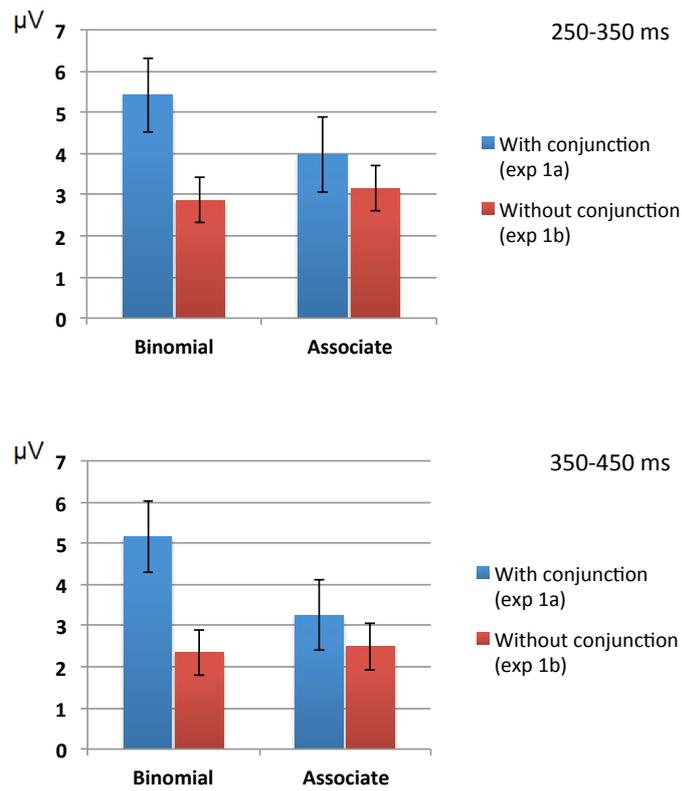


Figure 7. The interaction between Type (binomial vs. associate) and Experiment (1a – stimuli with conjunction vs. 1b – stimuli without conjunction) on mean voltages across the right lateral sites in the two time windows: 250-350 ms (top) and 350-450 ms (bottom).



Familiar, predictable phrases elicited a smaller N400 and a larger P300 than novel phrases. Reduced N400s suggest that familiar language is characterized by easier processing; larger P300s indicate that familiar phrases activate a 'template' that matches the upcoming information. Distinct mechanisms underlie the processing of highly predictable versus novel sequences.