Eye-tracking and ERPs in multi-word expression research

A state-of-the-art review of the method and findings

Anna Siyanova-Chanturia
Victoria University of Wellington

In recent years, there has been growing interest in the mechanisms that underlie on-line processing (comprehension and production) of units above the word level, known as multi-word expressions (MWEs). MWEs are a heterogeneous family of expressions that vary greatly in their linguistic properties but are perceived as highly conventional by native speakers. Extensive behavioural research has demonstrated that, due to their frequency and predictability, MWEs are processed differently from novel strings of language. At the very least, MWEs have been shown to be processed faster than matched control phrases. However, behavioural measures are limited in what they can tell us about MWE processing in the brain above and beyond the speed of processing. The present paper argues in favour of two powerful psycho- and neurolinguistic techniques — eye-tracking and event-related brain potentials (ERPs) — and presents a case for why these techniques are particularly suited for the investigation of phrasal frequency and predictive linguistic mechanisms. A number of studies that have drawn on these methods in their exploration of MWEs are reviewed, with a particular emphasis on the unique role of the method and its ability to tap into the underlying mechanisms implicated in MWE processing. It is argued that the two techniques complement, rather than duplicate each other, providing an ever richer account of the (psycho)linguistic phenomenon that MWEs are.

Keywords: multi-word expressions, on-line processing, behavioural, eye-tracking, event-related brain potentials

Much of the language we are exposed to on a daily basis is 'formulaic'. Despite the potentially infinite creativity of language, many words co-occur with some words more frequently than with other, seemingly synonymous, ones. In the present paper, we will refer to such frequently co-occurring word combinations as

The Mental Lexicon 8:2 (2013), **245–268.** DOI 10.1075/ml.8.2.06siy
ISSN 1871–1340 / E-ISSN 1871–1375

multi-word expressions (henceforth, MWEs). For the purpose of the present review, we adopt a very loose, rather inclusive definition of MWEs as highly familiar phrases that exhibit a certain degree of fixedness and are recognised as conventional by a native speaker. Arguably, MWEs can be of many different kinds, including but not limited to: collocations (*black coffee, morning sickness*), binomials (*bride and groom, heaven and hell*), multi-word verbs (*rely on, put up with*), idioms (*tie the knot, can't judge a book by its cover*), complex prepositions (*in support of, under the influence of*), speech formulae (*how's it going? good afternoon*), and so on. MWEs differ vastly in length (they can be between two and eight, or even more, words in length), literality (idioms are by definition figurative expressions, while many collocations and other types of MWEs are used literally), degree of fixedness (some allow internal modification, others do not), word class (closed-class vs. open-class), and so on. Even within a particular type, MWEs differ considerably (e.g., idioms can be decomposable or non-decomposable, ambiguous or non-ambiguous, etc.). Despite such vast differences, all MWEs can be said to be relatively frequent and, as a result, highly familiar and predictable word clusters. In the present paper, frequency and predictability will be shown to be quintessential characteristics of MWEs, ultimately defining the very nature of this remarkable (psycho)linguistic phenomenon and determining how such strings of language are processed in the brain.

The pervasiveness of MWEs in natural language has been taken to suggest that humans have the ability to "store" large numbers of frequent semi-preconstructed phrases (Sinclair, 1987). It has been argued that such word combinations form "chunks" in long-term memory (Ellis, 2001), and that it is easier and more economic to remember and use language in chunks rather than as a combination of single words (Langacker, 1987; Wray, 2002, 2008). In response to these (and other similar) claims, recent years have seen a surge in empirical, psycholinguistically-motivated, research on various kinds of MWEs.[1] A broad range of behavioural studies have demonstrated that, due to their frequency (and, hence, familiarity) and predictability, MWEs are processed differently from novel language. Comprehension and production studies with collocations, lexical bundles, phrasal verbs, and other types of MWEs have shown that, at the very least, MWEs enjoy quantitatively faster processing times compared to novel strings (Arnon & Snider, 2010; Conklin & Schmitt, 2008; Bannard & Matthews, 2008; Janssen & Barber, 2012; Janssen & Caramazza, 2011; Jiang & Nekrasova, 2007; Matlock & Hereida, 2002; Millar, 2011; Siyanova & Schmitt, 2008; Sosa & MacFarlane, 2002; Tremblay et al., 2011; Tremblay & Tucker, 2011). This finding, emphasising the role of phrasal frequency in language processing, has important implications for the nature of the mental lexicon and theories of language acquisition. Indeed, the processing advantage for MWEs found in behavioural studies has been proposed to argue against the traditional distinction between the lexicon, a collection of memorised forms, and grammar, a collection of rules (Arnon & Snider, 2010; Siyanova-Chanturia, Conklin, & van Heuven, 2011). Instead, it has been taken to support usage-based (Bybee, 1998; Goldberg, 2006; Tomasello, 2003) and exemplar-based models (Abbot-Smith & Tomasello, 2006; Bod, 2006; Pierrehumbert, 2001), according to which the basic unit of language acquisition is a construction (Goldberg, 2006; Tomasello, 2003). In line with these theories, all linguistic information (morphemes, words, or phrases) is represented and processed in a similar way, and thus it should be comparably affected by frequency manipulations.

Much of the research on MWEs to date has used various behavioural paradigms (e.g., self-paced reading, moving-window, grammaticality judgment, lexical decision, monitoring/word detection, sentence recall, priming, etc.). These techniques have been valuable in our gaining better understanding of MWE online processing.[2] Specifically, they have been useful in addressing the question of the speed of processing of MWEs versus control novel phrases in children and adults, first and second language speakers, as well as healthy individuals and patients with language impairment. However, behavioural methods are limited in that they can shed little light on MWE processing above and beyond the speed of processing. It is the aim of the present paper to review two powerful psycho- and neurolinguistic methods that have recently been used in the study of MWEs and have been shown to be particularly suited for the investigation of phrasal frequency and predictive linguistics mechanisms — the two quintessential characteristics of MWEs.

Present Review

A number of researchers have suggested using different techniques on the same or similar stimulus material in order to obtain converging evidence on a particular topic (Altarriba, Kroll, Scholl, & Rayner, 1996). As Rayner (1998) argues, any differences or similarities across different paradigms can deepen our understanding and enrich our knowledge of the processes and mechanisms involved. With this in mind, the present paper had as its aim to focus on two powerful methodologies that have only recently been used in MWE research: eye-tracking[3] and event-related brain potentials (ERPs).

The aim of the present paper is twofold. First, we briefly review each of the two methods in the context of MWEs,[4] demonstrating why these methods are particularly suited for the investigation of units above the word level and, critically, what they can tell us above and beyond the speed of processing. Second, we critically review a range of studies that have drawn on these methods in their investigation of various kinds of MWEs, putting emphasis on the unique role of the method and its ability to tap into the underlying mechanisms and neural correlates implicated in MWE processing. Throughout the paper, it will be shown that the two methods complement, rather than duplicate each other, providing an even richer account of the (psycho)linguistic phenomenon that MWEs are.

Eye Movements

During the course of reading, rapid eye movements, known as *saccades,* are made from one fixation point to another. In between saccadic movements, the eyes remain stationary. Such pauses are known as *fixations*. Unlike saccades, which do not contribute to information retrieval, fixations constitute rich data that can tell us what was fixated, for how long, how many times, and in what order (Liversedge, Paterson, & Pickering, 1998; Rayner, 2009). One of the benefits of the eye-tracking method is the ability to study not only words that were fixated during reading, but also those that were skipped. Not all words are fixated during natural reading. Some words are more likely to be skipped than others, especially higher frequency and shorter words (Altarriba et al., 1996; Brysbaert & Vitu, 1998; Brysbaert, Drieghe, & Vitu, 2005; Inhoff & Rayner, 1986; Rayner & Duffy, 1986). According to Carpenter and Just (1983), content words are fixated 85% of the time; in contrast, function words are fixated only 35% of the time. Predictability of a word given preceding (phrasal, sentential, or textual) context is one of the important factors (on a par with frequency and length) known to affect fixation durations, as well as the probability of being skipped (Balota, Pollatsek, & Rayner, 1985; Ehrlich & Rayner, 1981; Rayner & Well, 1996).

Another advantage that eye-tracking offers is the possibility to study regressive movements. Although most saccadic movements are from left to right (in alphabetic languages), around 10–15% of all saccadic movements are regressions — movements from right to left (Rayner, 2009). Ambiguity, polysemy, length, frequency, predictability of a word, and other factors are known to contribute to the number and length of regressions.

An important question in reading research concerns the size of the *perceptual span* — the amount of information a reader can extract during a given fixation. The size of the perceptual span in alphabetical languages is 3–4 characters to the left of a given fixation and 14–15 characters to the right of this fixation (McConkie & Rayner, 1975; Pollatsek, Rayner, & Balota, 1986; Rayner & Bertera, 1979; Rayner, Well, & Pollatsek, 1980). The relatively large perceptual span to the right of a fixation suggests that some information about the upcoming word may become available in the *parafovea* (Rayner, 1998). Higher frequency, shorter, and more predictable words are more likely to be processed parafoveally (and, potentially, skipped) than lower frequency, longer, and unexpected words (e.g., Inhoff & Rayner, 1986).

Arguably the greatest advantage of the eye-tracking paradigm is the possibility to analyse separately earlier and later stages of on-line reading. Early measures (e.g., first fixation duration and gaze duration) are sensitive to early processes in the comprehension of a text, such as lexical access and early integration of information. In contrast, late measures (e.g., total reading time and second pass reading time) are sensitive to later processes, such as information reanalysis, discourse integration, and recovery from processing difficulties (Paterson, Liversedge, & Underwood, 1999; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989). Because early and late measures tap into different processes, it is imperative to analyse both.

Given that the eye-movement method has the potential to offer one of the richest accounts of on-line language processing, one may want to analyse MWEs from a number of different perspectives. For example, the analysis of the number and duration of fixations, the length of saccades between fixations, the number of skippings and regressions within MWEs versus control novel phrases are likely to reveal the underlying mechanisms involved in the processing of highly familiar and uniquely predictable information. In addition, because MWEs are, by definition, at least two (and often more than two) words long, examination of a range of early and late measures is likely to offer critical insights into the various stages of phrasal processing.

Eye Movements in MWE Research

Although eye-tracking has long been used in the investigation of language processing at the word and sentence level, thus far, only three studies have employed this methodology in the context of MWEs. Of these, two have focussed on the processing of idiomatic expressions. This is hardly surprising, as idioms, in part due to their idiosyncrasy, atypicality, and saliency, are, arguably, the most archetypical of all MWEs. As pointed out above, idiom research, much unlike research on other types of MWEs, dates back to the 1970s. A number of hypotheses have been put forward to account for figurative versus literal meaning processing (in ambiguous idioms), and figurative versus novel phrase processing, the most prominent of which are the lexical representation hypothesis (Swinney & Cutler, 1979), the configuration hypothesis (Cacciari & Tabossi, 1988), and the idiom decomposition hypothesis (Gibbs, Nayak, & Cutting, 1989). So influential were the early theories that much of what has followed since has drawn on, or has been influenced by, them. However, much unlike the early (behavioural) idiom studies, the more recent ones, in particular, those employing the eye-tracking (and ERP) technique, have managed to address the issue of idiom processing and representation from a number of different perspectives.

Underwood, Schmitt and Galpin (2004) were the first ones to use the eye-movement paradigm to study how idioms are read in a first as well as second language. The target idioms were selected on the basis of (1) having a salient beginning, (2) the sequence not having a function word as its final word, (3) the sequence containing between four and eight words, and, finally, (4) the phrases being relatively frequent and predictable (although no cloze test was performed and so no measure was used to quantify the predictability of the chosen items). The experimental items were embedded in a story context. The use of the eye-tracking paradigm allowed the authors to present the stories in full, as one paragraph (rather than word by word, or sentence by sentence). Underwood et al. compared fixation count (the number of all fixations made within a given region of interest; not a measure of processing time) and fixation durations (a measure of processing time) for the terminal word of an idiomatic phrase (*honesty is the best policy*) and a novel phrase containing the same lexical item (*it seems that his policy of* …). A processing advantage for idioms was found for native speakers, with fewer and shorter fixations made in the idiom condition compared to the novel one. The authors argued that participants were better able to predict terminal words within idioms and, as a result, read them faster than the same words within novel phrases. For the non-native speaker group, a comparable advantage for the idiom was observed, although it only held true in the case of the number of fixations, and not for fixation durations — the only measure of processing time employed in the study.

The Underwood et al. (2004) findings are discussed in light of a prominent eye-movement model of reading, known as E-Z Reader (Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Rayner, & Pollatsek, 1999, 2003; Rayner, Reichle, & Pollatsek, 2000), and its five stages (familiarity check, lexical access, early saccadic programming, late saccadic programming, and saccadic movement). It is proposed that when a MWE (e.g., *the black sheep of the family*) is read, the constituent words become more predictable as they progress through the phrase, and the final word of the string (Word *n*) is almost redundant. The familiarity check is thus completed earlier than for the equivalent terminal word in novel context, resulting in more expedient reading (i.e., shorter fixation durations for native speakers). It is further argued that when looking at Word *n–1*, the familiarity check allows the reader to ascertain that the sequence is highly predictable, and the attention then moves to Word *n*. The familiarity check on Word *n*, done while the reader's eyes remain on Word *n–1*, also concludes that the word is familiar and predictable. If the check on the target Word *n* is completed during the early saccadic programming stage, then a decision may be reached to skip the target word altogether. Indeed, Underwood et al. (2004) report that not all target words (which were always content words) were fixated and thus it was concluded that the familiarity check enabled the skipping decision to be made. Unfortunately, no statistics are provided with regard to the skipping probability in the two experimental conditions and so no reliable conclusions can be made.

The Underwood et al. (2004) study is important because it is the first one to have used the eye-tracking method with MWEs. However, as with other studies that will be reviewed in the present paper, the Underwood et al. study has a number of shortcomings, on which it will be instructive to reflect. As was mentioned above, this method is unique in that one can use a range of (early and late) eye-movement measures. As Rayner (1998) succinctly put it, any single measure is "a pale reflection of the reality of cognitive processing" (p. 377). In the Underwood et al. study, only one measure of processing time was used (fixation durations). Further, we know that the eye-tracking technique is highly sensitive to frequency and length (of a word or a phrase), as well as the syntactic structure under investigation. This latter variable was not controlled for, potentially introducing a confound (e.g., *So as usual honesty is the best* <u>*policy*</u> *and Cindy was …* vs. *It seems that his* <u>*policy*</u> *of leaving things until the last minute …*. The target words within the idiomatic and novel phrase are clearly part of rather distinct syntactic structures and play different syntactic roles). Finally, it is not clear if the idioms used in the study were actually known to the non-native participants, which somewhat undermines the non-native results reported (the post-experiment test strongly suggested that up to half of the target idioms might have been unknown to non-native participants).

Another study that looked at idiom processing in a first and second language using the eye-tracking method is Siyanova-Chanturia, Conklin, and Schmitt (2011). These authors investigated idiom comprehension in a biasing story context. They looked at idiomatic expressions used figuratively (e.g., *at the end of the day* – 'eventually'), the same phrases used literally (e.g., *at the end of the day* – 'in the evening'), and matched novel phrases (e.g., *at the end of the war*). Again, the use of the eye-tracking technique allowed for naturalistic stimuli presentation (and reading) wherein the stories were read as one paragraph (rather than word by word, or sentence by sentence). Critically, the use of eye-tracking allowed the authors to conduct two types of analyses: (1) the entire region of interest and (2) before and after the recognition point (thus putting to test the configuration hypothesis [Cacciari & Tabossi, 1988]). Recognition point was defined as the point at which the expression becomes uniquely *recognisable* as idiomatic, and thus the words to the right of the recognition point may receive fewer fixations or be skipped.

The aim of the recognition point analysis was to find out where, in the course of reading, the processing advantage (or processing cost, for that matter) was largest — before or after the recognition point. Unlike Underwood et al. (2004), the study made use of three eye-tracking measures (two of which were measures of processing time):

1. *First pass reading time*: the sum of all fixation durations made within a region of interest before exiting either to the left or to the right (early measure)
2. *Total reading time*: the sum of all fixation durations made within a region of interest (late measure)
3. *Fixation count*: the number of all fixations made within a region of interest (not a measure of processing time)

The findings of this study confirmed the previously observed idiom processing advantage relative to novel language. It needs to be pointed out that although in native speakers, no significant differences were observed in the early measure, a processing advantage for idioms (*at the end of the day*) over novel phrases (*at the end of the war*) was found in the late measure and in the fixation count analysis. This was taken to suggest that not only were idiomatic expressions read faster than novel phrases, but that they also required less rereading and re-analysis. In addition, native speakers were not found to read the (lower frequency) literal meaning more slowly than the (higher frequency) figurative one. In contrast, L2 speakers did not read idioms more quickly than novel phrases. However, non-native speakers required more time to retrieve figurative senses than literal ones, even though the biasing context supported the figurative rendering. Interestingly, this slow-down was found to be mostly evident before the recognition point.

Siyanova-Chanturia, Conklin, and Schmitt (2011) touched upon an interesting finding — the absence of any significant differences in the early measure (when the differences were found to be significant in the late measure and the fixation count measure). It was concluded that early eye-tracking measures might not be suitable for the investigation of longer multi-word sequences (some idioms used in the study were up to eight words in length), and that a combination of late measures should be used instead.

Idioms aside, only one other type of MWEs has been investigated using the eye-movement paradigm. Siyanova-Chanturia, Conklin, and van Heuven (2011) examined the processing of frequent *binomial expressions* in native and non-native speakers. Binomial expressions are phrases formed by two content words from the same lexical class connected by the conjunction 'and', where one word order is more frequent than the other (e.g., *bride and groom* vs. *groom and bride*). The aim of the study was to find out whether speakers were sensitive to the frequency with which phrases occur in language. Participants read sentences containing binomial expressions (e.g., *bride and groom*) or their reversed forms (e.g., *groom and bride*), the two types of phrases being identical in syntax and meaning but having different phrasal frequency. Siyanova-Chanturia and colleagues used mixed-effects modelling with crossed random-effect factors for subjects and items and analysed three eye-tracking measures: first pass reading time, total pass reading time, and fixations count (the same measures as those used in Siyanova-Chanturia, Conklin, & Schmitt, 2011). The dependent variables were the three eye-tracking measures, while the predictors were as follows: phrasal frequency, content Word 1 frequency, content Word 2 frequency, phrase type (binomial or reversed), phrase length, association strength between content Word 1 and Word 2, and proficiency (native, non-native higher, and non-native lower proficiency). For each of the three dependent variables (separate models were built for the three eye-tracking measures), the authors started with a basic model, gradually adding or excluding predictors, based on whether or not they improved the model or entered in any meaningful interactions with other predictors. The mixed-effects modelling revealed that all three eye-tracking measures were significantly affected by phrase length, proficiency, phrase type, and phrasal frequency. Critically, a significant interaction was observed between proficiency and phrase type in the three measures, suggesting that proficiency (native, higher and lower non-native) played a crucial part in phrasal processing. Further, it was found that phrasal frequency significantly affected the eye-tracking measures, and that this effect was independent of the phrase type (which was partialed out). Overall, the authors demonstrated that both native and (higher proficiency) non-native speakers were sensitive to phrasal frequencies (i.e., frequent vs. infrequent) *and* phrasal configuration (i.e., binomial vs. reversed). The authors took their results to support the view according to which each and every occurrence of a linguistic unit (at the word, phrase, or sentence level) contributes to its degree of entrenchment in one's memory.

It is noteworthy that a comparable effect was found across all three measures analysed — first pass reading time (early), total reading time (late), and fixation count. This contrasts with the idiom results reported in Siyanova-Chanturia, Conklin, and Schmitt (2011), where the effect was observed in the late but not early measure. However, it is important to note that the phrases used in the binomial study were three words in length (Word 1 + and + Word 2) and were of a comparable syntactic structure; whereas those used in the idiom study were up to eight words in length, and varied in their syntactic structure. This highlights the importance of looking at a range of (early and late) eye-tracking measures, especially when the region of interest is larger than a single word (e.g., first pass reading time may be sensitive to frequency manipulations in shorter MWEs, such as binomials, but may not be an adequate measure to be used with longer MWEs, such as some idiomatic expressions).

The aim of the above review was to highlight the benefits of the eye-movement paradigm in the study of MWEs. Due to its ability to record skippings and regressions, availability of early and late processing measures, its allowing for a holistic, rather than word-by-word, mode of text presentation (which is particularly important given that MWEs are often thought to be represented as holistic units), as well as its ability to split the region of interest (e.g., into fixations made before and after the recognition point), the eye-tracking technique has the potential to become an invaluable tool in MWE enquiry. Undoubtedly, it offers unrivalled benefits compared to the many behavioural paradigms that have thus far been used far more frequently. There is, however, another method, borrowed from the field of neuroscience, that can offer comparably unique, rich, and informative, yet entirely new insights into MWE processing — electroencephalography. Below, this method and its application to MWE context are reviewed in some detail.

Event-Related Brain Potentials (ERPs)

Electroencephalography (EEG) is the recording of electrical activity produced by neurons in the brain. The observed EEG is believed to reflect the activity of a number of functionally distinct neuronal populations (Van Petten & Kutas, 1991). *Event-related brain potentials* (ERPs) are EEG responses time-locked to a particular stimulus and averaged over a large number of trials. One of the greatest advantages of the ERP methodology is its high temporal resolution. According to Kutas and Van Petten (1994), ERP measures are as close to immediate on-line processing as is technically possible. Another important benefit of ERPs is that not only can they tell us when something happened, but they can also inform us about the very nature of the cognitive processes involved, such as semantic or syntactic processing difficulty. As such, ERPs, unlike eye-tracking, are a *direct* reflection of the brain activity.

ERPs plotted against post-stimulus time are represented by a series of positive and negative peaks (Van Petten & Kutas, 1991). Such positive and negative waves are associated with different ERP components. A number of ERP components have been documented in psycholinguistic literature (LAN, N100, P200, P300, N400, P600, etc.). In what follows below, we briefly review two components that have been linked to MWE processing — the N400 and the P300.

N400

The N400 is a negative-going wave peaking between 300 and 500 ms following stimulus onset, most prominent in the centro-parietal area of the brain (Curran, Tucker, Kutas, & Posner, 1993). In the literature, the N400 has been shown to be sensitive to the violations of semantic information and real-world knowledge (Kutas & Hillyard, 1980; Hagoort, Hald, Bastiaansen, & Petersson, 2004). The N400 component was first described by Kutas and Hillyard (1980). In a set of experiments, ERPs evoked by semantically congruent sentence completions (e.g., *He spread the warm bread with butter*) were compared with those evoked by semantically incongruent ones (e.g., *He spread the warm bread with socks*). Incongruent completions elicited larger negativity most prominent over posterior scalp locations. The N400 component has also been linked to frequency and predictability manipulations, with more frequent and more predictable words eliciting smaller N400 amplitudes than less frequent and less predictable ones (frequency effects: Van Petten & Kutas, 1990; predictability effects: Kutas & Hillyard, 1984; Federmeier & Kutas, 1999; Wicha, Moreno, & Kutas, 2004).

P300

The P300 component, first discovered by Sutton, Braren, Zubin, & John (1965), is a positive deflection in voltage observed between 250 and 400 ms following stimulus onset. The P300 encompasses a number of distinct components, of which the P300a and the P300b are the most common. The P300a is associated with unexpected events; it is more anterior in its topography. The P300b, which is more posterior, is known to be elicited by infrequent task-relevant events. The P300 that is of interest to us (in light of MWE processing), and the one we will focus on below, is the P300b effect (henceforth, the P300).

Two theories have been proposed to account for the P300 effect: a *context-updating theory* (Donchin, 1981; Donchin & Coles, 1988; Donchin & Fabiani, 1991) and a *context-closure theory* (Verleger, 1988). Both accounts relate the P300 effects to expectancies that may arise during language 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 (i.e., event *n-1* implies that event *n* will follow). The major difference between the two theories is that 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).

A number of researchers have linked the P300 effect to "template matching" (Chao, Nielsen-Bohlman, & Knight, 1995; Ford, 1978). That is, participants may develop a neural representation, or a template, of the stimulus. The closer the match between the incoming information and the template, the larger the P300 amplitude (Kok, 2001). According to Kok (2001, p. 573), the P300 reflects "the awareness that a stimulus belongs or does not belong to the category of a certain memorised target event". In Roehm, Bornkessel-Schlesewsky, Rösler and Schlesewsky (2007, Experiment 1), participants read sentences such as *The opposite of black is …* which ended with: correct completion (*white*), related (*yellow*), or unrelated (*nice*). The expected completion (*white*) elicited a larger P300. The authors proposed that the P300 indexed "functionally distinct levels of predictive processing via distinct electrophysiological characteristics" (Roehm et al., 2007, p. 1260). Roehm and colleagues (2007, p. 1272) concluded that the highly expected antonymous adjective (*white*) elicited the P300 precisely 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)".

In sum, unlike eye-movement recordings, ERP components are considered to be a direct reflection of the brain activity (Luck, 2005), in that they allow us to investigate the very nature of the cognitive processes involved in language processing. As such, ERPs are likely to complement, rather than duplicate eye-tracking findings. Behavioural and eye-tracking studies have showed that MWEs are processed quantitatively faster than matched novel strings and may be (partially) skipped due to their high predictability. As noted by some researchers, behavioural measures represent the outcome of a set of cognitive processes, but are not, themselves, direct measures of these processes (e.g., Bartholow, 2010). In contrast, ERPs can more directly measure the responses that reflect cognitive processes of interest.

ERPs in MWE Research

Thus far, only a handful of studies have used ERPs in MWE enquiry. Of the many MWE types, one has been investigated in some detail — idioms. Others, such as collocations, lexical bundles, and complex prepositions, have received far less attention. As pointed out above, this is hardly surprising, as idioms are by far the most prototypical — and well-researched — type of MWEs.

The first studies to have used ERPs with MWEs are those that investigated the processing of figurative phrases versus matched novel phrases. In a recognition task with autistic patients and a control group, Strandburg, Marsh, Brown, Asarnow, Guthrie, and Higa (1993) recorded ERPs on the final word of meaningful idiomatic, meaningful literal (novel), and nonsensical phrases. The idioms used in the study had no literal equivalents. The stimuli were prepared such that the first word in the idiomatic phrase was also the first word in either the literal, or a nonsensical phrase (e.g., *vicious circle* [idiomatic] vs. *vicious dog* [literal]; *square deal* [idiomatic] vs. *square wind* [nonsensical]). All experimental stimuli were two-word sequences. The items were presented in isolation one word at a time, while participants decided whether or not each phrase was meaningful. Smaller N400 amplitudes were observed on frequent idiomatic phrases compared to literal and nonsensical ones, with an ordered increase in the N400 amplitudes from idiomatic to literal to nonsensical phrases in both autistic and control participants, suggesting progressive increases in their depth of processing. Thus, familiar, highly predictable phrases were processed more easily (at the electrophysiological level) compared to the other conditions. Interestingly, behavioural results[5] were found to mirror ERP data only in the case of the control group; autistic patients did not exhibit any processing differences between the three types of phrases at the behavioural level. This further emphasises the importance of a complementary use of different techniques.

In another study, Laurent, Denhières, Passerieux, Iakimovac, and Hardy-Baylé (2006) had participants perform a semantic relatedness task on French idioms that varied in their degree of salience (strongly and weakly salient). Salient meanings were defined as "coded meanings foremost on our mind due to conventionality, frequency, familiarity, or prototypicality" (Laurent et al., 2006, p. 151). Laurent and colleagues tested the hypothesis that it is not the literality, or the metaphoricalness, of idioms that is reflected in N400 amplitudes but their degree of salience. To this aim, they tested familiar ambiguous idioms (i.e., those that have both the figurative and literal meaning) that "enjoy a high degree of entrenchment or fixedness" (p. 153) and novel metaphorical expressions. The authors found that N400 amplitudes were significantly smaller on the last word of the strongly salient (i.e., conventional) idioms than on the last word of the weakly salient ones (i.e., new metaphor). Similar to Strandburg et al. (1993), reduced N400s on highly salient idioms were interpreted as easier integration and processing of frequent and predictable phrases relative to novel strings of language.

In a more recent study, Vespignani, Canal, Molinaro, Fonda, and Cacciari (2010) investigated the processing of Italian idioms before and after the recognition point (i.e., drawing on the configuration hypothesis [Cacciari & Tabossi, 1988]). Three conditions were investigated: idiomatic and two literal control conditions (substitution and violation). Phrases were embedded in sentence context and presented word by word. Akin to the two idiom studies reviewed above, Vespignani and colleagues found that idiomatic phrases elicited smaller N400s than literal phrases on the word that represented the recognition point of the phrase. Again, this was taken to suggest a processing advantage and easier semantic integration for conventional expressions.

The above studies (Strandburg et al., 1993; Laurent et al., 2006; Vespignani et al., 2010) clearly link conventionality and predictability of familiar idiomatic expressions to the N400 component. However, another ERP component — the P300 — has also been elicited by highly predictable incoming information. As was noted above, Roehm et al. (2007) showed that in a highly predictable context, such as, *The opposite of black is white*, where only one possible continuation is possible (*white*), the P300 effect was elicited on the highly predictable word (*white*). In contrast, the N400 effect was observed on the related word *yellow* and the unrelated word *nice* in the same context. In the above-mentioned study, Vespignani et al. (2010) hypothesised that the expectations driven by the activation of a prefabricated chunk (i.e., idiom) should be different from those driven by general discourse-based constraints. To test this hypothesis, ERP waveforms were compared for idioms and literal phrases before and after the recognition point. Before the recognition point, the difference between the literal and idiom conditions was attributed to the N400 (the finding reported above). However, following the recognition point, the idiomatic, but not literal, sentence completions resulted in the P300 effect. Because after the recognition point, only one correct idiom completion is possible (i.e., *take the bull by* … can only be completed by *the horns*), Vespignani and colleagues drew a parallel between their results and those of Roehm et al. (2007), where sentences like *The opposite of black is* … could only ever be completed by *white*. The authors concluded that the observed P300 effect is the result of *categorical template matching*, and that it "specifically operates for multi-word expressions … when the compositional analysis must be integrated with the retrieval of prefabricated meaning from the semantic memory" (p. 15).

The studies reviewed above have focused solely on one type of MWEs — idioms. Other types of MWEs have also been investigated using the ERP methodology, albeit to a lesser extent. Molinaro and Carreiras (2010) studied the comprehension of Spanish figurative and literal collocations (defined as sequences of words "which co-occur more often than would be expected by chance", p. 176) embedded in sentence context. ERPs were measured on the final word of the collocation and novel condition, the two being identical with the exception of the last word (e.g., literal collocation: *da segunda mano* "second-hand" vs. *da segunda vez* "second time"; figurative collocation: *en cuerpo y alma* "in body and soul" vs. *en cuerpo y mente* "in body and mind"). Molinaro and Carreiras found that both literal and figurative collocations elicited a P300 effect compared to novel phrases, suggesting that the recognition of the string leads to the pre-activation of the lexical item that completes the collocation. In contrast, meaning variation (i.e., figurative vs. literal) affected ERPs only in the N400 window (and not in the P300 window). The authors concluded that two distinct processing mechanisms were at work: an earlier one determined by high predictability of an upcoming word (P300), and a later one involved in semantic processing routines (N400).

One caveat with respect to this study pertains to the experimental items used. Although Molinaro and Carrerias (2010) administered a questionnaire to test the literality of the collocations (on the basis of which the items were assigned to either the literal and figurative condition), a closer inspection of the stimuli revealed that many of the "figurative collocations" used in the study were pure idioms (e.g., *sentir como Pedro por su casa* "feel as Pedro in his house" [feel at home]), while many of the "literal collocations" were, in fact, unambiguously figurative expressions (*entre la vida y muerte* "between life and death", *con la frente bien alta* "with your head high up" [be proud of yourself]). Thus, it is clear that the study's focus (at least, in part) was yet again on idiomatic expressions, similar to the studies described above. The somewhat blurred distinction between literal and figurative items used in the study may also explain why the authors failed to find an effect of literality of the collocation in the P300 time window.

Interestingly, the elicitation of the P300 component in highly constraining contexts, where there is a complete match between the expected and the observed configuration, has also been observed in other contexts, such as mathematical calculations (Fisher, Bassok, & Osterhout, 2010). Fisher et al. (2010) examined how people coordinated their arithmetic knowledge when they were presented with correct or erroneous additions and divisions (e.g., 12 + 3 = 15 vs. 12 + 3 = 14). As expected, erroneous calculations resulted in larger N400 amplitudes. Correct answers, on the other hand, exhibited pronounced positivity peaking around 300 ms after stimulus onset (for both divisions and additions). Although the authors did not go beyond explaining the observed effect in terms of larger N400s for arithmetic violations, and did not specifically focus on the positive peak elicited by correct answers, the morphological similarity between the positivity around 300 ms in Fisher and colleagues and the P300 found in MWE studies, is apparent. Thus, we can conclude that not only phrasal configurations, such idioms and collocations, can modulate the P300 component, but also a wider spectrum of other uniquely predictable scenarios (e.g., mathematical calculations).

Three more ERP studies merit our attention. They are important because they focus on different (from the above) types of MWEs — complex prepositions and lexical bundles. Unlike idioms and collocations, complex prepositions play the role of a function word in a sentence, implying that distinct (from the P300 and N400 reported above) neural mechanisms may be engaged in their processing. Complex prepositions are considered to be fixed-order lexical bundles with a prepositional function heading a complement within a prepositional phrase (Molinaro, Canal, Vespignani, Pesciarelli, & Cacciari, 2013). Molinaro, Vespignani, Canal, Fonda, & Cacciari (2008) measured ERPs on the highly expected final word (as suggested by its high cloze probability) within a complex preposition and an unexpected final word within a matched control phrase, embedded in the same context in Italian (e.g., *a sostegno di* "in support of" vs. *a sostegno per* "in support for"). The expected completion elicited an ERP complex associated with the processing of closed-class words — the N280 component followed by the N400–700. In contrast, the unexpected completion elicited only the N280, while the N400–700 was found to be reduced (unlike the N400 effect which is normally found to be larger [rather than smaller] for unexpected endings, compared to expected ones). The N280 is generally considered to reflect the lexical access to closed-class words (Neville, Mills, & Lawson, 1992), while the N400–700 is considered to be a vocabulary marker for closed-class words (Munte, Wieringa, Weyerts, Szentkuti, Matzke et al., 2001).

In another study, Molinaro et al. (2013) investigated the processing of complex prepositions (e.g., *in the hands of*) and the brain response when such frequent and familiar phrases were modified, for example by means of an adjective insertion (e.g., *in the capable hands of*). ERPs were measured on the noun (*hands*) when it was part of the formulaic unmodified and novel modified phrase. Molinaro and colleagues observed a smaller N400 on the noun in the modified condition, which was taken to suggest that the modification did not disrupt comprehension of the complex preposition. However, it was also found that phrase modification increased the processing load as evidenced by a (late) left anterior negativity (LAN) elicited on the noun.[6] These findings suggest that complex prepositions can be internally modified without losing their functional role. These findings are important as they imply that individual words within MWEs maintain their semantic and syntactic properties despite being part of a highly predictable and frequent phrase. The authors concluded that regular decompositional analyses were involved in the processing of MWEs (see Konopka & Bock, 2009 and Sprenger, Levelt, & Kempen, 2006 for a similar view). One criticism of the study regards the observed N400 effect. It is not clear why phrase modification (arguably rendering the phrase more difficult to process, given its divergence from the established pattern) elicited a *smaller* N400 effect compared to the unmodified phrase. If the insertion did not disrupt phrasal processing (as the authors argue), then one would expect to find comparable N400 amplitudes in the modified and unmodified condition. A smaller N400 in the modified condition implies that these novel phrases, in fact, enjoyed easier semantic integration than the more frequent and familiar unmodified phrases (which seems to go against the findings reported in other MWE studies).

Finally, Tremblay and Baayen (2010) investigated the processing of four-word lexical bundles (e.g., *in the middle of*) extracted from the British National Corpus, whose frequencies ranged from 0.03 to 105 occurrences per million. They found that the probability of occurrence of such MWEs, and individual word and trigram frequency, significantly affected participants' brain response. A frequency effect for the four-word sequence was found in the early time window — 110–150 ms after stimulus presentation — and was linked to the P1, N1, and P2 components. On the basis of the results obtained, the authors concluded that frequent MWEs "are retrieved in a holistic manner" (p. 170), and that more and less frequent four-word sequences are best viewed as two extremes on a frequency/probability continuum. However, a word of caution is warranted when interpreting the above results. First, the early effect observed in Tremblay and Baayen (2010) resulted in the modulation of the P1, N1, and P2 — the ERP components generally associated with attentional processes. Thus, it is unclear whether the findings reported in the study indeed support the proposition that MWEs "are retrieved in a holistic manner", or are suggestive of more general attentional processing. Second, in the ERP analysis, Tremblay and Baayen (2010) included only those variables that reached significance in the corresponding behavioural analysis. However, it has been documented in the literature that ERPs and behavioural data (elicited by the same stimuli) do not necessarily mirror each other's patterns of results (e.g., see Strandburg et al., 1993 and Siyanova-Chanturia, Pesciarelli, & Cacciari, 2012). Finally, given that existing ERP research with MWEs suggests modulations of later ERP components (e.g., the P300 and the N400), it is not clear why the authors chose to focus solely on the very early time windows and ERP components.

Conclusion

Unsurprisingly, of the many types of MWEs, only idioms have been investigated in some detail. Thus, any generalisations to other kinds of MWEs can only be tentative. However, the studies reviewed above do suggest that the processing of MWEs differs from novel language not only quantitatively — that is, in terms of the speed of processing; but also qualitatively — that is, fundamentally *distinct neural correlates* underlie on-line processing of novel language and MWEs. It is hoped that future research will endeavour to investigate a broader range of idiomatic and literal, compositional and non-compositional, fixed and semi-fixed MWEs including (but not limited to) collocations, binomials, lexical bundles, speech routines, multi-word verbs, and so on. Finally, it is worth noting that all ERP studies to date have focused entirely on MWE processing in a first language.

Whether the same or different neural correlates are involved in their processing in a second language is yet to be addressed in future research.

The aim of the present review was to show that eye-tracking and ERPs can be invaluable techniques in the study of the role that phrasal frequency and predictability play in natural language processing. The ability to detect skippings and regressions, availability of early and late processing measures, as well as naturalistic mode of stimuli presentation make the eye-tracking paradigm particularly useful in the investigation of MWEs; while a range of ERP components, being a direct reflection of the brain activity, can inform us about the very nature of the cognitive processes involved, offering not only *quantitative* (e.g., when something was activated) but also *qualitative* perspective (e.g., semantic vs. syntactic processing difficulties, general probabilistic expectations vs. categorical template matching). Thus, the two techniques complement, rather than duplicate each other, providing an ever richer account of the (psycho)linguistic phenomenon that MWEs are. Despite the still relative novelty of these two methods in MWE research, together, they have been able to contribute to our understanding of on-line processing of units above the word level in the way that behavioural measures have been unable to above and beyond the speed of processing.

Acknowledgements

The author would like to thank the two anonymous Reviewers as well as Phil Durrant for their helpful and insightful comments on the previous version of this paper.

Notes

1. It needs to be pointed out that (ambiguous) idioms and how they are processed given the availability of two distinct meanings (figurative and literal) have long been of interest to linguists and psycholinguists alike (Bobrow & Bell, 1973; Cacciari & Tabossi, 1988; Cutting & Bock, 1997; Gibbs, 1980; Gibbs & Gonzales, 1985; Gibbs, Nayak, & Cutting, 1989; Swinney & Cutler, 1979). In contrast, other types of MWEs (e.g., collocations, phrasal verbs, complex prepositions, lexical bundles, binomials, etc.) and issues pertinent to their processing had not been studied until only about a decade ago (Sosa & MacFarlane, 2002; Matlock & Hereida, 2002). At the time of writing, the vast majority of the psycholinguistic studies on MWEs that are not idioms appeared between 2008 and 2013.

2. We define on-line processing as one happening in real time, under significant time pressure, with participants not being aware of the purpose of the study and thus being unable to adopt strategies that may influence their processing. On-line studies contrast with off-line ones

(e.g., corpus studies or studies employing off-line measures, such as questionnaires). In on-line studies, reaction times and/or brain activity are recorded while participants perform a task in a laboratory setting.

3. In the present paper, we will use the terms eye-movement method and eye-tracking (a more colloquial term) interchangeably.

4. For a detailed overview of the eye-movement method, see Rayner (1998, 2009). For a detailed overview of the ERP methodology in the context of language processing, see Kaan (2007) and Osterhout, McLaughlin, and Bersick (1997).

5. In ERP experiments, it is customary to record both EEG and behavioural response (e.g., reaction times and accuracy).

6. A left anterior negativity (LAN) has been found in sentences with grammatical violations, for example, where the verb does not agree with the noun (Coulson, King, & Kutas, 1998), as well as in sentences with garden paths (Kaan & Swaab, 2003).

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.

Altarriba, J., Kroll, J., Scholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed language sentences: Evidence from eye fixation and naming times. *Memory and Cognition*, *24*, 477–492.

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

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.

Bartholow, B. D. (2010). On the role of conflict and control in social cognition: Event-related brain potential investigations. *Psychophysiology*, *47*, 201–212.

Bobrow, S., & Bell, S. (1973). On catching on to idiomatic expressions. *Memory and Cognition*, *1*(3), 343–346.

Bod, R. (2006). Exemplar-based syntax: How to get productivity from exemplars. *The Linguistic Review*, *23*, 291–320.

Brysbaert, M., & Vitu, E. (1998). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 125–148). Oxford: Elsevier.

Brysbaert, M., Drieghe, D., & Vitu, F. (2005). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Cognitive processes in eye guidance* (pp. 53–77). Oxford: Oxford University Press.

Bybee, J. (1998). The emergent lexicon. *Chicago Linguistic Society*, *34*, 421–435.

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, *27*, 668–683.
Carpenter, P., & Just, M. (1983). What your eyes do while your mind is reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 275–307). San Diego, CA: Academic Press.
Chao, L., Nielsen-Bohlman, L., & Knight, R. (1995). Auditory event-related potentials dissociate early and late memory processes. *Electroencephalography and Clinical Neurophysiology*, *96*, 157–168.
Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, *29*(1), 72–89.
Coulson, S., King, J., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*(1), 21–58.
Curran, T., Tucker, D., Kutas, M., & Posner, M. (1993). Topography of the N400: Brain electrical activity reflecting semantic expectancy. *Electroencephalography and Clinical Neurophysiology*, *88*, 188–209.
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.
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*, 357–374.
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 & M. Coles (Eds.), *Handbook of cognitive psychophysiology: Central and autonomic nervous system approaches* (pp. 471–498). New York: John Wiley.
Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word recognition and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*, 641–655.
Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33–68). Cambridge: Cambridge University Press.
Federmeier, K., & Kutas, M. (1999). A rose by any other name: Long term memory structure and sentence processing. *Journal of Memory and Language*, *41*, 469–495.
Fisher, K., Bassok, M., & Osterhout, L. (2010). When two plus two does not equal four: Event-related potential responses to incongruous arithmetic word problems. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1571–1576). Austin, TX: Cognitive Science Society.
Ford, J. (1978). Does P300 reflect template match 0mismatch? In D. A. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potential research* (pp. 181–183). Washington, DC: U.S. Government Printing Office.
Gibbs, R. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory and Cognition*, *8*, 449–456.
Gibbs, R., & Gonzales, G. (1985). Syntactic frozenness in processing and remembering idioms. *Cognition*, *20*, 243–259.
Gibbs, R., Nayak, N., & Cutting, C. (1989). How to kick the bucket and not decompose: Analysability and idiom processing. *Journal of Memory and Language*, *28*, 576–593.
Goldberg, A. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*, 438–441.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception and Psychophysics*, *40*, 431–439.

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.

Jiang, N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, *91*(3), 433–445.

Kaan, E. (2007). Event-Related potentials and language processing: A brief overview. *Language and Linguistics Compass*, *1*(6), 571–591.

Kaan, E., & Swaab, T. (2003). Electrophysiological evidence for serial sentence processing: A comparison between non-preferred and ungrammatical continuations. *Cognitive Brain Research*, *17*, 621–635.

Kok, A. (2001). On the utility of P300 amplitude as a measure of processing capacity. *Psychophysiology*, *38*, 557–577.

Konopka, A., & Bock, K. (2009). Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. *Cognitive Psychology*, *58*(1), 68–101.

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflectsemantic incongruity. *Science*, *207*, 203–205.

Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–163.

Kutas, M., & Van Petten, C. K. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83–143). San Diego, CA: Academic Press.

Langacker, R. (1987). *Foundations of cognitive grammar* (Vol. 1). Stanford: Stanford University Press.

Laurent, J., Denhières, G., Passerieux, C., Iakimovac, G., & Hardy-Baylé, M. (2006). On understanding idiomatic language. *Brain Research*, *1068*, 151–160.

Liversedge, S., Paterson, K., & Pickering, M. (1998). Eye movements and measures of reading time. In. G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 55–75). Oxford: Elsevier.

Luck, S. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Matlock, T., & Heredia, R. (2002). Understanding phrasal verbs in monolinguals and bilinguals. In R. Heredia & J. Altarriba (Eds.), *Bilingual sentence processing* (pp. 251–274). Amsterdam, Netherlands: North-Holland/Elsevier.

McConkie, G., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception and Psychophysics*, *17*, 578–586.

Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics*, *32*(2), 129–148.

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.

Molinaro, N., Canal, P., Vespignani, F., Pesciarelli, F., & Cacciari, C. (2013). Are complex function words processed as semantically empty strings? A reading time and ERP study of Collocational Complex Prepositions. *Language and Cognitive Processes*, *28*(6), 762–788.

Molinaro, N., Vespignani, F., Canal, P., Fonda, S., & Cacciari, C. (2008). Cloze-probability does not only affect N400 amplitude: The case of complex prepositions. *Psychophysiology*, *45*(6), 1008–1012.

Munte, T. F., Wieringa, B. M., Weyerts, H., Szentkuti, A., Matzke, M., & Johannes, S. (2001). Differences in brain potentials to open and closed class words: Class and frequency effects. *Neuropsychologia*, *39*, 91–102.

Neville, H. J., Mills, D. L., & Lawson, D. S. (1992). Fractionating language: Different neural subsystems with different sensitive periods. *Cerebral Cortex*, *2*, 244–258.

Osterhout, L., McLaughlin, J., & Bersick, M. (1997). Event-related brain potentials and human language. *Trends in Cognitive Sciences*, *1*(6), 203–209.

Paterson, K., Liversedge, S., & Underwood, G. (1999). The influence of focus operators on syntactic processing of short relative clause sentences. *The Quarterly Journal of Experimental Psychology*, *52*, 717–737.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–157). Amsterdam: John Benjamins.

Pollatsek, A., Rayner, K., & Balota, D. A. (1986). Inferences about eye movement control from the perceptual span in reading. *Perception and Psychophysics*, *40*, 123–130.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention during reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*, 1457–1506.

Rayner, K., & Bertera, J. H. (1979). Reading without a fovea. *Science*, *206*, 468–469.

Rayner, K., & Duffy, S. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition*, *14*, 191–201.

Rayner, K., Reichle, E. D., & Pollatsek, A. (2000). Eye movement control in reading: Updating the E-Z Reader model to account for initial fixation locations and refixations. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 701–719). Oxford, England: Elsevier.

Rayner, K., & Well, A. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin and Review*, *3*(4), 504–509.

Rayner, K., Sereno, S., Morris, R., Schmauder, A., & Clifton, C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognition Processes*, *4*, 21–49.

Rayner, K., Well, A., & Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Perception and Psychophysics*, *27*, 537–544.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*, 125–157.

Reichle, E. D., Rayner, K., & Pollatsek, A. (1999). Eye movement control in reading: Accounting for initial fixation locations and refixations within the E-Z Reader model. *Vision Research*, *39*, 4403–4411.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*, 445–526.

Riess Jones, M. (1988). What does expectancy mean? *Behavioral and Brain Sciences*, *11*, 387–388.
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*, 1259–1274.
Sinclair, J. M. (1987). Collocation: A progress report. In R. Steele & T. Threadgold (Eds.), *Language topics: Essays in honour of Michael Halliday* (Vol. 2, pp. 319–331). Amsterdam: John Benjamins.
Siyanova-Chanturia, A., Conklin, K., & 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*, 251–272.
Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. (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*, 776–784.
Siyanova-Chanturia A., Pesciarelli F., & Cacciari C. (2012). The electrophysiological underpinnings of processing gender stereotypes in language. *PLoS ONE*, *7*(12), e48712. doi:10.1371/journal.pone.0048712
Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, *64*(3), 429–458.
Sosa, A., & MacFarlane, J. (2002). Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word of. *Brain and Language*, *83*, 227–236.
Sprenger, S., Levelt, W., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, *54*, 161–184.
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.
Sutton, S., Braren, M., Zubin, J., & John E. (1965). Evoked potentials correlates of stimulus uncertainty. *Science*, *150*, 1187–1188.
Swiney, D., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behaviour*, *18*, 523–534.
Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, London: Harvard University Press.
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 (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151–173). London, UK: The Continuum International Publishing Group.
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.
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), 302–324.
Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 153–172). Amsterdam: John Benjamins.
Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*, *18*, 380–393.
Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context in open-and closed-class words. *Memory and Cognition*, *19*, 95–112.

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.

Wicha, N., Moreno, E., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*, 1272–1288.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.

Corresponding Address

Anna Siyanova-Chanturia, PhD
Victoria University of Wellington
Wellington, NEW ZEALAND

anna.siyanova@vuw.ac.nz